Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale

In Amministrazione, Finanza e Controllo

Tesi di Laurea

Utilizzo delle Reti Bayesiane per il monitoraggio della continuità aziendale

Relatore
Slanzi Debora

Laureando
Moschioni Silvia
Matricola 885892

Anno Accademico
2024 / 2025

Sommario

Introduzione **1**

1 Inquadramento teorico **3**

1.1 Cos'è la continuità aziendale - Contesto normativo *3*

1.1.1 Soglie e fattori da monitorare – ISA 570 E CNDCEC 6

1.2 Il modello predittivo *9*

1.2.1 Approccio tradizionale e approccio predittivo 9

1.2.2 Il contributo di Altman e Alberici 10

1.2.3 Il modello di Machine Learning Bayesiano 14

2 Le reti Bayesiane **19**

2.1 Teoria della probabilità *20*

2.2 Introduzione alle reti Bayesiane *25*

2.2.1 La struttura – componente qualitativa 26

2.2.2 I parametri – componente quantitativa 28

2.3 Dipendenza e indipendenza nelle reti Bayesiane *30*

2.4 Inferenza nelle reti Bayesiane *36*

2.4.1 Inferenza esatta e approssimata 39

2.4.2 Inferenza esatta – strutture a catena 41

2.4.3 Inferenza esatta – strutture a polialbero 43

2.4.4 Inferenza esatta – reti a connessioni multiple 48

2.5 Apprendimento delle reti Bayesiane *52*

2.5.1 Grafo noto 52

2.5.2 Grafo non noto 53

2.5.3 Approcci ibridi 58

2.6 Estensioni *60*

2.6.1 Gaussian Bayesian Network 60

2.6.2 Conditional Gaussian Bayesian Network 62

3 La continuità aziendale: caso di studio **65**

3.1 Raccolta e costruzione del database *65*

3.1.1 Software di supporto R 66

3.2 Pulizia dei dati e selezione delle variabili *67*

3.2.1 Gestione dei dati mancanti 69

3.2.2 Stima valori mancanti ... 72

3.3 Analisi descrittiva delle variabili ... 75

3.3.1 Area Geografica ... 75

3.3.2 Codice ATECO ... 78

3.3.3 Stato giuridico e Procedure di Cessazione ... 80

3.3.4 Numero dipendenti – ultimo anno disponibile ... 84

3.3.5 Indici e Voci di bilancio - ultimo anno disponibile ... 84

3.3.6 Indici e Voci di bilancio – andamento medio e mediano ... 92

3.4 Correlazioni tra variabili ... 102

3.4.1 Valutazione comparativa delle medie nei due gruppi aziendali ... 108

4 Le reti Bayesiane per lo studio della continuità aziendale ... 111

4.1 Apprendimento della rete ... 113

4.1.1 Rete Bayesiana ultimo anno ... 115

4.1.2 La Rete Bayesiana nei diversi anni contabili ... 118

4.1.3 Analisi orizzontale e verticale: confronto tra le reti ... 120

4.1.4 Il nodo continuità: la rete previsionale ... 134

4.2 Costruzione del modello ... 138

4.2.1 Matrice di confusione e affidabilità del modello ... 138

4.3 Inferenza ... 143

4.3.1 Analisi di scenario ... 144

4.3.2 Scenario evolutivo ... 152

Conclusione ... 159

Appendici ... 163

Appendice A – Indicatori di rischio secondo ISA 570 ... 163

Appendice B – Indici di bilancio da monitorare secondo il CNDCEC ... 165

Appendice C – Elenco variabili del database utilizzato nello studio ... 171

Appendice D – Calcoli di approfondimento, Area Geografica ... 177

Appendice E – Calcoli di approfondimento, codice ATECO ... 185

Appendice F – Dati completi, Stato giuridico e Procedure Concorsuali ... 187

Appendice G – Dati completi, Indici e voci di bilancio ... 193

Appendice H – Dati completi, correlazione e test di Mann-Whitney ... 197

Appendice I – Legenda grafici DAG *207*

Appendice L – Variabili utilizzate per l'apprendimento della rete *209*

Appendice M – Dati per analisi inferenziale predittiva e classificazione *211*

Appendice N – Codice R *211*

Elenco delle figure **226**

Elenco delle tabelle **230**

Bibliografia e Sitografia **232**

Introduzione

Il tema della stabilità economico-finanziaria delle aziende ha da sempre rappresentato un elemento cardine per la salute del sistema economico. Negli ultimi anni, tuttavia, è cresciuta in modo significativo l'attenzione verso lo sviluppo di strumenti e tecniche in grado di agevolare l'individuazione tempestiva di indicatori di squilibrio, potenzialmente in grado di condurre a situazioni di crisi grave e alla perdita del controllo dell'equilibrio economico-finanziario.

In questo contesto, l'approccio statistico assume un ruolo fondamentale nel tradurre fenomeni economici complessi in modelli interpretabili e utilizzabili a fini previsionali.

I modelli statistici tradizionali applicati allo studio del rischio di insolvenza hanno fornito contributi rilevanti in termini descrittivi ed esplicativi, ma risultano spesso limitati da ipotesi restrittive e da una struttura poco flessibile, che ne riduce l'efficacia nell'analisi di fenomeni complessi e dinamici. L'evoluzione delle metodologie statistiche ha tuttavia consentito di superare tali limiti, spostando l'attenzione da una lettura puramente retrospettiva dei dati a un approccio maggiormente previsionale.

In questo quadro si inseriscono le tecniche di Machine Learning, che permettono di analizzare grandi quantità di informazioni, cogliere interdipendenze tra le variabili e gestire elevati livelli di incertezza. L'analisi statistica assume così un ruolo centrale non solo nella descrizione dei fenomeni economico-aziendali, ma soprattutto nel supporto alle valutazioni prospettiche, rendendo possibile un monitoraggio più efficace della continuità aziendale.

L'obiettivo della presente tesi è analizzare l'applicabilità delle reti Bayesiane al monitoraggio della continuità aziendale, sviluppando un modello predittivo in grado di stimare la probabilità che un'impresa rispetti i presupposti di continuità sulla base dei dati di bilancio e di altre caratteristiche strutturali.

Attraverso l'analisi di scenari reali, il lavoro intende valutarne l'efficacia, evidenziandone potenzialità e limiti.

Il primo capitolo introduce i concetti fondamentali relativi alla continuità aziendale e presenta gli indicatori individuati dal Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili come principali strumenti di monitoraggio dello stato di salute dell'impresa.

Il secondo capitolo approfondisce i fondamenti statistici alla base dell'approccio adottato, partendo dalla probabilità condizionata fino ad arrivare alla struttura e al funzionamento delle reti Bayesiane. Viene illustrato come tali modelli consentano di rappresentare relazioni di dipendenza e indipendenza tra variabili e di effettuare inferenza anche in contesti caratterizzati da incertezza informativa.

Nel terzo capitolo vengono presentati i dati reali utilizzati nel caso di studio e le attività di pre-processing effettuate. Il capitolo approfondisce l'analisi descrittiva univariata delle variabili, dalle informazioni anagrafiche di natura categoriale fino all'analisi multivariata degli indicatori economico-finanziari selezionati. Viene inoltre analizzata la correlazione tra le variabili al fine di individuare eventuali fenomeni di multicollinearità e di valutare la relazione con la variabile target, rappresentativa dello stato di salute dell'impresa.

Nel quarto e ultimo capitolo viene sviluppato il modello di analisi previsionale basato su reti Bayesiane. La rete, strutturata come un DAG multilivello articolato su quattro livelli, uno per ciascun anno considerato, consente di stimare la probabilità che un'impresa, date specifiche caratteristiche economico-finanziarie, manifesti problematiche di continuità aziendale e permette di individuare i fattori con maggiore capacità predittiva, fornendo un supporto decisionale utile alla prevenzione delle crisi d'impresa.

1 Inquadramento teorico

1.1 Cos'è la continuità aziendale - Contesto normativo

Il principio di continuità aziendale può definirsi di recente definizione in quanto nonostante trovi applicazione nel Codice Civile sin dal 1991, è diventato un concetto di centrale importanza solo dal 2019 quando il Codice della Crisi ne ha chiarito gli aspetti principali.

Con il D.Lgs. n. 127 del 9 aprile 1991 venne introdotto l'articolo 2423-bis del Codice Civile ai sensi del quale "[…] la valutazione delle voci deve essere fatta secondo prudenza e nella prospettiva della continuazione dell'attività […]"[1] per la prima volta viene accostato ai principi classici di redazione del bilancio anche un principio che introduce il controllo dell'andamento futuro delle attività aziendali.

Nel Codice Civile troviamo un altro articolo che definisce in maniera più specifica quali sono i comportamenti che l'imprenditore deve mettere in atto ai fini del corretto adempimento dell'obbligo di continuità aziendale. Nello specifico l'articolo 2086 cita "*L'imprenditore, […], ha il dovere di istituire un assetto organizzativo, amministrativo e contabile […], anche in funzione della rilevazione tempestiva della crisi dell'impresa e della perdita della continuità aziendale, nonché di attivarsi senza indugio per l'adozione e l'attuazione di uno degli strumenti previsti dall'ordinamento per il superamento della crisi e il recupero della continuità aziendale*"[2], questo comma venne aggiunto nel 2019 con l'entrata in vigore del Codice della Crisi e dell'Insolvenza che introdusse numerose definizioni e precisazioni in tema di continuità aziendale; come si può evincere dall'articolo appena citato il legislatore attribuì all'imprenditore l'obbligo di stabilire un assetto organizzativo adeguato alla previsione tempestiva di situazioni di criticità, nonché la responsabilità di adottare in maniera tempestiva gli strumenti adeguati volti al superamento delle criticità riscontrate.

Su questi articoli si basa inoltre il principio di contabilità italiana numero 11 redatto dall'Organismo Italiano di Contabilità (OIC). Principio che nei paragrafi 21 e successivi

[1] Italia, Codice Civile, Regio Decreto 16 marzo 1942, n. 262, art. 2423-bis, introdotto dall'art. 6 del D.Lgs. 9 aprile 1991, n. 127, e successive modificazioni, consultabile su Gazzetta Ufficiale.

[2] Italia, Codice Civile, Regio Decreto 16 marzo 1942, n. 262, art. 2086, modificato dall'art. 375 del D.Lgs. 12 gennaio 2019, n. 14, *Codice della crisi d'impresa e dell'insolvenza*, consultabile su Gazzetta Ufficiale.

enuncia il criterio di continuità aziendale in relazione alla redazione del bilancio. Vengono specificati gli obblighi della direzione aziendale, tali per cui l'azienda deve essere in grado di continuare a costituire un complesso economico funzionante destinato alla produzione di reddito per un periodo di almeno dodici mesi[3].

Viene inoltre specificato che nei casi in cui non sia possibile garantire una previsione di continuità aziendale per il periodo di tempo stabilito, il bilancio deve comunque essere redatto secondo il principio di continuità aziendale, considerando, però, l'orizzonte temporale limitato.

Troviamo di particolare rilevanza anche il contenuto del principio ISA 570[4], che regolamenta tutti i comportamenti in capo al revisore per quanto riguarda la verifica della continuità aziendale nelle aziende revisionate. Nello specifico il revisore deve certificare che la direzione aziendale abbia operato in modo adeguato garantendo il proseguimento delle attività come un'entità in funzionamento e verificare che non siano scaturiti eventi tali da far sorgere un dubbio significativo da parte del revisore.

La "Legge fallimentare"[5] risalente al 1942 fu profondamente modificata con l'emanazione del Decreto Legislativo n. 14 del 12 gennaio 2019, "*Codice della Crisi e dell'Insolvenza*"[6], che però entrò in vigore solo il 15 luglio 2022 grazie alle modifiche apportate dal Decreto Legislativo del 17 giugno 2022, n. 83[7].

Si tratta di una legge che racchiude tutta la normativa in materia di crisi e insolvenza d'impresa, con l'obbiettivo ultimo di creare una "guida operativa" per fronteggiare situazioni di difficoltà economico-finanziaria nelle imprese Italiane.

---

[3] OIC 11, *Finalità e postulati del bilancio d'esercizio*, Organismo Italiano di Contabilità, ultimo aggiornamento 22 dicembre 2017.

[4] ISA Italia 570, *Continuità aziendale*, adottato con Determina della Ragioneria Generale dello Stato del 23 dicembre 2014, ai sensi dal D.Lgs. 39/2010.

[5] Italia, Regio Decreto 16 marzo 1942, n. 267, Disciplina del fallimento, del concordato preventivo, dell'amministrazione controllata e della liquidazione coatta amministrativa, e successive modificazioni, consultabile su Gazzetta Ufficiale.

[6] Italia, Decreto Legislativo 12 gennaio 2019, n. 14, *Codice della crisi d'impresa e dell'insolvenza*, consultabile su Gazzetta Ufficiale.

[7] Italia, Decreto Legislativo 17 giugno 2022, n. 83, Modifiche al codice della crisi d'impresa e dell'insolvenza, in attuazione della direttiva (UE) 2019/1023, in materia di ristrutturazione e insolvenza, consultabile su Gazzetta Ufficiale.

Il suddetto Codice inizia specificando alcune definizioni nell'art 2:

- "crisi", è definita come stato del debitore che rende probabile l'insolvenza e viene individuato con inadeguatezza dei flussi di cassa, viene definito anche un arco temporale pari a dodici mesi;
- "insolvenza", è definita come stato del debitore che non è più in grado di far fronte alle sue obbligazioni;
- "strumenti di regolazione della crisi e dell'insolvenza", vengono introdotte delle misure di risanamento di tipo continuativo diverse dalle misure di risanamento di tipo liquidatorio, quali, liquidazione giudiziale e controllata (ex fallimento); questi strumenti hanno l'obbiettivo di risanare l'impresa in crisi tramite la modifica della composizione, dello stato o della struttura delle sue attività, passività o del capitale.

La parte principale del codice definisce alcune procedure volte a prevedere tempestivamente lo stato di crisi e risanare l'azienda in tempo utile prima che sia necessario mettere in atto una procedura di liquidazione giudiziale.

Nello specifico il Titolo II del Codice definisce la composizione negoziata della crisi, quindi la facoltà dell'imprenditore in stato di squilibrio patrimoniale o economico-finanziario, di chiedere la nomina di un esperto che agevoli le trattative tra imprenditore e creditori, con il fine ultimo di individuare una soluzione per il superamento dello stato di difficoltà.

Il Titolo IV, invece, racchiude tutti gli "strumenti di regolazione della crisi e dell'insolvenza" e si suddivide in:

- Capo I, dove vengono regolamentati i piani di risanamento e gli accordi di ristrutturazione; il piano di risanamento consiste in un programma, predisposto dall'imprenditore in stato di crisi, volto a consentire il risanamento dell'esposizione debitoria e che deve essere sottoposto ad attestazione di veridicità e fattibilità da parte di un professionista indipendente.
  L'accordo di ristrutturazione dei debiti, invece, consiste in un insieme di accordi predisposti dall'imprenditore in crisi con i creditori che rappresentano almeno il sessanta percento dei crediti dell'azienda. Importante è sottolineare che l'accordo di ristrutturazione richiede l'omologazione, quindi l'intervento e l'approvazione, da parte del tribunale.

- Capo II, tratta i casi di *sovraindebitamento*, che, come definito dall'art. 2 dello stesso Codice, riguarda tutti i debitori non assoggettabili alle comuni misure di liquidazione; nello specifico si parla di ristrutturazione dei debiti e di concordati minori; i concordati minori consistono in accordi tra i debitori e i creditori dell'azienda e servono a prevedere il soddisfacimento dei creditori sotto qualsiasi forma.
- Capo III, disciplina la materia del concordato preventivo, che riguarda un accordo proposto dall'imprenditore in stato di crisi, realizzato sulla base di un piano di risanamento, che ha l'obbiettivo si soddisfare i creditori in misura non inferiore a quella che avrebbero in caso di liquidazione giudiziale; il concordato può essere di tipo continuativo o liquidatorio in base alla sostenibilità del business ancora in essere, nello specifico il concordato continuativo mira a mantenere l'attività aziendale, mentre il concordato liquidatorio prevede la vendita dei beni per soddisfare i creditori.

Dal codice della crisi possiamo quindi derivare una definizione di continuità aziendale intesa come la capacità dell'impresa di proseguire la propria attività senza dover ricorrere agli strumenti di regolazione della crisi, grazie ad una struttura patrimoniale, economica e finanziaria adeguata che permetta la corretta copertura delle obbligazioni prese per un arco di almeno dodici mesi.

1.1.1 Soglie e fattori da monitorare – ISA 570 E CNDCEC

Gli istituti internazionali di competenza, nei documenti ufficiali redatti ai fini della regolamentazione del concetto di continuità aziendale hanno a loro volta stabilito delle linee guida per aiutare i revisori e gli organismi aziendali a monitorare correttamente i valori economici finanziari e patrimoniali delle proprie organizzazioni.

In particolare, l'*International Auditing and Assurance Standards Board* (IAASB)[8], ovvero l'istituto che si occupa della redazione e della diffusione dei principi internazionali di revisione ISA, si è premurato di inserire all'interno dell'ISA 570 degli elenchi contenenti i fattori e gli eventi da analizzare che potrebbero costituire un rischio per la continuità aziendale. Per quanto

[8] Organismo indipendente, fondato nel 1978, che si occupa della redazione di standard internazionali per l'auditing – International Standard on Auditing (ISA), con il fine di promuovere la qualità e l'uniformità della revisione contabile a livello globale.

riguarda le procedure di valutazione del rischio, queste sono suddivide in: indicatori finanziari, dove vengono riportare le procedure di controllo relative alle situazioni patrimoniali e finanziarie dell'impresa; indicatori gestionali, dove si evidenziano le situazioni di difficoltà a livello di personale o di posizionamento nel mercato; altri indicatori, dove troviamo principalmente controlli relativi al rispetto della normativa e delle procedure legali. Gli elenchi vengono interamente riportati per completezza nell'Appendice A.

A livello nazionale, si è pronunciato anche il Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC)[9], che nel gennaio 2005 ha pubblicato il documento ufficiale "*Crisi d'impresa – Strumenti per l'individuazione di una procedura d'allerta"*, in cui vengono sintetizzati i principi e le linee guida che il Consiglio ritiene fondamentali per la gestione preventiva della crisi aziendale e per l'individuazione tempestiva dei segnali di allerta. Il documento nella terza sezione introduce l'analisi degli indicatori sottolineando che è necessario lavorare su una base di bilancio riclassificata al *fair value*[10], ossia valori correnti di mercato, in quanto fornisce una visione reale della situazione aziendale perché più coerente alla realtà economica.

La terza sezione del documento suddivide gli indici secondo tre equazioni di base che definiscono gli equilibri aziendali. Gli indici evidenziati derivano dalle componenti delle suddette equazioni e sono calcolati per individuare squilibri in queste componenti che conseguentemente generano squilibri nella situazione patrimoniale, finanziaria o economica:

- Equazione patrimoniale
    - *Indici di durata*, riguardano il monitoraggio della durata dell'esigibilità dei crediti e del pagamento dei debiti;
    - *Indici di solidità patrimoniale*, puntano alla verifica dell'equilibrio tra fonti di finanziamento e impieghi di tali fonti;

[9] CNDCEC – Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili, organismo italiano che recepisce e commenta i principi contabili e di revisione internazionali (ISA/IFRS) per l'applicazione nazionale, fornendo linee guida, interpretazioni e documenti di supporto agli iscritti.

[10] Secondo lo IASB (International Accounting Standards Board), il fair value è definito come "*il corrispettivo al quale un'attività può essere scambiata, o una passività estinta, tra parti consapevoli e disponibili, in un'operazione fra terzi indipendenti*".

  - *Indici di indebitamento*, servono per monitorare la composizione dei debiti e dei finanziamenti.
- Equazione monetaria
  - *Indici di liquidità finanziaria*, servono a comprendere la capacità dell'azienda di far fronte ai fabbisogni monetari tramite i mezzi propri.
- Equazione economica
  - *Indici di redditualità*, monitorano tutti i margini generati dall'attività;
  - *Indici finanziari*, viene verificata l'incidenza dei costi generati dal capitale di terzi.
- Indicatori extra-contabili, indicatori relativi a grandezze non economiche e contabili.

Per completezza vengono riportati per esteso tutti gli indici sopra accennati nell'Appendice B.

1.2 Il modello predittivo

Questa tesi si propone di sviluppare uno studio basato su un modello statistico di Machine Learning Bayesiano che permetta di calcolare la probabilità che un'impresa rispetti i criteri di continuità aziendale dati i suoi *trend* e valori di bilancio.

1.2.1 Approccio tradizionale e approccio predittivo

È importante introdurre la distinzione tra approccio statistico tradizionale e predittivo. Con approccio tradizionale si intende la modellazione esplicativa; quindi, l'insieme di metodologie statistiche volte alla comprensione dell'insieme di dati analizzati e all'analisi dei rapporti causali tra essi; invece, con approccio predittivo si intende la modellazione dei dati volti alla previsione futura dei fenomeni osservati.

I due approcci spesso vengono confusi alimentando il divario tra mondo accademico e pratico, contribuendo a creare barriere che limitano l'utilizzo di modelli statistici il cui impiego potrebbe agevolare determinati lavori predittivi nel mondo pratico. Per questo motivo è importante definirne la distinzione per un corretto uso statistico.

Secondo Galit Shmueli nel suo articolo "*To explain or to predict?*"[11] nella modellazione esplicativa la statistica viene utilizzata come strumento di supporto in quanto impiegata per la conferma di una teoria precedentemente elaborata, non viene infatti utilizzata per comprendere fattori di causalità all'interno di dati osservati, bensì la causalità viene assunta dalla teoria e il modello statistico misura se le relazioni tra i dati sono compatibili con l'ipotesi causale.

La "critica" alla base di questo articolo è che la modellazione esplicativa è l'approccio maggiormente utilizzato in vari campi di applicazione, trascurando invece le potenzialità di un approccio predittivo che, a differenza del precedente, viene usato per ipotizzare casi futuri o non osservabili e spesso grazie all'approccio predittivo vengono individuati nuovi pattern tra i dati che la teoria non aveva considerato.

[11] Galit Shmueli, *To Explain or to Predict?*, in *Statistical Science*, vol. 25, n. 3, 2010, pp. 289–310, Institute of Mathematical Statistics, DOI: 10.1214/10-STS330.

Un particolare importante da sottolineare secondo Shmueli è il fatto che la statistica descrittiva, che si occupa di rappresentare la struttura dei dati, non mira né a fare previsioni né ad individuare rapporti casuali, semplicemente si limita a descrivere come sono rappresentati i dati.

Questa distinzione non intende imporre la scelta di un approccio a discapito dell'altro, ma piuttosto sottolinearne le differenze e le potenzialità, così da favorirne un corretto utilizzo. L'approccio predittivo è particolarmente utile nell'attuale contesto dei *big data*, poiché risulta più efficace nel gestire grandi quantità di informazioni. Tuttavia, questi modelli possono configurarsi come delle vere e proprie *black box*: strumenti in grado di fornire risultati estremamente accurati, ma privi di trasparenza nei meccanismi che li hanno generati. Da qui l'importanza di trovare un equilibrio che consenta di sfruttare appieno le potenzialità predittive, senza rinunciare alla trasparenza e alle capacità esplicative.

1.2.2 Il contributo di Altman e Alberici

Per comprendere l'evoluzione degli studi sulla previsione del fallimento d'impresa è fondamentale richiamare i contributi pionieristici che hanno posto le basi della disciplina. Tra questi, i modelli proposti da Altman e successivamente da Alberici rappresentano un punto di riferimento essenziale per l'analisi della solidità aziendale e hanno influenzato in modo significativo lo sviluppo delle metodologie di valutazione del rischio di insolvenza.

Il primo studio in assoluto riconosciuto come pioniere nel campo è quello condotto da William H. Beaver[12] nel 1966. Al di la dei risultati ottenuti aveva tuttavia un grande limite, ovvero quello di utilizzare un approccio uni-variato, poiché analizza singolarmente ciascun indice di bilancio senza considerare l'effetto combinato delle variabili.

[12] W. H. Beaver, *Financial Ratios as Predictors of Failure*, *Journal of Accounting Research*, vol. 4, suppl. (1966), pp. 71–111.

Da qui Edward I. Altman[13] nel 1968 elaborò un'evoluzione del modello, proponendo un'analisi discriminate multivariata, in grado di combinare più indici di bilancio all'interno di un'unica funzione discriminante.

Lo studio si basa sul confronto tra due gruppi di imprese definiti da una variabile qualitativa dicotomica (aziende fallite e aziende sane), con l'obiettivo di individuare le variabili che, a livello di significatività statistica, mostrano le maggiori differenze tra i due gruppi. Dopo la fase di selezione delle variabili, Altman costruì una funzione di discriminazione che indica la forza del legame tra l'unità statistica e il gruppo di appartenenza.

La funzione di discriminazione restituisce dei valori detti z-score, che identificano di quante deviazioni standard un'osservazione si discosta dalla media di un gruppo. Sulla base di questi coefficienti discriminanti z-score ottenuti, Altman procede alla riclassificazione delle unità statistiche per valutare l'accuratezza del modello, ovvero la sua efficacia nella corretta classificazione.

I risultati ottenuti da Altman sono i seguenti:

- $Z > 2.99 \rightarrow$ zona sicura, basso o nullo rischio di insolvenza
- $1{,}81 < Z < 2{,}99 \rightarrow$ zona grigia, modello non attendibile
- $Z < 1{,}81 \rightarrow$ zona di pericolo, rischio alto di insolvenza

Lo studio di Altman si basò su un campione di 66 aziende, suddivise in due gruppi di 33 unità ciascuno: 33 aziende sane e 33 aziende fallite, le imprese selezionate avevano caratteristiche simili sia a livello di settore di appartenenza, settore manifatturiero statunitense, sia dimensionali, aziende quotate in borsa. Furono escluse le piccole e medie imprese a causa della scarsità di informazioni disponibili e per garantire una coerenza dimensionale con il gruppo selezionato di aziende fallite.

Altman analizzò i dati di bilancio relativi ai cinque anni precedenti il fallimento delle aziende oggetto di studio, evidenziando che l'accuratezza del modello risulta maggiore nei due anni antecedenti al fallimento.

[13] E. I. Altman, *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, *Journal of Finance*, vol. 23, n. 4 (1968), pp. 589–609.

Il modello dimostra un'elevata efficacia, classificando correttamente il 95% delle imprese. La funzione di discriminazione ottenuta è la seguente:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

dove:

- X_1 = Capitale circolante / totale attivo
- X_2 = Utili non distribuiti / totale attivo
- X_3 = Utili di gestione / totale attivo
- X_4 = Valore di mercato delle azioni / valore di libro del debito totale
- X_5 = Fatturato / totale attivo

Tuttavia, il modello presenta alcuni difetti, tra cui una ridotta affidabilità per i dati precedenti ai due anni dal fallimento. É stata inoltre identificata una zona grigia per le unità statistiche in cui il modello non è affidabile, con uno Z-score compreso tra 1.81 e 2.99**.** Inoltre, il modello è criticato per aver trascurato variabili qualitative come le finalità dei prestiti e delle garanzie offerte, o altre caratteristiche specifiche delle aziende che potrebbero fornire informazioni aggiuntive sulla situazione aziendale.

Da qui iniziarono diversi studi con l'applicazione di un modello di discriminazione multivariato che permetteva di ottenere una buona attendibilità dei risultati. A livello nazionale, in passato, i bilanci aziendali italiani non erano considerati affidabili come in altri Paesi esteri e gli studi svolti in materia avevano l'obbiettivo di confutare tali affermazioni.

A livello nazionale, uno studio di riferimento è il lavoro svolto da Adalberto Alberici[14] nel 1975, in quanto primo ad applicare modelli di discriminazione multivariata per la previsione del fallimento ad aziende manifatturiere italiane.

Nel suo studio analizzò delle difformità tra i dati ottenuti da un gruppo di aziende sane e quelli ottenuti da un gruppo di aziende a rischio di insolvenza e ne concluse che:

1. le aziende insolventi prestano una liquidità secca e generale peggiore di quelle sane

[14] A. Alberici, *Analisi dei bilanci e previsione delle insolvenze. Affidabilità bancaria e informativa del mercato mobiliare*, Milano, Isedi, 1975

2. Finanziano quote cospicue di immobilizzazioni mediante ricorso a forme di indebitamento a breve termine
3. Ricorrono maggiormente a forme di finanziamento esterne per coprire il loro fabbisogno finanziario

Nella prima fase del suo studio applica il *metodo dei quozienti* che prevede il calcolo di quozienti derivanti da rapporti tra voci di bilancio. Tali quozienti vengono successivamente confrontati e analizzati per comprendere quali siano i più indicativi a livello di previsione di insolvenza. Si tratta di una tecnica diagnostica e descrittiva che consente di analizzare la capacità delle imprese di individuare l'insolvenza con anticipo. È importante sottolineare che questo metodo non presenta un'elevata rilevanza previsionale rispetto al modello di Altman.

Alberici sviluppa, nella seconda parte dello studio, un modello di discriminazione multivariato attraverso il quale calcola cinque funzioni di discriminazione distinte per ciascuno dei cinque anni di dati raccolti:

$$Z_{-5} = -0{,}00401\,X_1 + 0{,}00203X_2 + 0{,}00346\,X_3 - 0{,}02201X_4 + 0{,}01374X_5 + 0{,}00108\,X_6 - 0{,}00417X_7 = 5{,}4945$$

$$Z_{-4} = 0{,}00164\,X_1 + 0{,}00350X_2 - 0{,}01659X_3 - 0{,}04353X_4 + 0{,}04026X_5 + 0{,}00013X_6 - 0{,}00105X_7 = 34{,}229$$

$$Z_{-3} = -0{,}00213\,X_1 + 0{,}00319X_2 + 0{,}00421\,X_3 - 0{,}02482X_4 + 0{,}01613X_5 + 0{,}00055\,X_6 + 0{,}00319X_7 = 120{,}222$$

$$Z_{-2} = 0{,}00004\,X_1 - 0{,}01528X_2 + 0{,}03013X_3 - 0{,}07389X_4 + 0{,}07658X_5 - 0{,}000446\,X_6 + 0{,}004828X_7 = 7.192{,}602$$

$$Z_{-1} = 0{,}00182\,X_1 - 0{,}02579X_2 + 0{,}00489\,X_3 - 0{,}05186X_4 + 0{,}00295X_5 + 0{,}03831\,X_6 + 0{,}01538X_7 = 92{,}708$$

Dove:

- X_1 = Mezzi propri / Capitale investito
- X_2 = Capitale circolante netto / Totale attivo
- X_3 = Mezzi propri / Debiti totali

- X_4 = Utile lordo / Vendite
- X_5 = Utile lordo / Capitale investito
- X_6 = Ricavi netti / Capitale investito
- X_7 = Disponibilità liquide / Debiti correnti

Sotto questo aspetto, si differenzia da Altman, il quale, invece elabora un'unica funzione che incorpora l'intero database relativo ai cinque anni di raccolta dati.

L'obiettivo dello studio di Alberici era confermare il fatto che anche nel contesto istituzionale italiano, il metodo dei quozienti e l'analisi discriminatoria multivariata consentivano di prevedere il fallimento di un'azienda con cinque anni di anticipo, commettendo un errore pari al 19% dei casi osservati.

In sintesi si ritiene che i contributi di Altman e Alberici siano di notevole importanza per l'evoluzione degli studi sulla previsione del fallimento aziendale in quanto, da un lato, Altman ha elaborato un modello di riferimento a livello internazionale introducendo la considerazione dell'influenza della relazione tra le variabili, dall'altra parte si ritiene significativo l'intervento di Alberici come ponte di congiunzione con l'economia italiana, comprovando il fatto che l'applicazione di tali modelli porta risultati attendibili anche nel contesto economico nazionale. Da questi studi emergono inoltre le misure che incidono in maniera più significativa sull'equilibrio aziendale e che potrebbero mettere a rischio i principi di continuità aziendale.

Le loro ricerche, come quelle di molti altri dopo di loro, hanno posto le basi per un filone di ricerca in cui la statistica si può affermare come strumento essenziale per la valutazione di fattori di rischio.

1.2.3 Il modello di Machine Learning Bayesiano

Sulla base delle osservazioni precedenti il modello oggetto di questo studio si baserà su due concetti principali che ne guideranno le assunzioni base, ovvero quello di *Machine Learning* e quello dei *Bayesian Networks*.

Si definisce come modello di machine learning un insieme di algoritmi che sulla base di un campione di dati forniti '*impara'* a riconoscere schemi e fare delle previsioni su insiemi di

nuovi dati, automatizzando e facilitando un processo decisionale caratterizzato da molteplici variabili.

La differenza principale rispetto ad altri modelli è che i modelli di machine learning sono progettati per trovare una funzione che predica y a partire da x, con l'obiettivo che questa funzione operi bene anche fuori dal campione di dati usato per addestrarla (out-of-sample). Nelle applicazioni economiche tradizionali, invece, l'obiettivo è stimare in modo accurato i parametri che descrivono la relazione tra x e y.

I modelli di machine learning sono quindi più adatti a compiti previsionali perché riescono a gestire relazioni non lineari e complesse e sono valutati in base alla loro capacità di generalizzare con nuovi dati, non solo sulla bontà dell'adattamento al campione di partenza.

Nell'articolo "*Machine Learning: an applied econometric approach*"[15] di S.Mullainathan e J.Spiess, viene spiegato come i modelli di machine learning vengono istruiti per funzionare nella previsione out-of-sample, il concetto viene spiegato con l'ausilio degli alberi decisionali. Nei modelli costruiti con alberi decisionali si parte da un insieme di dati che vengono suddivisi potenzialmente innumerevoli volte a seconda di varie caratteristiche che costringono una scelta. La suddivisione può avvenire fino a che ogni osservazione del campione iniziale trova la sua regola perfetta, questo però comporta un fenomeno di 'over fitting' ovvero il campione ha imparato a memoria i dati, seguendo un percorso ben preciso, di conseguenza, non sarà in grado di adattarsi in situazioni esterne con dati nuovi.

Sulla base di queste osservazioni Mullainathan e Spiess nel loro articolo evidenziano l'importanza di regolarizzare, quindi porre dei limiti di profondità e complessità al modello per ridurre i *rumors*. Per comprendere e analizzare il livello di accuratezza e generalizzazione del modello con dati al di fuori del campione si procede con una 'simulazione' di nuovi dati, ovvero si suddivide il campione di partenza in più sottocampioni, si crea il modello sulla base di alcuni di questi sottocampioni, escludendo i restanti che serviranno per il test out-of-sample. Quindi si procede con il test sul sottocampione escluso inizialmente procedendo poi con il medesimo

[15] Sendhil Mullainathan e Jann Spiess, *"Machine Learning: An Applied Econometric Approach"*, Journal of Economic Perspectives 31, n. 2 (Spring 2017): 87–106.

test finché non vengono utilizzati tutti i sottocampioni; questo metodo è detto *cross-validation* e serve per capire come il modello si comporta con dati che non sono stati utilizzati per addestrarlo, in modo da capire se sarà efficiente con dati nuovi.

Il modello di machine learning adottato si basa su un approccio Bayesiano, ovvero su un modello probabilistico che data una probabilità pregressa, sulla base di nuovi dati forniti, aggiorna la probabilità a posteriori guidando una scelta con informazioni aggiornate.
Questo tipo di modelli si basa sulla teoria di Bayes che nella sua formulazione elementare permette di calcolare la probabilità condizionata di un evento dato un altro[16]; nei modelli più complessi l'approccio si estende a più variabili interdipendenti come avviene nelle reti Bayesiane.

Questo tipo di modelli lavora in maniera più efficiente con variabili discrete, ovvero variabili quantitative o qualitative che assumono valori finiti e numerabili e di cui non esistono valori intermedi tra due possibili stadi della variabile, a differenza delle variabili continue che possono assumere qualsiasi valore in un intervallo definito. Questo perché i modelli Bayesiani sono pensati per lavorare con variabili categoriche per una gestione più semplice del calcolo. Tuttavia, l'approccio Bayesiano è generale e consente di modellare sia variabili discrete sia continue, scegliendo di volta in volta la distribuzione di probabilità più appropriata al tipo di dato.

Nell'articolo "*Using Bayesian Networks for Bankruptcy Prediction: some methodological issues*" di L.Sun e P.Shenoy[17] viene evidenziato il problema riportando come soluzione un metodo di discretizzazione per convertire le variabili continue oggetto dell'analisi, principalmente indicatori contabili come l'indice di redditività o di liquidità, in variabili discrete. Nello studio sopracitato sono state suddivise le variabili discrete in stati ed è stato

[16]
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$
- $P(A)$ e $P(B)$: probabilità dell'evento A e dell'evento B separatamente
- $P(A|B)$: probabilità aggiornata di A sulla base delle nuove evidenze di B
- $P(B|A)$: probabilità che avvenga l'evento B dato l'evento A

[17] Lili Sun e Prakash P. Shenoy, "Using Bayesian Networks for Bankruptcy Prediction: Some Methodological Issues," *European Journal of Operational Research* 180, no. 2 (2007): 738–753

analizzato che il numero migliore fosse due o tre, per evitare una problematica di over-fitting. Tutte le osservazioni sono quindi state suddivise secondo un livello che può essere ad esempio "basso", "medio" o "alto" per l'indice di redditività, così facendo la variabile continua viene trasformata in categoriale e processata in maniera più efficiente dal modello.

Uno dei valori aggiunti più significativi dei modelli Bayesiani è la possibilità di modellare le relazioni di dipendenza condizionale tra le variabili poiché questo tipo di modelli tiene in considerazione, oltre all'influenza della singola variabile sul risultato, anche delle influenze reciproche che le variabili hanno tra loro.

Nell'articolo "*Toward a better expert system for auditor going concern opinions using Bayesian network inflation factors*"[18] viene evidenziato come recenti studi abbiano appurato che i modelli Bayesiani ottengono risultati migliori sulle scelte previsionali rispetto a modelli di regressione classici.

Inoltre, l'approccio Bayesiano permette di incorporare nei modelli decisionali fattori osservabili e non osservabili. Questa caratteristica è molto importante in questo studio in quanto ci possono essere dei fattori che influenzano la continuità aziendale di un'azienda che non emergono esplicitamente nei bilanci, come ad esempio alcune scelte manageriali oppure determiniate caratteristiche della rete relazionale esterna.

In conclusione, da recenti studi si può osservare che il modello che meglio si può adattare ad uno studio di tipo probabilistico sul fallimento aziendale sia appunto un modello Bayesiano; nel corso di questa tesi verrà impiegato questo tipo di modello in applicazione ad uno studio che vuole analizzare la probabilità che un'azienda rispetti i principi di continuità aziendale e che sia trovi o meno nella condizione di dover attuare una misura di risanamento prevista dalla normativa.

[18] Vikram Desai, Anthony C. Bucaro, Joung W. Kim, Rajendra Srivastava e Renu Desai, *Toward a better expert system for auditor going concern opinions using Bayesian network inflation factors*, International Journal of Accounting Information Systems, Vol. 49, 2023, 100617.

2 Le reti Bayesiane

In questa tesi si intende sviluppare un'analisi dei dati finalizzata al supporto dei processi decisionali mediante metodologie statistiche. In scenari simulati tali decisioni possono essere assunte sulla base di informazioni complete e perfettamente osservabili; tuttavia, nelle applicazioni reali, le informazioni disponibili risultano spesso incomplete e caratterizzate da un elevato grado di incertezza. Diventa pertanto necessario adottare metodologie in grado di mantenere un adeguato equilibrio tra il rigore statistico e la rappresentazione dell'incertezza dei fenomeni osservati.

L'approccio Bayesiano interpreta la probabilità come misura del grado di credenza razionale di un decisore e fornisce un quadro formale per l'aggiornamento delle credenze alla luce di nuove evidenze. Tuttavia, in contesti reali caratterizzati da un elevato numero di variabili, la rappresentazione e l'utilizzo diretto delle distribuzioni di probabilità congiunte risulta complesso dal punto di vista computazionale.

Per ovviare a tali limitazioni, si ricorre alle reti Bayesiane, modelli grafici che consentono di rappresentare in modo compatto le relazioni di dipendenza e indipendenza tra variabili. Attraverso la combinazione di grafi diretti aciclici (*Directed Acyclic Graphs*, DAG) e dell'insieme di distribuzioni di probabilità condizionate, le reti Bayesiane permettono di rendere l'inferenza probabilistica più interpretabile e più efficiente dal punto di vista computazionale.

In questo capitolo viene introdotto formalmente il modello delle reti Bayesiane da un punto di vista metodologico, descrivendone le assunzioni di indipendenza e i principi di inferenza probabilistica. L'obiettivo è fornire il quadro teorico necessario per comprendere il loro utilizzo nei capitoli successivi.

2.1 Teoria della probabilità

Per formalizzare a livello quantitativo l'incertezza associata ai fenomeni analizzati, si utilizza la teoria della probabilità, in quanto essa consente di esprimere in maniera coerente il grado di incertezza relativo al verificarsi di determinati eventi. In particolare, la probabilità fornisce uno strumento matematico per modellare situazioni in cui non è possibile prevedere con certezza l'esito di un fenomeno, ma è comunque possibile valutarne la plausibilità.

I concetti fondamentali della teoria della probabilità costituiscono la base teorica dell'approccio Bayesiano e risultano indispensabili per la costruzione e l'interpretazione delle reti Bayesiane, nelle quali l'incertezza viene modellata attraverso distribuzioni di probabilità e probabilità condizionate.

Definizione 2.1: (Prova). La prova (o esperimento aleatorio) è un esperimento che ha due o più possibili risultati e in cui c'è un certo grado di incertezza su quale dei risultati si verificherà.

Definizione 2.2: (Evento). Gli eventi sono i possibili risultati della prova, definiti dal simbolo ω.

Definizione 2.3: (Spazio campionario). L'insieme di tutti i possibili eventi ω *viene chiamato spazio campionario ed è indicato con il simbolo* Ω.

Definizione 2.4: (Probabilità). La probabilità è una funzione di insieme che associa ad ogni evento $A \in \Omega$ *un numero reale dell'intervallo [0,1]. La probabilità indicata con P(A).*

Il calcolo delle probabilità deve rispettare alcuni assiomi fondamentali, dai quali derivano le principali proprietà utilizzate nell'analisi dei fenomeni aleatori. Questi concetti costituiscono la base teorica su cui si fonda l'approccio Bayesiano e risultano indispensabili per la costruzione e l'interpretazione delle reti Bayesiane, nelle quali l'incertezza viene modellata attraverso distribuzioni di probabilità e probabilità condizionate.

Con il termine assioma si indica un enunciato fondamentale assunto come vero senza necessità di dimostrazione, dal quale è possibile derivare, tramite deduzione logica, ulteriori proprietà e risultati. Nel caso specifico, gli assiomi del calcolo delle probabilità sono i seguenti.

Assioma 2.1: La probabilità associata all'intero spazio campionario, che contiene tutti i possibili eventi, è pari alla probabilità massima ed è uguale a 1:

$$P(\Omega) = 1$$

Assioma 2.2: La probabilità di un qualsiasi evento non può assumere valori negativi:

$$\forall A \subseteq \Omega, P(A) \geq 0$$

Assioma 2.3: dati due eventi, A e B, tra loro esclusivi, la probabilità della loro unione è data dalla somma delle rispettive probabilità.

$$\forall A, B \subseteq U, \text{se } A \cap B = \emptyset, \qquad \text{allora } P(A \cup B) = P(A) + P(B)$$

Qualsiasi funzione che soddisfi i suddetti assiomi può essere una funzione di probabilità.

Poiché nella teoria della probabilità gli eventi sono formalmente definiti come insiemi di uno spazio campionario, le relazioni tra eventi possono essere descritte attraverso le operazioni dell'algebra degli insiemi. In combinazione con gli assiomi di probabilità, tali operazioni permettono di derivare alcune proprietà fondamentali del calcolo delle probabilità, tra cui:

- (*Insieme vuoto*) La probabilità associata ad un insieme vuoto, privo di elementi, è nulla
$$P(\emptyset) = 0$$
- (*Evento complementare*) La probabilità di un evento complementare all'evento A, indicato con $\bar{A}$, è data da:
$$(A=) = 1 - P(A)$$
- (*Monotonicità*) Dati due eventi A e B, tali che $B \subseteq A$, allora la probabilità dell'evento B sarà sempre inferiore alla probabilità dell'evento A:
$$\forall A, B \; dove \quad B \subseteq A, allora \, P(B) \leq P(A)$$
- (*Unione di due eventi*) Dati due eventi A e B appartenenti allo spazio campionato, l'unione delle probabilità dei due eventi viene calcolato come:
$$\forall A, B \subseteq \Omega, P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Le ipotesi appena descritte individuano particolari situazioni in cui lo spazio campionario è completamente noto e tutti gli eventi sono equiprobabili. In tali casi, la probabilità può essere calcolata attraverso il cosiddetto approccio classico.

Definizione 2.5 (Probabilità classica). Dato uno spazio campionario noto e data la probabilità degli eventi dello stesso spazio come tutti ugualmente probabili, la probabilità di un evento A può essere calcolata come il rapporto tra i casi favorevoli dell'evento e il numero di casi possibili.

$$P(A) = \frac{numero\ dei\ casi\ favorevoli}{numero\ dei\ casi\ possibili}$$

In contesti reali è frequente disporre di informazioni parziali che modificano le probabilità iniziali associate agli eventi in relazione al verificarsi di altri eventi. La teoria della probabilità introduce quindi il concetto di probabilità condizionata, che consente di calcolare la probabilità di un evento alla luce del verificarsi di un altro evento.

Definizione 2.6 (Probabilità condizionata). Siano A e B due eventi tali per cui P(B) ≠ Ø. La probabilità condizionata di A dato B, indicata come P(A|B), è data da:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

L'interpretazione alla base risiede nel fatto che la probabilità condizionata può essere interpretata come la probabilità dell'evento A calcolato su uno spazio campionario ridotto, costituito dall'intersezione tra l'insieme A e l'insieme B. Poiché si assume che l'evento B si sia verificato, l'insieme degli eventi possibili si restringe a quelli compatibili con B. Il numeratore rappresenta quindi la probabilità degli eventi in cui si verificano simultaneamente A e B, mentre il denominatore normalizza tale valore rispetto alla probabilità dell'evento che condiziona, ovvero la probabilità dell'evento B.

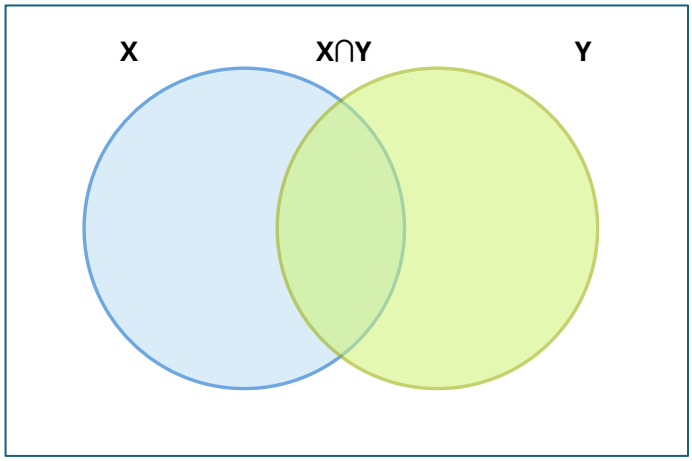


Figura 2.1: Rappresentazione grafica dell'intersezione tra insiemi. Fonte: elaborazione grafica personale

La probabilità condizionata consente dunque di formalizzare il concetto di aggiornamento dell'informazione, permettendo di ricalcolare le probabilità alla luce di nuove evidenze. A partire da tale definizione si ottiene il cosiddetto principio delle probabilità composte, espresso di seguito.

Definizione 2.7 (Principio delle probabilità composte). Dati due eventi A e B tali per cui P(A) >0 e P(B) > 0:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Secondo questo principio, la probabilità congiunta di due eventi può essere espressa come il prodotto tra la probabilità di un evento e la probabilità condizionata dell'altro. Tale relazione evidenzia come la probabilità di eventi complessi possa essere scomposta in termini più semplici, combinando informazioni a priori e probabilità condizionate.

Il concetto di probabilità condizionata permette inoltre di introdurre formalmente la nozione di indipendenza tra eventi. Due eventi si dicono indipendenti quando il verificarsi di uno non modifica la probabilità dell'altro; in tal caso, non sussiste alcuna relazione di condizionamento tra gli eventi.

Definizione 2.8 (Indipendenza). Due eventi A e B si dicono indipendenti se e solo se $A \perp B$, quindi:

$$P(A|B) = P(A)$$

Al contrario, se due eventi sono dipendenti, il verificarsi di uno, influenza la probabilità dell'altro e risulta quindi necessario introdurre un meccanismo di aggiornamento delle probabilità.

Il teorema di Bayes fornisce una regola formale per aggiornare la probabilità di un'ipotesi alla luce di una nuova evidenza. Formalmente:

Definizione 2.9 (Teorema di Bayes). Dato un insieme esclusivo ed esaustivo di k eventi B_1, B_2, B_3, … B_k ed un evento A, si definisce:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

$$= \frac{P(A|B_i)P(B_i)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) + \cdots + P(B_k)P(A|B_k)}$$

- *Un insieme di eventi si dice esclusivo se gli eventi in questione sono incompatibili:*

$$B_i \cap B_j = \emptyset \; per \; ogni \; i \neq j$$

- *Un insieme di eventi si dice esaustivo se l'unione di tali eventi coincide con lo spazio campionario:*

$$\bigcup_{i=1}^{k} B_i = \Omega$$

Questi elementi rappresentano la base del ragionamento probabilistico in situazioni di incertezza che si manifestano nei contesti reali, nei quali le informazioni disponibili sono spesso incomplete. Tuttavia, in presenza di un numero elevato di variabili, l'applicazione diretta di tali relazioni risulta frequentemente complessa sia dal punto di vista computazionale sia da quello interpretativo. Le reti Bayesiane rispondono a questa esigenza fornendo un quadro strutturato che consente di applicare in modo sistematico i principi basati sul teorema di Bayes, rendendo possibile l'analisi e l'inferenza in sistemi complessi.

2.2 Introduzione alle reti Bayesiane

Dopo aver introdotto i principali concetti della teoria della probabilità, è ora possibile presentare le reti Bayesiane.

Le reti Bayesiane sono modelli grafici probabilistici che combinano i principi della teoria della probabilità con quelli della teoria dei grafi. Esse consentono di rappresentare in modo compatto le relazioni di dipendenza e indipendenza tra un insieme di variabili casuali, fornendo al contempo uno strumento efficace per il ragionamento probabilistico sotto incertezza. Grazie a questa doppia natura, le reti Bayesiane risultano particolarmente adatte all'analisi di fenomeni reali, nei quali le informazioni sono incomplete o incerte.

Gli elementi principali che permettono la costruzione delle reti Bayesiane sono:

- *Nodi*: presentano la variabile casuale, o una quantità incerta;
- *Archi diretti*: indicano l'esistenza di dipendenze dirette tra le variabili collegate, la cui forza è quantificata da probabilità condizionali;
- *Assenza di archi*: indica l'indipendenza tra due variabili.

Formalmente una rete Bayesiana è definita come:

Definizione 2.10 (Rete Bayesiana). Si consideri un insieme con n variabili casuali $X_1, X_2, X_3, \dots X_n$, dove ogni variabile X è associata a un nodo di un grafo aciclico diretto (DAG). Il grafo è una rete Bayesiana se:

$$P(X_1, X_2, X_3, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid P_a\ (X_i))$$

Dove $P_a\ (X_i)$ definisce l'insieme dei genitori del nodo X_i, ovvero l'insieme di tutte le variabili per le quali esiste un nodo diretto verso la variabile X_i.

Una rete Bayesiana è generalmente definita come BN = (G, P), dove G rappresenta la struttura del grafo aciclico diretto e P la distribuzione delle probabilità congiunte dell'insieme delle variabili.

2.2.1 La struttura – componente qualitativa

Vengono ora introdotti in modo più approfondito gli elementi che compongono una rete Bayesiana, a partire dalla sua struttura grafica, che rappresenta la componente qualitativa del modello.
Tale struttura è espressa mediante un grafo aciclico diretto (Directed Acyclic Graph, DAG) e ha il compito di descrivere le relazioni di dipendenza e indipendenza condizionata tra le variabili casuali del sistema.

Definizione 2.11 (Grafo). Un grafo, indicato come G (N, A) è un insieme di nodi N, connessi tra loro da un insieme di archi A, che possono essere o meno direzionati.

Nello specifico nelle reti Bayesiane le strutture grafiche devono soddisfare una condizione fondamentale definita come segue.

Definizione 2.12 (Grafo aciclico diretto, DAG). Un grafo aciclico diretto è un grafo orientato, nel quale gli archi presentano una direzione specifica di influenza, e nel quale non esistono cicli diretti. In particolare, non è possibile partire da un nodo e ritornare allo stesso seguendo un percorso coerente con la direzione degli archi.

L'assenza di cicli rappresenta un vincolo essenziale per le reti Bayesiane, poiché consente di definire una fattorizzazione ben definita della distribuzione di probabilità congiunta. Dal punto di vista concettuale, tale vincolo impedisce la presenza di dipendenze circolari tra le variabili, che renderebbero ambigua l'interpretazione probabilistica del modello.

In una rete Bayesiana, ciascun nodo del DAG rappresenta una variabile casuale, mentre ciascun arco diretto esprime una dipendenza probabilistica diretta tra le variabili collegate. In particolare, la presenza di un arco diretto dal nodo X_j al nodo X_i indica che la variabile X_j ha un'influenza diretta sulla variabile X_i, nel senso che la distribuzione di probabilità di X_i dipende dal valore assunto da X_j. Tale relazione non deve essere interpretata come una relazione causale, ma come una dipendenza rilevante dal punto di vista probabilistico.

L'assenza di un arco diretto tra due nodi, invece, rappresenta un'ipotesi di indipendenza condizionata tra le corrispondenti variabili, dato l'insieme delle variabili che influenzano direttamente ciascun nodo. In questo modo, la struttura del grafo consente di codificare in

maniera esplicita le ipotesi di indipendenza del modello, permettendo una rappresentazione compatta ed efficiente della distribuzione di probabilità congiunta.

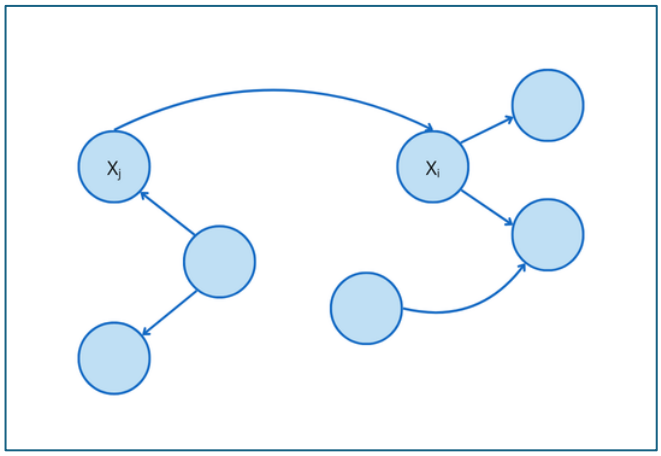


Figura 2.2: Rappresentazione di un grafo aciclico diretto semplificato. Fonte: elaborazione grafica personale

All'interno del DAG, si definiscono alcune relazioni tra cui:

Definizione 2.13 (Nodo genitore). Si definisce nodo genitore (parent) ogni nodo dal quale parte un arco diretto per raggiungere un altro nodo. L'insieme dei nodi genitori di un dato nodo X_i, come $P_a(X_i)$.

Definizione 2.14 (Nodo figlio, discendente). Si definisce nodo figlio (child) ogni nodo che può essere raggiunto direttamente da ciascuno dei suoi genitori. Sono inoltre definiti discendenti (descedants) tutti i nodi che possono essere raggiunti a partire da un dato nodo seguendo un percorso diretto nel grafo.

Un nodo privo di nodi genitori è definito *nodo radice (root)*, mentre un nodo privo di nodi figli è definito *nodo foglia (leaf)*.

Queste relazioni strutturali sono di fondamentale importanza nelle reti Bayesiane, poiché determinano l'insieme delle variabili rispetto alle quali ciascun nodo deve essere condizionato. In particolare, esse consentono di specificare localmente le distribuzioni di probabilità condizionata associate ai nodi della rete e giustificano la fattorizzazione della distribuzione di probabilità congiunta del sistema, che costituisce uno dei principali vantaggi delle reti Bayesiane nella modellazione di sistemi complessi.

2.2.2 I parametri – componente quantitativa

Affinché una rete Bayesiana possa essere correttamente utilizzata per svolgere inferenza probabilistica, alla componente qualitativa, rappresentata dalla struttura del grafo, deve essere affiancata una componente quantitativa, costituita dall'insieme delle tabelle di probabilità condizionata (*Conditional Probabilitity Tables,* CPT).

Definizione 2.15 (Tabella di probabilità condizionata). Per ogni nodo X_i di una rete Bayesiana, avente come genitori le n variabili Y_1, Y_2, … Y_n, la tabella di probabilità condizionata (Conditional Probabilitity Tables, CPT), indicata con $P(X_i \mid Y_1, Y_2, \ldots Y_n)$, contiene la probabilità associata a ciascuna possibile combinazione tra gli stati assunti da X_i e da tutti i suoi nodi genitori.

Le tabelle di probabilità condizionata hanno quindi il compito di quantificare l'influenza probabilistica esercitata dai nodi genitori su ciascun nodo figlio, traducendo in termini numerici le relazioni di dipendenza individuate dalla struttura del grafo.

Nel caso dei nodi radice, ovvero nodi privi di genitori, la CPT coincide con la distribuzione di probabilità a priori della variabile. Essa rappresenta una probabilità marginale, in quanto descrive la distribuzione degli stati della variabile in assenza di qualunque informazione o condizionamento derivante dalle altre variabili della rete.

Per i nodi interni o foglia, che possiedono uno o più genitori, la CPT specifica invece tutte le probabilità condizionate della variabile rispetto a ciascuna possibile configurazione degli stati dei genitori, garantendo che, per ogni configurazione, la somma delle probabilità associate agli stati del nodo sia pari a uno.

L'insieme di tutte le tabelle di probabilità condizionata associate a tutti i nodi della rete, consente di definire la distribuzione di probabilità congiunta della rete.

Definizione 2.16 (Distribuzione di probabilità congiunta). Data una rete Bayesiana definita sull'insieme delle variabili X= (X_1, X_2, … X_n), la distribuzione di probabilità congiunta può essere definita come prodotto delle distribuzioni locali associate a ciascun nodo:

$$P(X_1, X_2, .., X_n) = \prod_{i=1}^{n} P(X_i \mid P_a\ (X_i))$$

Dove $P_a(X_i)$ indica l'insieme dei nodi genitori del nodo X_i.

Questa fattorizzazione rappresenta uno degli aspetti fondamentali delle reti Bayesiane, poiché permette di descrivere sistemi complessi mediante un insieme di distribuzioni di probabilità locali, riducendo significativamente la complessità rispetto alla rappresentazione esplicita della distribuzione congiunta.

In conclusione, la struttura della rete Bayesiana ha il compito di individuare le relazioni di dipendenza e di indipendenza condizionale tra le variabili del sistema, codificando in modo esplicito quali variabili influenzano direttamente le altre. Tuttavia, la sola struttura non è sufficiente a descrivere il comportamento probabilistico del fenomeno analizzato, poiché essa non fornisce alcuna informazione sull'intensità o sulla natura quantitativa di tali relazioni.

Questo ruolo è svolto dalle tabelle di probabilità condizionata, associando a ciascun nodo una distribuzione di probabilità locale condizionata ai suoi genitori. L'insieme delle CPT consente così di ricostruire la distribuzione di probabilità congiunta dell'intero sistema attraverso la fattorizzazione Bayesiana, rendendo il modello efficiente a livello computazionale anche in presenza di un elevato numero di variabili.

Questa combinazione rende la rete Bayesiana uno strumento efficace per l'inferenza probabilistica, in grado di aggiornare i propri dati sulla base di nuove evidenze osservate.

2.3 Dipendenza e indipendenza nelle reti Bayesiane

In questo paragrafo si intende fare chiarezza su un concetto fondamentale delle reti Bayesiane, ovvero il concetto di dipendenza e indipendenza tra le variabili. Nelle reti Bayesiane, infatti, la presenza o l'assenza di archi tra i nodi non ha un significato puramente descrittivo, ma esprime assunzioni precise di indipendenza condizionata tra le variabili del modello.

In ambito probabilistico, il concetto di indipendenza assume particolare rilevanza nella sua forma condizionata. In particolare, due variabili casuali X e Y si dicono indipendenti condizionatamente a una terza variabile Z se, una volta nota l'informazione fornita da Z, la conoscenza di X non apporta alcuna informazione aggiuntiva su Y, e viceversa.

Definizione 2.17 (Indipendenza condizionata). Due variabili X e Y sono indipendenti condizionalmente a Z, se

$$P(X, Y \mid Z) = P(X \mid Z)\, P(Y \mid Z)$$

Il legame tra la struttura grafica di una rete Bayesiana e le relazioni di indipendenza probabilistica è formalizzato dalla proprietà di Markov, che rappresenta uno dei fondamenti teorici del modello.

Definizione 2.18 (Proprietà di Markov). Data una rete Bayesiana definita su un insieme di variabili casuali $X = (X_1, X_2, \ldots, X_n)$, ogni variabile X_i è condizionatamente indipendente da tutti i suoi non discendenti dato l'insieme dei suoi genitori. Formalmente:

$$X_i \perp NonDisc(X_i) \mid P_a(X_i)$$

Dove $NonDisc(X_i)$ indica l'insieme dei nodi non discendenti di X_i, e $P_a(X_i)$ indica l'insieme dei nodi genitori di X_i.

Questa proprietà implica che, una volta note le variabili che influenzano direttamente X_i, nessun'altra variabile del sistema fornisce informazioni aggiuntive rilevanti su X_i. In altre parole, in una rete Bayesiana ogni variabile è indipendente da tutte le variabili che non ne sono conseguenza diretta o indiretta, a condizione che siano noti i valori delle sue cause immediate.

All'interno delle reti Bayesiane, le possibili relazioni tra le variabili sono definite da specifiche configurazioni elementari dei nodi, tali strutture consentono alla rete Bayesiana di codificare il

flusso di informazione tra le variabili e di determinare le relazioni di dipendenza e indipendenza condizionata implicite nel modello.

In particolare, esistono tre configurazioni fondamentali, ciascuna delle quali identifica un pattern distinto ed è associata a specifiche proprietà di indipendenza condizionata: le connessioni seriali, le connessioni divergenti e le connessioni convergenti.

1. *Connessioni seriali (catena).* Il primo pattern è rappresentato dalla connessione seriale, in cui una variabile X influenza una variabile intermedia Y, che a sua volta influenza una variabile Z:

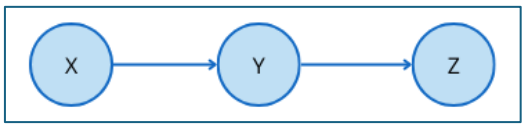


Figura 2.3: Rappresentazione grafica di una catena causale. Fonte: elaborazione grafica personale

In questa configurazione, l'influenza di X su Z è interamente mediata dalla variabile Y. Di conseguenza, una volta osservato lo stato di Y, la variabile X non fornisce ulteriori informazioni rilevanti su Z. Formalmente, vale la seguente relazione di indipendenza condizionata:

$$P(Z \mid X \wedge Y) = P(Z \mid Y) \equiv X \perp Z \mid Y$$

Ciò significa che conoscere il valore assunto da Y rende irrilevante la conoscenza di X ai fini della determinazione della probabilità di Z. Questo pattern è tipico dei processi causali sequenziali, in cui l'effetto si propaga lungo una catena di variabili intermedie.

2. *Connessioni divergenti (causa comune).* Il secondo pattern è la connessione divergente, in cui una variabile Y agisce come causa comune di due variabili X e Z:

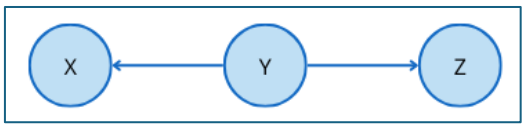


Figura 2.4: Rappresentazione grafica di connessioni divergenti. Fonte: elaborazione grafica personale

In assenza di informazioni su Y, le variabili X e Z risultano in generale dipendenti, poiché condividono una causa comune. Tuttavia, una volta osservato lo stato di Y, tale dipendenza viene eliminata: la conoscenza di X non apporta informazioni aggiuntive su

Z, e viceversa. Anche in questo caso si ottiene una relazione di indipendenza condizionata:

$$P(Z \mid X \wedge Y) = P(Z \mid Y) \equiv Z \perp Z \mid Y$$

Questo pattern è tipico di situazioni in cui una stessa variabile influenza più effetti, e l'osservazione della causa comune è sufficiente a spiegare le variazioni dei singoli effetti.

3. *Connessioni convergenti (effetto comune o collider).* Il terzo pattern fondamentale è la connessione convergente, detta anche struttura a *effetto comune* o *collider*, in cui due variabili X e Z influenzano una stessa variabile Y:

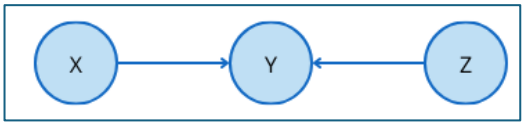


Figura 2.5: Rappresentazione grafica di connessioni divergenti. Fonte: elaborazione grafica personale

A differenza dei casi precedenti, in assenza di osservazioni su Y, le variabili X e Z risultano marginalmente indipendenti, poiché non esiste un percorso che consenta il passaggio di informazione tra di esse.

Tuttavia, quando la variabile Y (o uno dei suoi discendenti) viene osservata, si instaura una dipendenza tra X e Z. Questo fenomeno è noto come *explaining away*: la conoscenza dello stato di una delle due cause modifica la probabilità dell'altra causa, dato l'effetto osservato.

$$P(Z \mid X \wedge Y) \neq P(X \mid Y) \equiv X \not\perp Z \mid Y$$

Questo pattern rappresenta uno degli aspetti più controintuitivi delle reti Bayesiane ed è fondamentale per comprendere il comportamento del modello durante il processo di inferenza.

I criteri introdotti nel paragrafo precedente consentono di identificare le relazioni di dipendenza e indipendenza in contesti semplificati, caratterizzati da un numero limitato di variabili e connessioni. Tuttavia, nelle applicazioni reali, in cui le reti Bayesiane presentano numerose variabili e relazioni complesse, risulta necessario disporre di un criterio generale che permetta

di determinare se due variabili siano indipendenti, dato un insieme arbitrario di evidenze. Tale criterio è noto come *d-separation* (*directional separation*).

La *d-separation* è una proprietà puramente grafica che consente di stabilire le relazioni di indipendenza condizionata tra le variabili di una rete Bayesiana analizzando esclusivamente la struttura del grafo, senza fare esplicito riferimento ai valori delle probabilità. In particolare, essa definisce quando un percorso tra due nodi è in grado di trasmettere informazione e quando, invece, risulta bloccato in presenza dell'osservazione di determinate variabili.

Definizione 2.19 (D-Separation). Siano X e Z due nodi di una rete Bayesiana e sia E un ulteriore insieme di nodi osservati. Un cammino tra X e Z si dice bloccato dato E se lungo il cammino esiste almeno un nodo Y che soddisfa una delle seguenti condizioni:

- *Y è contenuto nell'insieme delle evidenze E, ed è in una connessione seriale o divergente lungo il cammino.*
- *Y ne nessuno dei suoi discendenti appartengono all'insieme delle evidenze E, ed è in una connessione convergente (collider) nel suo cammino.*

Se tutti i cammini tra X e Z risultano bloccati dato E, allora i due nodi sono detti d-separati e risultano condizionatamente indipendenti dato E, formalmente:

$$X \perp Z \mid E$$

Al contrario, se esiste almeno un cammino non bloccato tra X e Z, i nodi sono d-connessi e quindi probabilisticamente dipendenti.

$$X \not\perp Z \mid E$$

Il criterio di *d-separation* consente quindi di determinare, a partire dalla sola struttura del grafo, quali relazioni di indipendenza condizionata siano valide tra le variabili del modello. Tale proprietà rende le reti Bayesiane uno strumento particolarmente potente, poiché permette di analizzare sistemi complessi senza fare riferimento esplicito ai valori numerici delle probabilità.

In particolare, la *d-separation* fornisce la base teorica per individuare sottoinsiemi rilevanti di variabili ai fini dell'inferenza, riducendo la complessità computazionale e migliorando l'interpretabilità del modello.

Definizione 2.20 (Markov Blanket). Sia G= (V, A) una rete Bayesiana definita su un insieme di variabili casuali V e sia X ∈ V un nodo della rete. Il Markov Blanket di X, indicato con MB(X), è il sottoinsieme minimo di nodi S ⊆ V | {X} tale per cui:

$$X \perp V \setminus (\{X\} \cup S) \mid S$$

X è indipendente da tutte le variabili dell'insieme escluso il nodo X unito a S, ovvero l'insieme di variabili facenti parte del Markov Blanket di X.

Il Markov Blanket di un nodo X è composto dai genitori di X, i figli di X e i co-genitori dei figli di X.

Queste variabili rappresentano l'unico sottoinsieme della rete in grado di influenzare direttamente o indirettamente la distribuzione di probabilità di X. Di conseguenza, una volta osservati i nodi appartenenti al Markov Blanket, tutte le altre variabili della rete risultano irrilevanti ai fini dell'inferenza su X.

Il concetto di Markov Blanket riveste un ruolo centrale nell'inferenza probabilistica, poiché individua l'insieme minimo di variabili necessarie per descrivere completamente il comportamento probabilistico di un nodo di interesse. Una volta fissata l'evidenza sul Markov Blanket, tutte le altre variabili della rete risultano irrilevanti ai fini dell'inferenza su quel nodo.

Nel contesto applicativo di questa tesi, il Markov Blanket verrà utilizzato per focalizzare l'analisi sui principali fattori di influenza delle variabili chiave, consentendo di interpretare i risultati della rete Bayesiana in modo più chiaro ed efficace dal punto di vista decisionale.

Definizione 2.21(Independency Map). Un grafo è detto independency map (I-Map) se tuttte le relazioni di indipendenza condizionata deducibili tramite il criterio di d-separation sono valide anche nella distribuzione di probabilità associata, definita sull'insieme di variabili casuali.

Nel contesto delle reti Bayesiane, l'ipotesi che il grafo costituisca una independency-map (I-map) risulta particolarmente rilevante dal punto di vista applicativo. Essa implica che le relazioni di indipendenza condizionata deducibili dalla struttura grafica siano coerenti con quelle presenti nella distribuzione di probabilità del sistema. In altri termini, l'assenza di un arco tra due nodi rappresenta un'assunzione esplicita di indipendenza condizionata, che semplifica sia l'interpretazione del modello sia il processo di inferenza.

Questa proprietà è fondamentale, poiché garantisce la coerenza tra la rappresentazione grafica del modello e le relazioni probabilistiche sottostanti. Inoltre, essa costituisce il presupposto teorico che consente di utilizzare la struttura della rete per guidare in modo efficiente il processo di inferenza probabilistica.

Nel seguito della tesi si assumerà che la rete Bayesiana costruita soddisfi tale proprietà, consentendo di utilizzare la struttura del grafo come strumento affidabile per l'analisi delle relazioni tra le variabili e per l'identificazione dei fattori rilevanti nel processo decisionale.

2.4 Inferenza nelle reti Bayesiane

Dopo aver definito la struttura della rete (DAG) e aver quantificato le relazioni tra variabili tramite le tabelle di probabilità condizionata, una rete Bayesiana può essere utilizzata per svolgere inferenza probabilistica, ossia per aggiornare le distribuzioni di probabilità delle variabili quando si osservano nuove informazioni. In questo contesto, l'inferenza coincide con un processo di *belief updating*: a partire dalle probabilità a priori implicite nel modello, si calcolano le probabilità a posteriori condizionate alle evidenze disponibili.

Definizione 2.22 (Evidenza). Sia $X = (X_1, X_2, \ldots, X_n)$ *l'insieme delle variabili casuali di una rete Bayesiana. Un'evidenza è un sottoinsieme* $E \subseteq X$ *di variabili osservate, per le quali sono noti i valori assunti. L'evidenza induce un vincolo sulla distribuzione di probabilità della rete e viene utilizzata per aggiornare le probabilità delle variabili non osservate mediante inferenza probabilistica.*

L'evidenza rappresenta l'insieme delle informazioni osservate sul sistema e costituisce il punto di partenza del processo di inferenza nelle reti Bayesiane. Introdurre un'evidenza nel modello significa vincolare una o più variabili a determinati stati, modificando di conseguenza la distribuzione di probabilità delle altre variabili della rete in modo coerente con la struttura del grafo e con le tabelle di probabilità condizionata.

Il processo di aggiornamento delle probabilità che segue l'introduzione dell'evidenza prende il nome di *belief updating* e consente di passare dalle probabilità a priori alle probabilità a posteriori, riflettendo l'effetto informativo delle osservazioni disponibili.

Definizione 2.23 (Probabilità a posteriori). Data una rete Bayesiana con n nodi definiti sulla base di n variabili casuali X, e dato un insieme di evidenze osservate E. La probabilità che una variabile X_i *assuma lo stato* x_i *è definita come:*

$$Bel(X_i = x_i) \triangleq P(x_i|e)$$

Nella forma più comune, l'evidenza è detta *hard evidence* o *evidenza specifica* quando l'osservazione assegna con certezza uno stato a una variabile. In tal caso, la probabilità dello stato osservato è pari a uno, mentre quella degli altri stati è posta pari a zero. Questo tipo di evidenza è tipico delle osservazioni dirette e completamente affidabili.

In alcune applicazioni è tuttavia possibile disporre di informazioni parziali o incerte, che non consentono di determinare con certezza lo stato di una variabile. In tali situazioni si parla di *soft evidence* (o evidenza virtuale), che non fissa il valore della variabile osservata, ma ne modifica la distribuzione di probabilità in modo graduale, riflettendo il grado di affidabilità dell'informazione disponibile. Questo tipo di evidenza risulta particolarmente utile in contesti decisionali reali, nei quali le osservazioni possono essere rumorose o soggettive.

A seconda della direzione lungo cui l'informazione si propaga nella rete Bayesiana e del ruolo assunto dalle variabili osservate, è possibile distinguere differenti tipologie di inferenza probabilistica. Ciascuna di esse corrisponde a una specifica modalità di ragionamento e consente di rispondere a domande di natura diversa all'interno del modello.

In particolare, si distinguono tre principali tipi di inferenza: predittiva, diagnostica e inter-causale.

1. Inferenza predittiva (*forward evidence*)
   Consiste nel propagare l'informazione dalle cause agli effetti. In questo caso, l'evidenza viene osservata su una o più variabili genitrici e l'obiettivo è stimare la distribuzione di probabilità delle variabili discendenti.
   Questo tipo di inferenza è tipico dei problemi di previsione, nei quali si vogliono valutare le conseguenze probabili di determinate condizioni iniziali. Ad esempio, osservando lo stato di alcune variabili che rappresentano fattori di input o decisioni strategiche, è possibile stimare la probabilità di differenti esiti futuri all'interno del sistema modellato.

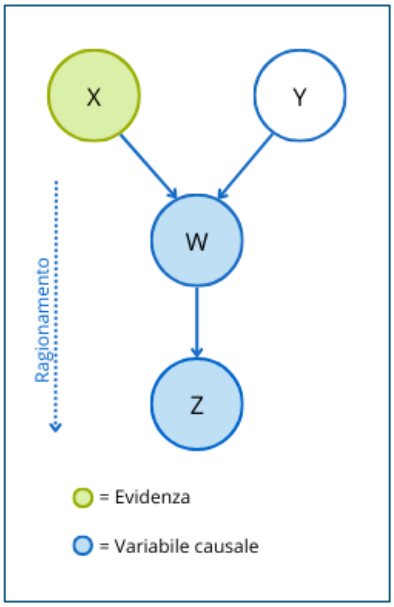


Figura 2.6: Inferenza predittiva.
Fonte: elaborazione grafica personale

2. Inferenza diagnostica (*backward evidence*)

   Opera nella direzione opposta rispetto a quella predittiva, propagando l'informazione dagli effetti alle cause. In questo caso, l'evidenza viene osservata su una o più variabili figlie e l'obiettivo è inferire le probabilità a posteriori delle variabili genitrici.

   Tale modalità di inferenza è tipica dei problemi di diagnosi e analisi delle cause, nei quali si osserva un determinato risultato e si desidera valutare quali condizioni o fattori abbiano maggiormente contribuito al suo verificarsi.

   Questo tipo di ragionamento risulta particolarmente utile nei contesti decisionali in cui si analizzano fenomeni complessi a partire da indicatori osservabili.

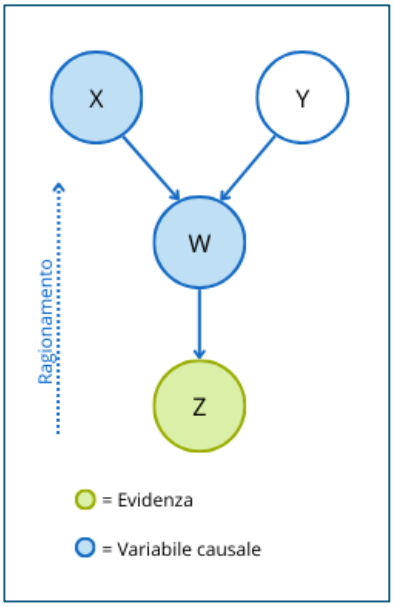


Figura 2.7: Inferenza diagnostica. Fonte: elaborazione grafica personale

3. Inferenza inter – causale (*explaning way*)

   L'inferenza inter-causale si verifica in presenza di strutture convergenti, in cui più variabili genitrici influenzano una stessa variabile figlia. In questo caso, l'osservazione dell'effetto comune induce una dipendenza tra le cause, anche se queste risultano marginalmente indipendenti in assenza di evidenza.

   Questo fenomeno, noto come *explaining way*, implica che la conoscenza dello stato di una delle cause modifica la probabilità delle altre, dato l'effetto osservato.

   L'inferenza inter-causale rappresenta uno degli aspetti più controintuitivi delle reti Bayesiane ed evidenzia la capacità del modello di catturare dinamiche complesse di interazione tra variabili.

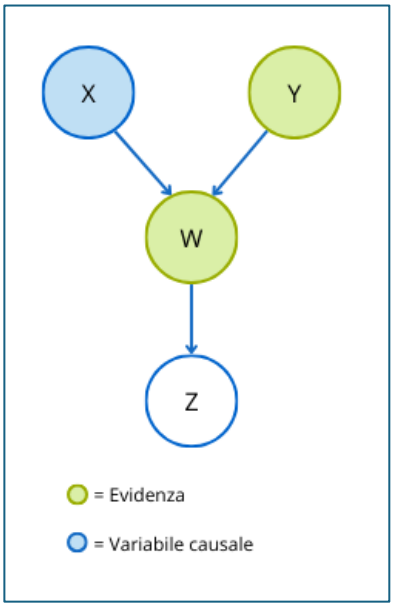


Figura 2.8: Inferenza inter-causale. Fonte: elaborazione grafica personale

Le diverse tipologie di inferenza non sono mutuamente esclusive, ma coesistono all'interno di una stessa rete Bayesiana e possono essere attivate simultaneamente in funzione dell'evidenza disponibile. Questa flessibilità rende le reti Bayesiane uno strumento particolarmente efficace per il supporto alle decisioni, consentendo di analizzare sia gli effetti di determinate scelte sia le possibili cause dei risultati osservati.

Le tipologie di inferenza descritte definiscono la natura delle domande che è possibile porre a una rete Bayesiana e la direzione lungo cui l'informazione viene propagata nel grafo. Tuttavia, tali schemi concettuali non specificano come il calcolo delle probabilità a posteriori venga effettivamente eseguito dal punto di vista computazionale.

Per rispondere operativamente alle domande di inferenza è quindi necessario introdurre algoritmi in grado di calcolare le distribuzioni di probabilità condizionate a partire dalla struttura della rete e dalle evidenze osservate. In funzione delle caratteristiche del grafo e dei vincoli computazionali, tali algoritmi possono essere distinti in metodi di inferenza esatta e metodi di inferenza approssimata.

2.4.1 Inferenza esatta e approssimata

L'inferenza esatta nelle reti Bayesiane consiste nel calcolo preciso delle probabilità a posteriori delle variabili di interesse, dato un insieme di evidenze osservate, senza introdurre approssimazioni. Questo approccio garantisce risultati teoricamente corretti e pienamente coerenti con le assunzioni probabilistiche della rete Bayesiana.

Tuttavia, il principale limite dell'inferenza esatta risiede nella sua complessità computazionale. In generale, il problema del calcolo esatto delle probabilità a posteriori è a livello computazionale, in quanto eccessivamente oneroso, poiché il numero di configurazioni possibili cresce esponenzialmente con il numero di variabili e con il grado di connessione del grafo.
Di conseguenza, l'applicabilità pratica dell'inferenza esatta dipende in larga misura dalla struttura della rete Bayesiana, rendendo necessario distinguere tra reti a struttura semplice e reti caratterizzate da connessioni multiple.

Definizione 2.24 (Rete Bayesiana a struttura semplice). Una rete Bayesiana si dice a struttura semplice se il suo DAG presenta al più un cammino non diretto tra ogni coppia di nodi. Ovvero, ignorando la direzione degli archi, il grafo sottostante non contiene cicli.

In questi casi, l'informazione si propaga lungo la rete senza ambiguità, consentendo l'applicazione efficiente di algoritmi di inferenza esatta.

Definizione 2.25 (Rete Bayesiana a connessioni multiple). Una rete Bayesiana si dice a connessioni multiple quando il suo DAG presenta due o più cammini non diretti tra almeno una coppia di nodi. Ovvero, ignorando la direzione degli archi, il grafo sottostante contiene uno o più cicli.

In questi casi, la presenza di connessioni multiple implica dipendenze strutturali che impediscono l'applicazione diretta degli algoritmi di inferenza esatta, rendendo necessario l'utilizzo di tecniche di inferenza più articolate per garantire il calcolo esatto delle probabilità posteriori.

In contrapposizione all'inferenza esatta, l'inferenza approssimata viene adottata in tutte quelle situazioni in cui il costo computazionale richiesto dagli algoritmi esatti risulta eccessivo o non sostenibile. Questo scenario si presenta tipicamente nel caso di reti Bayesiane particolarmente complesse o caratterizzate da un'elevata densità di connessioni.

In tali contesti, l'esecuzione di algoritmi di inferenza esatta può diventare impraticabile a causa dell'elevato numero di operazioni necessarie per il calcolo delle probabilità a posteriori. Di conseguenza, si rende necessario ricorrere a metodologie approssimative, in grado di fornire stime delle probabilità aggiornate riducendo sensibilmente il carico computazionale, a scapito di una precisione assoluta.

Gli algoritmi sviluppati per questo scopo sono numerosi e differiscono per approccio, ipotesi di base e caratteristiche operative. Alcuni metodi si fondano sull'esecuzione di simulazioni stocastiche, altri sfruttano particolari rappresentazioni grafiche del modello, l'impiego di evidenze virtuali oppure valutazioni approssimate del processo inferenziale. Tra le tecniche più note e utilizzate in letteratura si possono citare, a titolo esemplificativo, il *logic sampling*, il *likelihood weighting* e i metodi basati su *Markov Chain Monte Carlo* (*Korb & Nicholson, 2011*).

2.4.2 Inferenza esatta – strutture a catena

Catena a due nodi

La struttura a catena a due nodi rappresenta la configurazione più semplice di una rete Bayesiana ed è costituita da due variabili X e Y collegate da un arco diretto X→Y. Tale struttura descrive una dipendenza probabilistica diretta, nella quale la variabile X influenza la variabile Y.

Figura 2.9: Relazione a catena con evidenza su nodo genitore. Fonte: elaborazione grafica personale

In questa situazione quando l'evidenza viene osservata su un nodo genitore, ponendo X= *e*, la probabilità a posteriori del nodo figlio Y coincide direttamente con la probabilità condizionata:

$$P(Y|X = e)$$

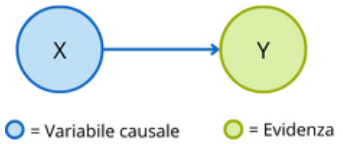


Figura 2.10: Relazione a catena con evidenza su nodo figlio. Fonte: elaborazione grafica personale

Al contrario, qualora l'evidenza sia disponibile sul nodo figlio, ovvero Y=*e*, l'aggiornamento della distribuzione di probabilità del nodo genitore richiede l'applicazione del teorema di Bayes:

$$\begin{aligned} Bel(X = x) &= P(X = x|Y = y) \\ &= \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)} \\ &= \alpha P(x)\lambda(x) \end{aligned}$$

Dove con P(x) si indica la probabilità a priori marginale per il nodo X, P(X=x), e con $\lambda = P\ (Y = e|X = x)$ la probabilità condizionata.

Poiché la somma delle probabilità dei valori di X deve essere 1, il termine α è la costante di normalizzazione, calcolata come:

$$\alpha = \frac{1}{P(Y = y)}$$

Questo caso elementare consente di illustrare il meccanismo di base dell'inferenza esatta, mostrando come l'osservazione di una variabile modifichi la distribuzione di probabilità dell'altra attraverso l'applicazione della probabilità condizionata.

Catena a tre nodi

Il procedimento di inferenza esatta illustrato per la struttura a due nodi può essere esteso in modo naturale al caso di una catena composta da tre variabili casuali X→Y→Z. In questa configurazione, il processo di aggiornamento delle probabilità dipende dalla posizione del nodo su cui viene osservata l'evidenza.

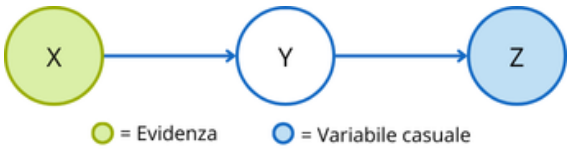


Figura 2.11: Relazione a catena con evidenza su nodo radice. Fonte: elaborazione grafica personale

Quando l'evidenza è disponibile sul nodo radice, ponendo X=x, l'aggiornamento della distribuzione di probabilità del nodo finale Z avviene nella stessa direzione degli archi del grafo. In questo caso, l'inferenza predittiva può essere effettuata applicando la regola del prodotto e sfruttando le indipendenze condizionate implicite nella struttura della rete, ottenendo:

$$Bel(Z) = P(Z|X = x) = \sum_{Y=y} P(Z|Y)P(Y|X = x)$$

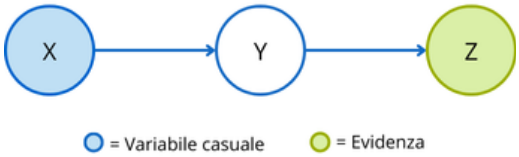


Figura 2.12: Relazione a catena con evidenza su nodo foglio. Fonte: elaborazione grafica personale

Al contrario, quando l'evidenza è osservata sul nodo foglia, ponendo Z=z, l'aggiornamento della distribuzione di probabilità del nodo radice X richiede un processo di inferenza

diagnostica.

In questo caso, la probabilità a posteriori di X si ottiene applicando il teorema di Bayes in combinazione con la regola della catena:

$$
\begin{aligned}
Bel(X = x) &= P(X = x|Z = z) \\
&= \frac{P(Z = z|X = x)}{P(Z = z)} \\
&= \frac{\sum_{Y=y} P(Z = z|Y = y, X = x)P(Y = y|X = x)P(X = x)}{P(Z = z)} \\
&= \frac{\sum_{Y=y} P(Z = z|Y = y)P(Y = y|X = x)P(X = x)}{P(Z = z)} \quad (Z \perp X \mid Y) \\
&= \alpha P(x)\lambda(x)
\end{aligned}
$$

Dove

$$
\lambda(\mathrm{x}) = \mathrm{P}(\mathrm{Z} = \mathrm{z}|\mathrm{X} = \mathrm{x}) = \sum_{Y=y} P(Z = z|Y = y)P(Y = y|X = x)
$$

2.4.3 *Inferenza esatta – strutture a polialbero*

Le strutture a polialbero rappresentano una generalizzazione delle strutture a catena e sono costituite da reti Bayesiane nelle quali esiste al più un cammino non diretto tra ogni coppia di nodi. Questa proprietà strutturale consente di estendere il meccanismo di inferenza esatta illustrato per le catene a reti di dimensione maggiore, mantenendo un calcolo computazionalmente efficiente.

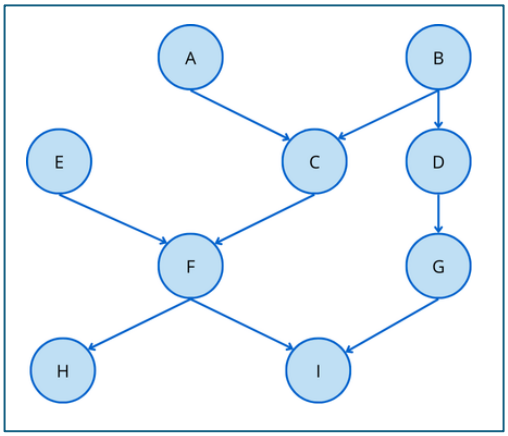


Figura 2.13: Rappresentazione grafica di una struttura a polialbero. Fonte: elaborazione grafica personale

Nei polialberi, l'inferenza esatta si basa su un principio di calcolo locale, secondo il quale ciascun nodo aggiorna la propria distribuzione di probabilità utilizzando esclusivamente l'informazione proveniente dai nodi direttamente connessi. Come nel caso delle strutture a catena, l'informazione si propaga lungo gli archi del grafo attraverso due flussi distinti: un flusso predittivo, che si muove dai nodi genitori verso i nodi figli, e un flusso diagnostico, che si propaga dai nodi figli verso i nodi genitori.

In presenza di evidenza su uno o più nodi della rete, il processo inferenziale tiene conto delle dipendenze probabilistiche locali codificate nelle tabelle di probabilità condizionata e delle indipendenze condizionate implicite nella struttura del grafo. Le variabili non osservate vengono gestite mediante operazioni di marginalizzazione, analogamente a quanto avviene nel caso delle catene, evitando il calcolo esplicito della distribuzione congiunta su tutte le variabili della rete.

L'assenza di cicli non diretti garantisce che la propagazione dell'informazione non generi ambiguità o dipendenze circolari, consentendo la convergenza del processo di aggiornamento delle probabilità.

Il meccanismo di inferenza esatta nei polialberi, descritto in termini qualitativi come propagazione locale dell'informazione, può essere formalizzato introducendo una rappresentazione esplicita dei contributi informativi che convergono su un generico nodo della rete. Seguendo l'impostazione proposta da Kim e Pearl, l'insieme delle evidenze che influenzano un nodo può essere suddiviso in due componenti distinte, corrispondenti ai flussi informativi predittivo e diagnostico già discussi.

Questa formalizzazione consente di esprimere il processo di aggiornamento delle probabilità a posteriori in modo compatto e sistematico, rendendo esplicita la logica di calcolo sottostante all'algoritmo di propagazione dei messaggi nei polialberi.

L'algoritmo di propagazione dei messaggi di Kim e Pearl consente di aggiornare localmente la belief di ciascun nodo di un polialbero combinando informazioni provenienti dai nodi adiacenti e dalle relazioni probabilistiche locali. Per un generico nodo X, l'algoritmo richiede il mantenimento delle seguenti tre tipologie di parametri:

- *Supporto predittivo*

  Per ciascun arco entrante $U_i \rightarrow X$, viene definito il contributo di supporto predittivo

  $$\pi_x(U_i) = P(U_i \,|E_{U_i/x})$$

  Dove $E_{U_i/x}$rappresenta l'insieme delle evidenze connesse al nodo genitore U_i, escludendo quelle che transitano attraverso X. Tale quantità esprime l'influenza informativa che il nodo U_i esercita su X sulla base delle evidenze disponibili.
- *Supporto diagnostico*

  Per ciascun arco uscente $X \rightarrow Y_j$, viene definito il contributo di supporto diagnostico

  $$\lambda_{Y_j}(X) = P\left(E_{Y_J/X}\middle|X\right)$$

  dove $E_{Y_j/X}$ indica l'insieme delle evidenze associate al nodo figlio Y_j, al netto delle informazioni che passano attraverso X. Questo termine sintetizza l'informazione che risale dai nodi figli verso il nodo X

- *Tabella di probabilità condizionata (CPT)*

  L'algoritmo utilizza inoltre la tabella di probabilità condizionata $P\ (X|U_i, \ldots, U_n)$ che descrive la relazione probabilistica locale tra il nodo X e l'insieme dei suoi nodi genitori $U_i, \ldots, U_)$.

Questi tre elementi vengono combinati per aggiornare la belief del nodo X e per generare nuovi messaggi da trasmettere ai nodi adiacenti, questo processo viene sviluppato in tre fasi.

1. *Belief updating*:

   In questa fase, il nodo X aggiorna il proprio supporto predittivo combinando i messaggi $\pi_x(U_i)$ ricevuti dai nodi genitori $U_i, \ldots, U_n$ con la propria tabella di probabilità condizionata $P\ (X|U_i, \ldots, U_n)$. Questo passaggio consente di integrare l'informazione che fluisce dall'alto verso il basso nella rete, riflettendo l'influenza delle variabili genitrici sul nodo X, applicando la formula:

$$Bel(x_i) = \alpha\lambda(x_i)\pi(x_i)$$

Dove,

$$\lambda(x_i) = \begin{cases} 1 & se\ l'evidenza \text{ è } X = x_i \\ 0 & se\ l'evidenza \text{ è } per\ un\ altro\ x_i \\ \prod_j \lambda_{Y_j}(x_i), & in\ tutti\ gli\ altri\ casi \end{cases}$$

Questa equazione specifica il criterio secondo cui viene assegnato il valore al termine λ(xi) in funzione dell'evidenza disponibile. In particolare, nel caso in cui l'osservazione riguardi direttamente la variabile X e assuma il valore x_i, il corrispondente contributo diagnostico è fissato pari a uno. Qualora, invece, l'evidenza osservata sia associata a un valore alternativo x_j, il parametro assume valore nullo. In assenza di osservazioni dirette su X, il contributo diagnostico è determinato aggregando le informazioni provenienti dai nodi discendenti, attraverso il prodotto dei messaggi ricevuti.

Il calcolo del supporto predittivo è invece descritto nell'equazione successiva, che esprime $\pi(x_i)$ come una combinazione delle probabilità condizionate locali e dei messaggi trasmessi dai nodi genitori. In particolare, il valore di $\pi(x_i)$ si ottiene considerando tutte le possibili configurazioni delle variabili genitrici e sommando, per ciascuna di esse, il prodotto tra la probabilità condizionata del nodo e i relativi contributi predittivi:

$$\pi(x_i) = \sum_{u_1,\dots,u_n} P(x_i) \mid (u_1, \dots, u_n) \prod_i \pi_X(u_i)$$

Il fattore α svolge infine il ruolo di costante di normalizzazione, assicurando che la distribuzione di probabilità a posteriori della variabile X risulti correttamente normalizzata, $\sum_{x_j} Bel(X = x_i) = 1$

2. *Bottom-up propagation:*
   in questa fase il nodo X elabora le informazioni diagnostiche ricevute dai propri nodi figli e le trasmette ai nodi genitori sotto forma di messaggi λ. L'informazione fluisce quindi dal basso verso l'alto nella struttura della rete.

Il contenuto del messaggio $\lambda(u_i)$ inviato da X a un generico nodo genitore U_i è determinato combinando il supporto diagnostico complessivo ricevuto dai figli, i contributi predittivi provenienti dagli altri nodi genitori e i valori della tabella di probabilità condizionata associata al nodo X, secondo la relazione:

$$\lambda(u_i) = \sum_{x_i} \lambda(x_i) \sum_{u_k; k \neq i} P(x_i \,|u_1, \dots, u_n) \prod_{k \neq 1} \pi_X(u_k)$$

I messaggi λ trasmessi ai nodi genitori sintetizzano pertanto tre differenti sorgenti informative: le evidenze provenienti dai nodi figli di X, le informazioni predittive ricevute dagli altri genitori e le dipendenze probabilistiche locali descritte dalla distribuzione condizionata del nodo.

3. *Top-down propagation*

   Questa fase riguarda la generazione dei messaggi predittivi che il nodo X invia ai propri nodi figli. In questo caso, l'informazione si propaga dall'alto verso il basso nella rete e riflette l'effetto combinato delle evidenze disponibili sul nodo X e sulle variabili a esso collegate.

   Il messaggio $\pi_{Y_j}(x_i)$ inviato dal nodo X a un nodo figlio Y_j è definito in funzione dello stato di evidenza associato alla variabile X. In particolare, esso assume valore unitario qualora l'evidenza osservata corrisponda allo stato x_i, mentre risulta nullo se l'evidenza riguarda uno stato alternativo x_j. In assenza di osservazioni dirette su X, il messaggio è determinato combinando il supporto diagnostico proveniente dagli altri nodi figli, il supporto predittivo ricevuto dai nodi genitori e la tabella di probabilità condizionata del nodo, secondo la seguente espressione:

$$f(x) = \begin{cases} 1 & se\ il\ valore\ dell'evidenza\ \text{è}\ x_i \\ 0 & se\ l'evidenza\ riguarda\ un\ altro\ valore\ di\ x_j \\ \alpha \left[\prod_{k \neq j} \lambda_{Y_k}(x_i)\right] \sum_{u_1,\dots,u_n} P(x_i \,|u_1, \dots, u_n) \prod_i \pi_x(u_i) = \dfrac{\alpha Bel(x_i)}{\lambda_{Y_j}(x_j)} \end{cases}$$

Le tre fasi di aggiornamento descritte mostrano come, nelle reti Bayesiane a struttura polialbero, l'inferenza esatta possa essere realizzata attraverso operazioni locali di propagazione dei

messaggi. Grazie all'assenza di cicli non diretti, l'informazione fluisce in modo coerente e il processo di aggiornamento converge a una distribuzione di probabilità a posteriori corretta.

Questo meccanismo costituisce il fondamento teorico dell'algoritmo di Kim e Pearl e rappresenta il riferimento principale per l'inferenza esatta nelle reti Bayesiane a struttura semplice (*Korb & Nicholson, 2011*).

2.4.4 Inferenza esatta – reti a connessioni multiple

Nella pratica applicativa, i fenomeni di interesse sono raramente riconducibili a strutture elementari come le catene o i polialberi. Nella maggior parte dei casi, la modellazione richiede l'utilizzo di reti Bayesiane più articolate. In questo ambito si collocano le reti a connessioni multiple, note in letteratura come *multiply connected networks* (MCN).

Una rete Bayesiana viene definita a connessioni multiple quando esiste almeno una coppia di nodi collegata da più di un percorso non direzionato. Questa proprietà implica che l'influenza esercitata da una variabile su un'altra possa propagarsi attraverso cammini alternativi all'interno del grafo, generando interazioni più articolate rispetto a quelle osservabili nelle reti a connessione singola.

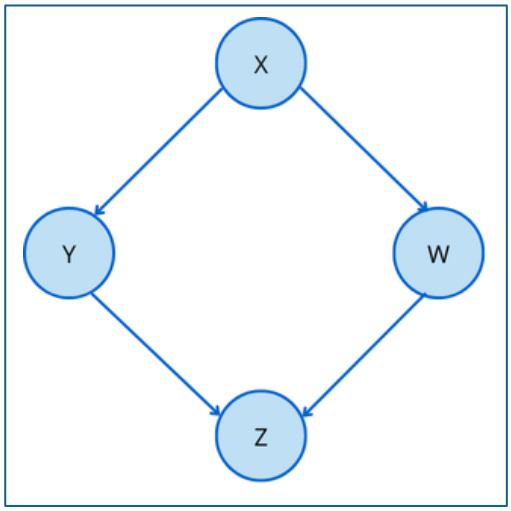


Figura 2.14: Rappresentazione rete a connessioni multiple. Fonte: elaborazione grafica personale

In presenza di reti di questo tipo, l'algoritmo di propagazione dei messaggi proposto da Kim e Pearl non risulta direttamente applicabile. La causa principale di questa limitazione è riconducibile alla presenza di percorsi multipli attivi, che possono condurre a una trasmissione ridondante dell'informazione.

Più precisamente, un'evidenza introdotta su un nodo, ad esempio X, può generare messaggi che raggiungono il nodo Z attraverso più cammini distinti. Questo fenomeno provoca una duplicazione dell'informazione propagata, violando le assunzioni di indipendenza condizionata alla base degli algoritmi di inferenza locale e compromettendo la correttezza dell'aggiornamento delle probabilità.

Per ovviare a queste difficoltà, sono stati sviluppati metodi di inferenza specificamente progettati per trattare reti Bayesiane con connessioni multiple. Tali approcci si basano su tecniche di raggruppamento delle variabili (*clustering*) e consentono di trasformare la rete originale in una struttura equivalente, più adatta alla propagazione corretta dell'evidenza. Tra questi metodi, un ruolo centrale è rivestito dall'algoritmo dei *junction tree*, ampiamente utilizzato per l'inferenza esatta in reti complesse (*Korb & Nicholson, 2011*).

L'algoritmo di *junction tree* clustering rappresenta uno degli strumenti più efficaci per l'inferenza probabilistica nelle reti Bayesiane a connessioni multiple. Esso appartiene alla classe dei metodi di *clustering*, che operano attraverso una riorganizzazione strutturale del modello al fine di eliminare i percorsi multipli responsabili della propagazione ridondante dell'informazione.

Il principio alla base dell'algoritmo consiste nel trasformare la rete Bayesiana originale in una nuova struttura ad albero, nella quale l'inferenza può essere eseguita mediante tecniche di propagazione locale dei messaggi. Tale trasformazione è sempre possibile, poiché ogni *multiply connected network* può essere ricondotta a una rappresentazione equivalente che preserva la distribuzione di probabilità congiunta del modello iniziale.

Il procedimento si articola in sei fasi fondamentali:

1. *Moralizzazione del grafo*
   In una prima fase, il grafo diretto della rete Bayesiana viene convertito in un grafo non direzionato. A tal fine, tutti i nodi che condividono un figlio comune vengono collegati tra loro mediante un arco, qualora tale collegamento non sia già presente. Il grafo risultante prende il nome di *moral graph*. Ad esempio, nella struttura mostrata in Figura 2.6, i nodi Y e W vengono connessi poiché entrambi sono genitori del nodo Z.

2. *Triangolazione*

   Il grafo moralizzato viene successivamente modificato introducendo archi aggiuntivi in modo tale che ogni ciclo di lunghezza superiore a tre presenti almeno una corda. Il risultato di questa operazione è un *grafo triangolato*. La scelta della triangolazione non è univoca e ha un impatto diretto sulla complessità computazionale del metodo, poiché triangolazioni differenti possono generare cluster di dimensioni diverse.

3. *Costruzione del junction tree*

   A partire dal grafo triangolato vengono individuate le cricche massimali, che costituiscono i nodi della nuova struttura ad albero, detta *junction tree*. Ogni nodo del junction tree rappresenta un insieme di variabili trattate congiuntamente nel processo inferenziale.

4. *Identificazione dei separatori*

   Per ogni arco che collega due nodi del junction tree viene definito un separatore, dato dall'intersezione delle variabili contenute nelle due cricche adiacenti. I separatori garantiscono la coerenza globale del modello e fungono da supporto per lo scambio dei messaggi durante l'inferenza.

5. *Ridefinizione dei parametri probabilistici*

   A ciascun nodo del junction tree viene associata una nuova tabella di probabilità, costruita sulla base delle variabili che compongono la cricca. In particolare:

   a. si individua una cricca Y che contenga una variabile X e tutti i suoi genitori;
   b. la probabilità locale $P\left(X\middle|P_a(X)\right)$ viene incorporata nella tabella associata alla cricca Y.

6. *Aggiornamento delle credenze*

   Nell'ultima fase, l'evidenza viene introdotta nel modello e le probabilità vengono aggiornate attraverso un algoritmo di propagazione dei messaggi applicato alla struttura del junction tree, concettualmente analogo a quello di Kim e Pearl, ma adattato ai nodi composti.

Al termine di questa procedura, la struttura ottenuta è in grado di rappresentare la medesima distribuzione di probabilità congiunta della rete Bayesiana di partenza, consentendo di eseguire inferenza esatta anche in presenza di connessioni multiple.

Per un maggiore approfondimento delle tecniche di inferenza nelle reti Bayesiane, si rimanda ai lavori fondamentali di Kim e Pearl, in particolare *A Computational Model for Causal and*

Diagnostic Reasoning in Inference Systems (1983), nei quali vengono introdotti i principi della belief propagation, successivamente inquadrati in un contesto teorico più generale da Pearl in *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (1988) e formalizzati in forma moderna e operativa in testi di riferimento quali *Bayesian Artificial Intelligence* di Korb e Nicholson (2011).

2.5 Apprendimento delle reti Bayesiane

In contesti applicativi reali spesso né le relazioni di dipendenza tra le variabili né le distribuzioni di probabilità che le quantificano sono completamente note. Diventa quindi necessario introdurre un processo di apprendimento, finalizzato alla costruzione della rete Bayesiana a partire dalle informazioni disponibili.

Con il termine *apprendimento delle reti Bayesiane* si intende l'insieme delle metodologie che consentono di stimare il modello sulla base di dati osservati, di conoscenze esperte, oppure di una combinazione di entrambe le fonti. In generale, il problema dell'apprendimento può essere scomposto in due aspetti distinti: da un lato, l'apprendimento della struttura, che consiste nell'individuazione del grafo aciclico diretto più adeguato a rappresentare le relazioni di dipendenza tra le variabili del sistema; dall'altro, l'apprendimento dei parametri, che riguarda la stima delle tabelle di probabilità condizionata associata a una struttura nota.

In questo paragrafo si andranno quindi a descrivere le metodologie di apprendimento per le diverse situazioni possibili.

2.5.1 Grafo noto

Nel caso in cui la struttura della rete Bayesiana sia nota, il problema dell'apprendimento si riduce alla sola stima dei parametri quantitativi del modello, ovvero delle tabelle di probabilità condizionata associate a ciascun nodo. In questa situazione, si assume che le relazioni di dipendenza tra le variabili siano state già definite, mentre resta da determinare l'entità probabilistica di tali relazioni sulla base delle informazioni disponibili.

Nel caso di dati completi, ovvero quando per ogni osservazione sono noti i valori assunti da tutte le variabili del modello, la stima delle probabilità condizionate può essere effettuata tramite le frequenze relative osservate nel campione. In un'ottica Bayesiana, tali stime possono essere integrate mediante l'introduzione di distribuzioni a priori, che consentono di incorporare informazioni preliminari sul comportamento delle variabili e di agire come meccanismo di regolarizzazione, riducendo l'instabilità delle stime ottenute esclusivamente dalle frequenze osservate.

In contesti reali, tuttavia, è frequente che i dati presentino criticità legate alla stima dei parametri. Anche in presenza di dataset completi, può accadere che alcune configurazioni delle variabili risultino scarsamente rappresentate o del tutto assenti nel campione, portando all'assegnazione di probabilità nulle nelle tabelle di probabilità condizionata. Per evitare questo problema si ricorre a tecniche di *smoothing*, il cui obiettivo è stabilizzare la stima delle probabilità e migliorare la capacità di generalizzazione del modello (*Korb & Nicholson, 2011*).

Una delle tecniche di *smoothing* più diffuse è il *Laplace smoothing*, che consiste nell'aggiungere una quantità costante a ciascun conteggio osservato, evitando così l'assegnazione di probabilità nulle e riducendo l'eccessiva dipendenza del modello dalle frequenze empiriche.

Una problematica distinta si presenta quando il dataset contiene valori mancanti, ossia quando per alcune osservazioni non sono noti i valori assunti da una o più variabili. In questo caso, l'assenza parziale delle informazioni rende inapplicabile il calcolo diretto delle frequenze. Per affrontare tale situazione si utilizza l'algoritmo di *Expectation–Maximization* (EM), che consente di stimare i parametri del modello. Nella fase di *Expectation*, i valori mancanti vengono stimati in modo probabilistico sulla base dei parametri correnti della rete, mentre nella fase di *Maximization* tali stime vengono utilizzate per aggiornare le tabelle di probabilità condizionata. Il procedimento viene ripetuto fino alla convergenza, permettendo di ottenere una parametrizzazione coerente del modello.

Per approfondimenti in tema di apprendimento dei parametri in presenza di dati mancanti, si rimanda alla trattazione proposta in *Bayesian Artificial Intelligence* di Korb e Nicholson (2011), che descrive l'utilizzo dell'algoritmo di Expectation–Maximization nelle reti Bayesiane.

2.5.2 Grafo non noto

Nel caso in cui la struttura della rete Bayesiana non sia nota a priori, il problema dell'apprendimento si estende dalla sola stima dei parametri alla determinazione della struttura del grafo, ovvero all'individuazione degli archi e della loro direzione. L'obiettivo è costruire un grafo aciclico diretto che rappresenti in modo adeguato le relazioni di dipendenza tra le variabili del sistema, sulla base delle informazioni contenute nei dati osservati.

L'apprendimento della struttura è un problema intrinsecamente più complesso rispetto all'apprendimento dei parametri, poiché lo spazio delle possibili reti cresce in modo esponenziale con il numero di variabili. Per questo motivo non è possibile effettuare una ricerca esaustiva tra tutte le strutture ammissibili e si rende necessario il ricorso ad approcci approssimati. In letteratura, le metodologie di apprendimento strutturale delle reti Bayesiane vengono generalmente classificate in approcci basati su vincoli (*constraint-based*), approcci basati su funzioni di scoring (*score-based*) e approcci ibridi, che combinano informazioni provenienti dai dati e dalla conoscenza esperta

Approcci constraint-based

Gli approcci *constraint-based* si fondano sull'analisi delle relazioni di indipendenza condizionata tra le variabili, individuate mediante test statistici applicati ai dati. Sulla base dei risultati di tali test, viene costruita una struttura grafica che rispetti le dipendenze e le indipendenze emerse, facendo esplicito riferimento ai concetti di *d-separation* e alla proprietà di Markov.

Il procedimento tipico di questi metodi prevede una prima fase di identificazione dello scheletro del grafo, inizialmente non orientato, seguita da una fase di orientamento degli archi compatibile con i risultati dei test di indipendenza e con il vincolo di aciclicità.

Tra gli approcci *constraint-based*, uno degli algoritmi più noti e utilizzati per l'apprendimento della struttura delle reti Bayesiane è *il PC algorithm* (*Korb & Nicholson, 2011*). Questo metodo si basa sull'analisi delle relazioni di indipendenza condizionata tra le variabili, ricavate mediante test statistici applicati ai dati osservati, e sfrutta in modo esplicito i concetti di *d-separation* e di proprietà di Markov.

Il funzionamento del *PC algorithm* può essere descritto come un processo articolato in più fasi. In una prima fase, l'algoritmo costruisce uno scheletro non orientato della rete, partendo da un grafo completo in cui tutte le variabili sono inizialmente collegate tra loro. Successivamente, attraverso una serie di test di indipendenza condizionata, vengono progressivamente rimossi gli archi tra coppie di variabili che risultano indipendenti dato un opportuno sottoinsieme delle altre variabili. Al termine di questa fase si ottiene un grafo non orientato che rappresenta le dipendenze statisticamente significative emerse dai dati.

Nella fase successiva, il *PC algorithm* si occupa di orientare gli archi dello scheletro ottenuto, individuando in particolare le cosiddette *v-structure*, ovvero configurazioni in cui due variabili marginalmente indipendenti diventano dipendenti quando si condiziona su una terza. L'orientamento degli archi avviene nel rispetto dei risultati dei test di indipendenza e del vincolo di aciclicità, evitando la formazione di cicli diretti nel grafo.

Il risultato finale del *PC algorithm* è un grafo che appartiene alla classe di equivalenza delle strutture compatibili con le indipendenze osservate nei dati. L'accuratezza dell'algoritmo dipende in modo rilevante dalla qualità dei test di indipendenza condizionata e dalla numerosità del campione: con dataset di dimensioni elevate il metodo tende a produrre strutture affidabili, mentre in presenza di campioni ridotti o dati rumorosi può risultare sensibile agli errori statistici. Nonostante questi limiti, il *PC algorithm* rappresenta un riferimento fondamentale tra i metodi *constraint-based* ed è ampiamente utilizzato come base per lo sviluppo di approcci più avanzati.

Approcci score based

Gli approcci *score-based* affrontano il problema dell'apprendimento strutturale definendo una funzione di valutazione che misura la qualità di una rete Bayesiana rispetto ai dati osservati. A ciascuna struttura candidata viene associato un punteggio che tiene conto sia della capacità del modello di adattarsi ai dati sia della sua complessità, al fine di evitare fenomeni di *overfitting*.

L'apprendimento della struttura viene quindi formulato come un problema di ottimizzazione, in cui si ricerca la rete che massimizza la funzione di scoring.

Di seguito riportati alcuni dei metodi maggiormente utilizzati.

Hill Climbing. È una strategia di ricerca che esplora lo spazio delle strutture modificando iterativamente il grafo corrente (aggiunta, rimozione o inversione di archi), selezionando a ogni passo la struttura che massimizza la funzione di scoring.

K2 (algoritmo). Algoritmo *score-based* che assume un ordinamento causale prefissato delle variabili e ricerca, per ciascun nodo, il miglior insieme di genitori massimizzando una funzione di punteggio locale.

Maximum Likelihood (ML). Criterio di valutazione che seleziona la struttura che massimizza la verosimiglianza dei dati osservati, senza introdurre penalizzazioni per la complessità del modello, risultando quindi soggetto a *overfitting*.

$$ScoreML = P(D|\Theta_G, G_i)$$

Dove G_i indica la struttura candidata e θ_G il corrispondente vettore di parametri.

L'output ottenuto corrisponde alla stima del set di parametri θ che massimizza la funzione di maximum likelihood associata alla rete G. Tuttavia, questo criterio presenta il limite di favorire strutture più complesse, in quanto una maggiore complessità consente in genere una migliore aderenza ai dati osservati D.

BIC (Bayesian Information Criterion). Funzione di scoring che approssima il log-posterior Bayesiano combinando bontà di adattamento ai dati e penalizzazione della complessità del modello, favorendo strutture più parsimoniose.

$$BIC = logP(D|\theta_G) - \frac{d}{2}logN$$

dove *d* rappresenta il numero di parametri della rete Bayesiana e N la dimensione del campione disponibile nel dataset. A differenza della *maximum likelihood*, questa funzione introduce un termine di penalizzazione che scoraggia modelli eccessivamente complessi; ciò comporta tuttavia il rischio opposto di privilegiare strutture troppo semplici. Si osservi inoltre che il criterio BIC non richiede la specificazione delle probabilità a priori delle variabili.

MDL (Minimum Description Length). Principio secondo cui la struttura ottimale è quella che minimizza la lunghezza complessiva della descrizione del modello e dei dati, formalmente equivalente, in molti casi, al criterio BIC.

BS (Bayesian Score). Funzione di scoring completamente bayesiana che valuta una struttura integrando la verosimiglianza sui parametri, tenendo conto di distribuzioni a priori sulla struttura e sui parametri del modello.

$$P(G_i|D) = P(G_i)P(D|G_i)/P(D)$$

dove $P(G_i)$ indica la probabilità a posteriori associata alla struttura *Gi*, che può essere assunta uniforme in assenza di informazioni preliminari, oppure specificata sulla base di conoscenze fornite da esperti del dominio.

BDe (Bayesian Dirichlet equivalent). Variante del Bayesian Score basata su priori di tipo Dirichlet, progettata per garantire l'equivalenza di punteggio tra strutture appartenenti alla stessa classe di equivalenza.

Considerata una generica configurazione in cui $X_i=k$ dato $Pa(X_i)=j$, e indicando con N′ la dimensione del campione, il punteggio associato a tale configurazione è determinato dalla seguente espressione:

$$N_{ijk} = P(X_i = k, P_a(X_i) = j|G_i, \theta_G)xN'$$

K2 score. Funzione di scoring associata all'algoritmo K2, che assume una distribuzione a priori uniforme sui parametri e valuta localmente ogni nodo dato l'insieme dei suoi genitori.

Rappresenta una semplificazione del criterio BS e consiste nel decomporre il punteggio complessivo, calcolandolo separatamente per ciascuna variabile X_i, una volta noto l'insieme dei suoi genitori $Pa(X_i)$, ottenendo:

$$S_i = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}!$$

Per una formulazione matematica dettagliata dei criteri di scoring si rimanda a Korb e Nicholson (2011).

Approcci ibridi

Nei contesti applicativi reali, la disponibilità di dati può essere limitata o affetta da rumore, rendendo difficoltosa l'identificazione di una struttura affidabile basata esclusivamente su procedure automatiche. Per questo motivo, è frequente l'adozione di approcci ibridi, che combinano l'apprendimento dai dati con l'integrazione di conoscenze fornite da esperti di dominio.

La conoscenza esperta può essere utilizzata in modo restrittivo, imponendo vincoli sulla presenza o sull'orientamento di specifici archi e riducendo lo spazio delle strutture ammissibili, oppure in modo propositivo, validando o raffinando una struttura inizialmente ipotizzata. Questo tipo di approccio consente di migliorare l'interpretabilità del modello e di ottenere reti Bayesiane più coerenti con il fenomeno analizzato, specialmente in presenza di dataset di dimensioni ridotte.

2.5.3 Approcci ibridi

L'apprendimento automatico della struttura di una rete Bayesiana può incontrare notevoli difficoltà quando la quantità di informazioni disponibili è limitata o quando i dati risultano rumorosi. In tali situazioni, affidarsi esclusivamente ai dati empirici può condurre a modelli poco affidabili o scarsamente interpretabili. Per questo motivo, nella pratica applicativa si ricorre spesso all'integrazione di conoscenze di dominio, fornite da esperti del settore, che permettono di supportare e guidare il processo di costruzione della rete.

L'apporto degli esperti non si limita alla sola definizione della struttura, ma può risultare rilevante anche nella stima dei parametri, soprattutto quando alcune distribuzioni di probabilità sono difficili da determinare in modo accurato a partire dai dati. Per quanto riguarda specificamente l'apprendimento della struttura, l'informazione a priori può essere incorporata secondo due strategie concettualmente distinte.

- **Utilizzo vincolante della conoscenza esperta**
  In questa modalità, le informazioni fornite dagli esperti vengono impiegate come vincoli strutturali, che delimitano l'insieme delle reti ammissibili. Tali vincoli permettono di ridurre sensibilmente la complessità del problema, orientando gli algoritmi di apprendimento verso soluzioni compatibili con le relazioni causali ritenute plausibili. A titolo esemplificativo, la conoscenza di dominio può essere utilizzata per definire precedenze causali tra variabili, fissare o escludere l'orientamento di determinati archi, consentire relazioni direzionali alternative o impedire collegamenti diretti tra specifiche coppie di nodi.
- **Validazione empirica di strutture esperte**
  In alternativa, la struttura della rete può essere inizialmente formulata sulla base delle conoscenze degli esperti e successivamente sottoposta a valutazione tramite i dati. In questo caso, l'apprendimento non ha lo scopo di esplorare lo spazio delle possibili

topologie, ma di verificare se la rete proposta sia coerente con le evidenze osservate, consentendo eventuali aggiustamenti o conferme del modello iniziale.

In sintesi, le reti Bayesiane forniscono un quadro teorico rigoroso e al tempo stesso flessibile per rappresentare sistemi complessi caratterizzati da incertezza, integrando in modo coerente struttura grafica e informazione probabilistica. Attraverso i concetti di dipendenza condizionata, d-separation e Markov Blanket, esse consentono di interpretare le relazioni tra le variabili non solo dal punto di vista quantitativo, ma anche strutturale, rendendo espliciti i meccanismi di propagazione dell'informazione all'interno del modello. Queste proprietà costituiscono il fondamento dei processi di inferenza e apprendimento che verranno approfonditi nei paragrafi successivi, nei quali si mostrerà come tali modelli possano essere concretamente utilizzati per l'analisi dei dati e il supporto alle decisioni in contesti reali.

2.6 Estensioni

Nei paragrafi precedenti sono state presentate le reti Bayesiane nel loro formalismo di base, mettendo in evidenza la rappresentazione grafica delle dipendenze probabilistiche e la fattorizzazione della distribuzione congiunta in termini di distribuzioni condizionate locali. Tale impostazione fornisce un quadro teorico solido e ampiamente utilizzato, ma risulta limitata quando il dominio di interesse richiede una modellazione più articolata delle variabili e delle relazioni che le legano.

In numerosi contesti applicativi, infatti, le grandezze osservate sono naturalmente espresse su scale numeriche continue o derivano dalla combinazione di informazioni di natura diversa, qualitative e quantitative. In questi casi, per una rappresentazione più fedele dello scenario è necessaria l'adozione di modelli probabilistici in grado di trattare in modo diretto tali caratteristiche senza ricorrere a semplificazioni preliminari sui dati.

Per rispondere a queste esigenze, la letteratura ha proposto estensioni del formalismo delle reti Bayesiane che mantengono la struttura a grafo aciclico diretto e i meccanismi di inferenza probabilistica, introducendo al contempo ipotesi più flessibili sulla forma delle distribuzioni condizionate locali. Tra queste rivestono un ruolo centrale le *Gaussian Bayesian Networks* e le *Conditional Linear Gaussian Bayesian Networks*, che permettono rispettivamente di modellare domini continui e domini misti, come verrà discusso nei paragrafi seguenti.

2.6.1 Gaussian Bayesian Network

Le *Gaussian Bayesian Networks* (GBN) rappresentano una classe di reti Bayesiane definite su insiemi di variabili casuali continue, nelle quali tutte le distribuzioni condizionate locali sono gaussiane. In tali modelli, la distribuzione di probabilità congiunta delle variabili è una distribuzione normale multivariata e la struttura a grafo aciclico diretto codifica le relazioni di dipendenza condizionata tra le variabili, analogamente a quanto avviene nel caso discreto.

In una GBN, ciascuna variabile continua è modellata come una funzione lineare delle variabili appartenenti al proprio insieme di genitori, a cui si aggiunge un termine di errore gaussiano indipendente.

Definizione 2.26 (Distribuzione di probabilità – linear Gaussian). Sia Y una variabile continua con genitori continui $X_1, \dots, X_k$, si dice che Y segue un modello Gaussiano lineare se esistono parametri $\beta_0, \dots, \beta_k$ e una varianza σ^2 tali per cui

$$p(Y|x_1, \dots, x_k) = \mathcal{N}\left(\beta_0 + \beta_1 x_k + \cdots + \beta_k x_k;\ \sigma^2\right)$$

Questa formulazione afferma che Y è una funzione lineare delle variabili $X_1, \dots, X_k$, a cui si aggiunge un errore Gaussiano con media zero e varianza σ^2.

Nello specifico,

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

dove ε è una variabile aleatoria Gaussiana con media zero e varianza σ^2, che rappresenta il rumore del sistema. (Koller e Friedman,2009)

L'intero modello è quindi completamente determinato dalla struttura del DAG e dall'insieme dei parametri associati a ciascun nodo, costituiti dai coefficienti della relazione lineare e dalla varianza del rumore. Come nel caso delle reti Bayesiane discrete, la struttura grafica codifica assunzioni di indipendenza condizionata che consentono di fattorizzare la distribuzione normale multivariata in termini di distribuzioni condizionate locali.

Un aspetto fondamentale delle Gaussian Bayesian Networks è che la famiglia delle distribuzioni gaussiane è chiusa rispetto alle operazioni di moltiplicazione, marginalizzazione e condizionamento. Di conseguenza, tutte le distribuzioni che emergono durante il processo di inferenza mantengono una forma gaussiana. Questa proprietà consente di estendere al caso continuo gli algoritmi di inferenza esatta sviluppati per le reti Bayesiane discrete.

Per rendere tali operazioni computazionalmente efficienti, Koller e Friedman propongono una rappresentazione alternativa delle distribuzioni gaussiane, nota come *forma canonica* o *information form*, nella quale una distribuzione normale è espressa mediante una funzione esponenziale quadratica dei parametri. Tale rappresentazione facilita l'applicazione di algoritmi di propagazione dei messaggi e di eliminazione delle variabili, permettendo di effettuare inferenza esatta in modo efficiente per reti di dimensione moderata.

Le Gaussian Bayesian Networks costituiscono pertanto un'estensione naturale delle reti Bayesiane discrete al caso continuo e rappresentano il punto di partenza concettuale per la definizione di modelli più generali in grado di trattare domini misti, come le *Conditional Linear Gaussian Bayesian Networks*, discusse nel paragrafo seguente.

In particolare, nel caso studio presentato in questa tesi verrà adottata una modellazione basata su *Gaussian Bayesian Networks*, poiché la natura continua delle variabili considerate, unitamente alla struttura dei dati e agli obiettivi dell'analisi, rende questo approccio particolarmente efficace per gli scopi del lavoro.

2.6.2 Conditional Gaussian Bayesian Network

Le *Conditional Linear Gaussian Bayesian Networks* (CLG) costituiscono un'estensione delle *Gaussian Bayesian Networks* al caso di domini misti, nei quali coesistono variabili discrete e variabili continue. Tale classe di modelli è introdotta per consentire la rappresentazione congiunta di informazioni qualitative e quantitative all'interno di un unico formalismo probabilistico, mantenendo la struttura a grafo aciclico diretto tipica delle reti Bayesiane.

Definizione 2.27 (Distribuzione di probabilità – Conditional linear Gaussian). Sia X una variabile continua, siano $U = \{U_1, \dots, U_m\}$ *i suoi genitori discreti e siano* $Y = \{Y_1, \dots, Y_k\}$ *i suoi genitori continui. Si dice che X ha una distribuzione condizionata Gaussiana lineare (Conditional Linear Gaussian, CLG) se, per ogni configurazione* $u \in Val(U)$*, esiste un insieme di* $k + 1$ *coefficienti* $a_{u,0}, \dots, a_{u,k}$ *e una varianza* σ_u^2 *tali per cui*

$$p(X|u, y) = \mathcal{N}\left(a_{u,0} + \sum_{i=1}^{k} a_{u,i}\, y_i\, ; \sigma_u^2\right)$$

(Koller e Friedman, 2009)

La definizione evidenzia che, fissata una specifica configurazione delle variabili discrete genitrici, la variabile continua X è modellata come una combinazione lineare dei propri genitori continui, a cui si aggiunge un termine di errore gaussiano con varianza costante per quella configurazione discreta. Ne consegue che ogni configurazione delle variabili discrete seleziona un diverso modello gaussiano lineare per la variabile continua considerata.

Definizione 2.28 (Conditional linear Gaussian Bayesian Network). Una rete Bayesiana è detta rete CLG (Conditional Linear Gaussian) se ogni variabile discreta ha esclusivamente genitori discreti e ogni variabile continua è associata a una distribuzione condizionata Gaussiana lineare (CLG). (Koller e Friedman,2009)

Per quanto riguarda le variabili discrete, le distribuzioni condizionate locali sono definite mediante tabelle di probabilità condizionata, analogamente a quanto avviene nelle reti Bayesiane discrete. Le variabili continue, invece, sono modellate tramite distribuzioni gaussiane condizionate, i cui parametri dipendono dalla configurazione assunta dalle variabili discrete genitrici. In particolare, fissata una configurazione delle variabili discrete, ciascuna variabile continua è descritta da un modello gaussiano lineare in funzione dei propri genitori continui.

Il formalismo CLG impone un vincolo strutturale rilevante: le variabili continue non possono avere figli discreti. Tale restrizione è necessaria per garantire che la distribuzione di probabilità congiunta indotta dalla rete sia ben definita e computazionalmente trattabile. In presenza di tale vincolo, un modello CLG induce una distribuzione di probabilità congiunta che può essere interpretata come una combinazione di distribuzioni gaussiane, ciascuna associata a una specifica configurazione delle variabili discrete della rete e ponderata dalla probabilità di tale configurazione.

Grazie a questa struttura, una volta fissate le variabili discrete, le operazioni di inferenza sulle variabili continue mantengono la chiusura delle distribuzioni gaussiane rispetto alle operazioni di marginalizzazione e condizionamento. Ciò consente di estendere i metodi di inferenza esatta delle reti Bayesiane al caso di domini misti, rendendo le Conditional Linear Gaussian Bayesian Networks particolarmente adatte alla modellazione di sistemi nei quali fattori qualitativi influenzano grandezze quantitative in modo strutturato e interpretabile.

Nel contesto del monitoraggio della continuità aziendale, le *Conditional Linear Gaussian Bayesian Networks* risultano particolarmente adatte in quanto consentono di modellare in modo congiunto variabili qualitative e quantitative tipicamente presenti in tali sistemi. Fattori discreti quali stati operativi o livelli di rischio possono influenzare grandezze continue come costi operativi o indicatori di performance. Il formalismo CLG permette di rappresentare queste relazioni in maniera interpretabile, mantenendo al contempo la possibilità di effettuare

inferenza probabilistica efficiente. Ciò rende tali modelli uno strumento flessibile ed efficace per l'analisi e il supporto decisionale in scenari complessi e dinamici.

3 La continuità aziendale: caso di studio

3.1 Raccolta e costruzione del database

I dati utilizzati per il presente caso di studio sono stati raccolti tramite la banca dati *AIDA* – Analisi Informatizzata delle Aziende Italiane[19].

Si tratta di una banca dati sviluppata originariamente da una società belga[20], oggi parte del gruppo internazionale Moody's Analytics, divisione della Moody's Corporation, una delle cosiddette *"Big Three"* a livello mondiale, leader nei servizi di analisi finanziaria, valutazione del rischio di credito e fornitura di dati economici.

AIDA si occupa di raccogliere i dati economico finanziari delle società di capitali Italiane[21]. Tramite la piattaforma è possibile reperire informazione di natura anagrafica (come la ragione sociale, la natura giuridica o il numero di dipendenti) e contabile tra cui i bilanci d'esercizio e gli indicatori economici e patrimoniali.

Il fatto che le informazioni derivino esclusivamente da società di capitali presenta numerosi vantaggi: da un lato, consente una selezione naturale del campione in termini dimensionali e strutturali; dall'altro, garantisce un elevato grado di affidabilità e qualità dei dati, poiché tali

[19] Bureau van Dijk – Moody's Analytics, *AIDA – Analisi Informatizzata delle Aziende Italiane*, banca dati.

[20] *Bureau van Dijk* è una società fondata a Bruxelles nel 1987, specializzata nella raccolta e diffusione di informazioni economico-finanziarie sulle imprese.

[21] Società per Azioni (S.p.A.); Società in Accomandita per Azioni (S.A.p.A.); Società a Responsabilità Limitata (S.R.L.)

società sono tenute per legge alla redazione, approvazione e deposito del bilancio d'esercizio presso il Registro delle Imprese[22].

La disponibilità di dati economici e finanziari rappresenta un elemento indispensabile e pienamente coerente con gli obiettivi della presente ricerca, che mira ad analizzare il rischio di perdita della continuità aziendale attraverso l'esame degli indicatori di bilancio. Inoltre, la possibilità di disporre congiuntamente di informazioni quantitative e qualitative costituisce un ulteriore punto di forza: consente di ampliare lo spettro dell'analisi, includendo non solo la valutazione di variabili economico-finanziarie (indici e voci di bilancio), ma anche di variabili anagrafiche, come il settore di appartenenza o l'area geografica dell'impresa.

In sintesi, la banca dati *AIDA* rappresenta una fonte informativa di elevata attendibilità e coerenza metodologica, capace di assicurare la qualità e la comparabilità dei dati necessari all'analisi empirica. Le sue caratteristiche strutturali e la natura ufficiale delle informazioni la rendono particolarmente idonea per lo studio del rischio di perdita della continuità aziendale e per l'elaborazione di modelli basati su indicatori di bilancio.

3.1.1 Software di supporto R

Per lo svolgimento delle analisi è stato utilizzato *R*, un linguaggio di programmazione open-source[23] ampiamente impiegato in ambito accademico e professionale per l'elaborazione statistica dei dati e la produzione di rappresentazioni grafiche.

La sua struttura, basata su un vasto ecosistema di pacchetti dedicati, consente di affrontare in modo flessibile e personalizzabile attività di analisi quantitativa, modellizzazione e gestione di database anche di grandi dimensioni.

---

[22] Ai sensi degli artt. 2423 e seguenti del *Codice Civile*, le società di capitali sono tenute alla redazione, approvazione e deposito del bilancio d'esercizio presso il Registro delle Imprese, costituendo così una fonte ufficiale e comparabile di dati economico-finanziari.

[23] Si definisce un linguaggio di programmazione *open-source*, un software il cui codice sorgente è pubblico: chiunque può usarlo e modificarlo.

La scelta di questo software è stata dovuta alla sua capacità di garantire procedure riproducibili, alla solidità degli strumenti messi a disposizione e alla versatilità che offre nell'adattarsi a esigenze analitiche specifiche.

3.2 Pulizia dei dati e selezione delle variabili

Nella fase iniziale di estrazione dei dati dalla piattaforma *AIDA* sono stati creati due campioni distinti.

Il primo campione è composto da 90.000 imprese selezionate casualmente, accomunate dall'unica caratteristica di essere attive e in stato regolare, quindi considerate imprese sane. Il secondo campione, anch'esso costituito da 90.000 osservazioni, include invece imprese con stato giuridico irregolare, ossia soggetti coinvolti in procedure di concordato preventivo, liquidazione giudiziale o altri provvedimenti previsti dalla legge a tutela della continuità aziendale.

La selezione delle variabili da considerare nell'analisi si è basata su studi, analisi e direttive ufficiali nonché legislative analizzate nel primo capitolo di questo elaborato; nell'*Appendice C* si riporta l'elenco delle variabili che sono state scelte durante l'esportazione dei dati tra le numerose a disposizione in AIDA.

I database inizialmente esportati si presentavano quindi come due dataset composti da 90.000 osservazioni e da 164 variabili ciascuno. Il primo è un database caratterizzato da imprese che non hanno problemi di continuità aziendale, che chiameremo d'ora in avanti "`imprese_sane`", il secondo database è composto da imprese caratterizzate da uno stato giuridico che identifica delle problematiche di continuità aziendale, che chiameremo "`imprese_cattive`" o imprese problematiche.

Tutti i dati sono stati successivamente importati nel software di analisi R.

In questa fase, tutte le variabili sono state automaticamente codificate come "*character*"[24], a causa della presenza di identificativi automatici di Excel e di valori "*n.d.*" (non determinabili) e "*n.s.*" (non significativi), utilizzati da *AIDA* per indicare dati mancanti, che sono invece stati interpretati da R come caratteri testuali.

Per poter effettuare correttamente le analisi. È necessario associare a ciascuna variabile la corretta codifica a seconda della natura della variabile. Questo consente di poter poi utilizzare le adeguate tecniche di analisi.

Come primo intervento di pulizia, si è proceduto alla correzione delle date, poiché l'importazione riportava i numeri seriali di Excel anziché le date effettive. Una volta riportata la corretta notazione, si è passati alla ricodifica delle restanti variabili.

Per ottenere una mappatura corretta di tutte le variabili, è stato necessario un lavoro preliminare di categorizzazione. Tramite R, le variabili sono state suddivise in "numeriche pure*"*, "solo testo*"* e "miste*"*, in base alla presenza di caratteri numerici, testuali o di entrambi i tipi. Questa suddivisione si è rivelata indispensabile per individuare la codifica più appropriata.

Successivamente le variabili "numeriche pure" sono state convertite in variabili "*numeric*"[25] e le variabili "solo testo" sono state mantenute come "*character"*.

Il lavoro principale ha riguardato le variabili "miste", nelle quali tutti i dati non determinabili (*n.d.*) o non significativi (*n.s.*) sono stati convertiti in valori *NA*, ovvero dati che il software R riconosce come mancanti e non come caratteri. Successivamente, è stata ripetuta la procedura di categorizzazione per verificare la nuova composizione delle variabili e procedere con l'attribuzione della categoria corretta.

[24] In R, le variabili di tipo *character* rappresentano stringhe di testo, valori non numerici, interpretati come carattere.

[25] In R, le variabili di tipo *numeric* rappresentano valori numerici, interi o decimali. Questo tipo di dato consente di eseguire operazioni aritmetiche.

Come ultima fase del processo di codifica, alcune variabili sono state convertite in tipo "*factor*"[26], tra cui:

- L'indicazione della provincia di appartenenza della sede legale [`provincia`];
- L'indicazione della regione di appartenenza della sede legale [`indirizzo_sede_legale_regione`];
- Lo stato giuridico di ogni singola azienda [`stato_giuridico`];
- La specifica della procedura concorsuale o di cessazione in atto [`procedura_cessazione`].

3.2.1 Gestione dei dati mancanti

In tema di gestione dei dati mancanti (NA), la prima fase consiste nel comprendere la natura dell'assenza dei dati, così da individuarne le cause e classificare i casi secondo le tipologie riconosciute in letteratura: MCAR, MAR e MNAR *(Korb & Nicholson, 2010)*.

I dati *MCAR* (*Missing Completely at Random*) rappresentano i casi in cui l'assenza di dati è del tutto casuale e indipendente da qualunque altra variabile osservata o non osservata. In tali situazioni, la mancanza può interessare un'intera riga di osservazioni, senza alcuna regolarità sottostante. Poiché in questo scenario l'eliminazione delle osservazioni non introduce *bias*[27], si considera come scelta migliore la rimozione delle stesse.

I dati *MAR* (*Missing at Random*) si verificano quando la probabilità che un valore manchi dipende da altre variabili osservate, ma non dal valore stesso mancante. In questo caso è possibile stimare un valore plausibile basandosi sulle informazioni disponibili. Le tecniche di imputazione dei dati mancanti adottabili includono:

[26] In R, le variabili di tipo *factor* rappresentano dati categoriali, ossia variabili che assumono un numero limitato di valori distinti (livelli) e sono utilizzate per gestire i dati qualitativi.

[27] Con il termine *bias* si indica una distorsione sistematica che altera i risultati di un'analisi o di una stima.

- imputazione semplice (media per distribuzioni simmetriche, mediana per distribuzioni asimmetriche, moda per variabili categoriali, oppure approcci *hot-deck*[28] o casuali);
- modelli di regressione;
- metodi basati sui *K-Nearest Neighbors*, che imputano i valori mancanti utilizzando i k casi più simili in termini di distanza euclidea[29].

Infine, i dati *MNAR* (*Missing Not at Random*) rappresentano i casi in cui la mancanza non è casuale e dipende da fattori non osservabili o dallo stesso valore mancante. In tali circostanze non è possibile stimare in modo attendibile i dati mancanti, e l'unica operazione appropriata consiste nel segnalarli e valutarne l'impatto.

Per comprendere quindi a quale categoria i dati *NA* del presente studio facessero parte è stata condotta un'analisi esplorativa quantitativa finalizzata a individuare quante imprese presentassero assenza totale o quasi totale di informazioni.

Nel database delle imprese sane sono state individuate 17.418 osservazioni completamente prive di dati (ossia con tutte le variabili *NA*). Considerata la natura della mancanza e l'impossibilità di effettuare una stima attendibile dei valori, si è deciso di eliminare tali osservazioni dal dataset.

Nel database delle imprese con problemi di continuità aziendale non sono state rilevate osservazioni interamente mancanti. Si è pertanto proceduto con una seconda analisi, volta a individuare le osservazioni con più del 70% di variabili mancanti, anch'esse eliminate, per un totale di 14.358 casi. La scelta deriva dal fatto che, all'aumentare del numero di dati mancanti, risulta più complessa e meno precisa qualsiasi procedura di stima, motivo per cui si è preferito escludere tali osservazioni al fine di preservare l'accuratezza dell'analisi.

Conclusa l'analisi "per osservazioni", è stata avviata una valutazione "per variabili", calcolando la percentuale di valori mancanti per ciascuna di esse. Le variabili sono state ordinate in senso

---

[28] La tecnica di imputazione dei valori mancanti "*hot-deck*" consiste nell'individuare un'osservazione simile completa e sostituire il valore mancante con il valore corrispondente a quello del soggetto simile identificato.

[29] La *distanza euclidea* è una misura di similarità tra due punti in uno spazio multidimensionale, calcolata come la radice quadrata della somma dei quadrati delle differenze tra le rispettive coordinate.

decrescente rispetto alla quantità di dati mancanti e si è fissata una soglia limite di accettazione pari al 40%. Le variabili con una percentuale di valori mancanti superiore a tale limite sono state eliminate, poiché un livello così elevato di assenze avrebbe compromesso l'affidabilità di eventuali procedure di imputazione.

Per le variabili con una percentuale di valori mancanti inferiore al 40% è stata invece condotta un'analisi puntuale, al fine di valutarne l'importanza informativa e stabilire se fosse opportuno mantenerle ed eventualmente stimarle, oppure se potessero essere eliminate senza perdita rilevante di informazione. Ad esempio, si è notato che la maggior parte delle variabili relative al quinto e quarto anno antecedente l'ultimo bilancio disponibile, fossero in particolar modo carenti di informazioni, per questo motivo si è scelto di eliminarle riducendo l'arco temporale di analisi.

Si riporta nell'*Appendice C*, l'elenco delle variabili che sono state mantenute dopo questa fase di analisi.

Al termine di queste procedure, le dimensioni dei due database risultavano le seguenti:

- per quanto riguarda il database di imprese sane: `db_IMP.SANE_def`: 72.418 osservazioni e 82 variabili;
- per quanto riguarda, invece, il database di imprese che hanno riscontrato problematiche di continuità aziendale: `db_IMP.CATTIVE_def`: 75.642 osservazioni e 84 variabili.

La differenza nel numero di variabili tra i due database è dovuta esclusivamente alla presenza, nel dataset delle imprese con problemi di continuità aziendale, di due variabili aggiuntive relative alla procedura concorsuale in atto e alla data di inizio della stessa, ovvero le variabili denominate come:

- `procedura_cessazione`
- `data_di_inizio_procedura_cessazione`

È evidente che tali variabili non possano esistere nel database delle imprese sane. Tutte le restanti sono le medesime in entrambi i database e sono state selezionate incrociando le caratteristiche comuni ai due insiemi.

I database sono ora caratterizzati da un numero leggermente più ridotti di osservazioni e da una riduzione di variabili per coerenza e significatività, sono ancora presenti però piccole percentuali di dati mancanti all'interno di entrambi i database.

3.2.2 Stima valori mancanti

Vista la natura dei dati mancanti residui, classificabili come MAR (*Missing At Random)*, per il processo di stima dei valori mancanti si è ritenuto opportuno adottare un modello statistico di tipo predittivo. A tal fine è stato scelto il pacchetto "*missRanger"* di R, basato sull'algoritmo Random Forest (*Stekhoven, Daniel J., and Peter Buehlmann, 2011; Wright, Marvin, and Andreas Ziegler, 2017*).

Il modello analizza congiuntamente tutte le variabili del dataset, trattando ogni colonna contenente valori mancanti come variabile dipendente e costruendo, per ciascuna di esse, un modello di regressione Random Forest che utilizza le restanti variabili come predittori. Per ogni dato mancante, il modello stima una misura plausibile sulla base dei pattern esaminati nelle altre osservazioni.

Il Random Forest è un modello di apprendimento supervisionato basato su un insieme di alberi decisionali. Ogni albero, costruito su un sottoinsieme casuale di osservazioni e variabili, genera una previsione per la variabile in esame; al termine del processo, il modello aggrega le previsioni di tutti gli alberi producendo un output medio o di voto maggioritario. Si parla di modello supervisionato poiché, durante la fase di addestramento, l'algoritmo apprende relazioni da dati per i quali la risposta è nota *(Breiman, 2001)*.

Questo approccio si rivela concettualmente coerente con il modello Bayesiano adottato nella parte successiva dello studio, poiché entrambi i modelli condividono una logica predittiva e probabilistica: entrambi apprendono dai dati e aggiornano le proprie previsioni in funzione delle nuove evidenze. Si differenziano per il fatto che mentre il modello Bayesiano opera in termini di probabilità aggiornate, il Random Forest stima la probabilità di un determinato risultato sulla base delle combinazioni di variabili osservate, aggregando le stime di una moltitudine di modelli decisionali elementari.

Tuttavia, l'applicazione di tale metodo si è rivelata impraticabile a causa delle limitazioni di memoria RAM del dispositivo utilizzato e dell'elevato volume di dati.

Si è pertanto tentato un secondo approccio mediante l'algoritmo *K-Nearest Neighbors* (KNN), basato sul calcolo della distanza euclidea tra osservazioni *(Cover & Hart, 1967).* Tale metodo sostituisce i valori mancanti con la media dei k casi più simili, preservando la coerenza statistica e la distribuzione originaria dei dati. La logica sottostante risulta anch'essa coerente con l'approccio Bayesiano, in quanto entrambi i metodi stimano il valore più probabile di una variabile in funzione delle evidenze osservate e delle relazioni condizionali presenti nel campione. Anche in questo caso, tuttavia, la procedura è risultata eccessivamente lenta e quindi non applicabile a un dataset di tali dimensioni.

Si è scelto quindi di procedere con una strategia di campionamento, selezionando inizialmente le 50.000 osservazioni con la minor percentuale di valori mancanti. Successivamente è stato nuovamente applicato l'algoritmo Random Forest, ma le difficoltà computazionali sono rimaste significative, in particolare a causa della presenza di variabili categoriali. Si è dunque ridotto il campione a 30.000 osservazioni e si è limitata l'analisi alle sole variabili contenenti gli indici di bilancio utili ai successivi studi. Inoltre, per alleggerire il calcolo, è stata disattivata la funzione di *predictive mean matching*, che confronta i valori stimati con i k valori più simili presenti nel dataset: in questo modo il modello ha sostituito i valori mancanti direttamente con i valori predetti, riducendo i tempi di elaborazione, la nota negativa di questa decisione risiede nel fatto che questa funzione serve per mantenere la distribuzione originaria dei dati, così facendo i valori stimati risultano più artificiali.

Un altro intervento adottato per mantenere il modello leggero è stato quello di limitare il numero di alberi generati a 100. In tali condizioni il modello ha mostrato stabilità e tempi di esecuzione accettabili. Successivamente si è reintrodotta la funzione *predictive mean matching* con un valore di *pmm.k = 1*, in modo da mantenere un compromesso tra leggerezza computazionale e accuratezza di stima: il modello, in questo caso, sceglie il valore esistente nel dataset più simile al valore predetto.

Il modello ha retto anche su un campione di 50.000 osservazioni, consentendo successivamente di reintegrare le variabili precedentemente escluse e di ottenere due database completi:

- per quanto riguarda il database di imprese sane: `db_IMP.SANE.comp`: 50.000 osservazioni e 82 variabili;

- per quanto riguarda, invece, il database di imprese che hanno riscontrato problematiche di continuità aziendale: `db_IMP.CATTIVE.comp`: 50.000 osservazioni e 84 variabili.

Per le imprese sane, la percentuale di valori mancanti residui risultava pari allo 0,0083%, imputabile principalmente a variabili geografiche; mentre per le imprese con problemi di continuità aziendale, la percentuale era dello 0,087%, sempre imputabili a caratteristiche anagrafiche delle imprese.
Sono state infine eliminate tutte le osservazioni contenenti valori mancanti nelle variabili descrittive quali oggetto sociale, regione o provincia di appartenenza, ottenendo due database completamente privi di valori mancanti:

- database di imprese sane: `db_IMP.SANE.na.omit`: 49.675 osservazioni e 82 variabili;
- database di imprese che hanno riscontrato problematiche di continuità aziendale: `db_IMP.CATTIVE.na.omit`: 46.379 osservazioni e 84 variabili.

3.3 Analisi descrittiva delle variabili

In questa fase di analisi preliminare verranno esaminate in modo approfondito tutte le variabili selezionate, con l'obiettivo di comprenderne la struttura, la composizione interna e le possibili relazioni.
L'analisi consentirà inoltre di orientare in modo più consapevole le fasi successive dello studio.

3.3.1 Area Geografica

La prima analisi condotta è quella relativa all'area geografica, per comprendere se questa variabile influisse in qualche modo sulla salute dell'impresa, condizionando o meno il verificarsi di problematiche legate alla continuità aziendale.

Per prima cosa si è scelto di categorizzare tutte le osservazioni per aree geografiche Italiane: Nord, Centro e Sud. Le isole sono state inserite nell'area geografica Sud Italia.

Si sono quindi ottenute le seguenti suddivisioni.

	Imprese Sane "db.IMP.SANE.na.omit"	*%*	Imprese con problematiche di continuità aziendale "db.IMP.CATTIVE.na.omit"	*%*	Totale	*%*
Nord	22.812	*23,75 %*	32.175	*33,50 %*	54.987	*57,25 %*
Centro	13.596	*14,15 %*	8.409	*8,75%*	22.005	*22,91 %*
Sud	13.267	*13,81 %*	5.795	*6,03%*	19.062	*19,85 %*
Totale	*49.675*	*51,72 %*	*46.379*	*48,28 %*	*96.054*	*100%*

Tabella 3.1: Frequenze della distribuzione delle osservazioni per area geografica. Fonte: elaborazione personale su dati Aida

Da una prima analisi si può osservare come la maggior parte delle imprese in genarle sia situata nell'area del Nord Italia, pari al 57,25%: questo dato è coerente vista la maggior

industrializzazione di questa area, come riportato dall'Istat (2024). Il Nord-ovest concentra la quota più elevata di imprese e addetti dell'industria italiana (28,9% e 34,6%)[30].

Possiamo osservare che nell'area del Nord Italia c'è una preponderanza di aziende con problematiche (33,05 %), rispetto ad aziende sane (23,75 %), con un gap di 9,3 punti percentuali.
Nell'area del Centro e del Sud invece, le aziende sane caratterizzano la maggioranza con degli scostamenti rispettivamente di 5,4 (centro) e 7,78 (sud) punti percentuali rispetto alle imprese con problematiche nella stessa area. Si può quindi concludere che l'area del Centro Italia presenti la maggiore omogeneità nella distribuzione tra imprese sane e imprese con problematiche di continuità aziendale.

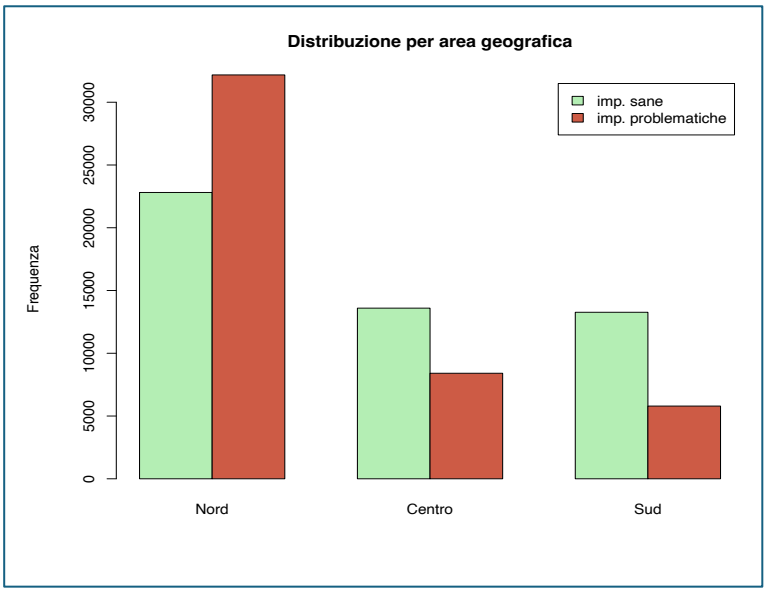


Figura 3.1: Bar Plot della distribuzione delle osservazioni per area geografica. Fonte: elaborazione personale su dati AIDA

Dai dati emerge una relazione potenziale tra numerosità delle aziende e continuità d'impresa: le aree caratterizzate da una maggiore concentrazione di imprese mostrano una più elevata incidenza di criticità legate alla continuità aziendale, mentre al diminuire della numerosità produttiva prevalgono le imprese in stato giuridico sano.

[30] Istat, "Rapporto Annuale 2024 – La situazione del Paese", Capitolo 14, "Struttura e dinamica delle imprese industriali e dei servizi", Roma, Istituto Nazionale di Statistica, 2024, p. 567.

Da una ricerca più approfondita condotta secondo la provincia di appartenenza dell'impresa, si può osservare un potenziale trend che conferma quanto sopra riportato; infatti, per le province che presentano una numerosità particolarmente elevata, superiore alle 1.000 aziende per provincia, è più frequente avere aziende con problematiche di continuità aziendale.

È stato scelto il valore di 1.000 unità in quanto si è osservato che fosse il punto di discontinuità in tutte le situazioni osservate, ovvero il punto nel quale le tendenze tipiche dei dati si invertissero.

Si riportano in *Appendice D* tutti i dati estratti relativi alla distribuzione geografica. Dalle analisi, si può osservare quanto di seguito riportato.

Ragionando sull'intero territorio nazionale emerge che le province con una numerosità superiore alle 1.000 imprese sono 23: tra queste, 13 presentano una maggioranza di imprese con problemi di continuità aziendale. Ne deriva una probabilità del 56,52% che una provincia con più di 1.000 imprese abbia una maggioranza di aziende con problematiche di continuità, contro una probabilità del 43,48% che la maggioranza delle imprese risulti sana.

Le province con una numerosità inferiore alle 1.000 imprese sono invece 88: di queste, 19 presentano una maggioranza di aziende che hanno riscontrato problematiche economico-finanziarie, mentre 69 presentano una maggioranza di imprese sane. Pertanto, una provincia con meno di 1.000 imprese ha una probabilità del 78,41% di essere caratterizzata da una prevalenza di imprese sane, contro una probabilità del 21,59% che la maggioranza delle sue aziende presenti problematiche di continuità aziendale.

Questi risultati confermano che le aree geografiche con una maggiore concentrazione di imprese tendono a presentare un rischio più elevato di perdita della continuità aziendale, dato dalla maggiore probabilità che la prevalenza delle imprese riscontri questo tipo di problema. Tale fenomeno può essere ricondotto al maggiore grado di competitività che caratterizza i mercati più popolati, dove l'elevato numero di operatori economici può accentuare le difficoltà nel mantenere condizioni di stabilità e solidità nel medio periodo.

Le percentuali riportate rappresentano probabilità empiriche derivate dalla distribuzione osservata nel campione e non costituiscono risultati previsionali. Esse vanno pertanto

interpretate esclusivamente come indicazioni descrittive sul comportamento generale della variabile.

3.3.2 Codice ATECO

Il codice ATECO permette la suddivisione delle imprese in base all'attività economica che svolgono. È usato dall'Agenzia delle Entrate per scopi contributivi e dall'ISTAT per ricerche statistiche.

Il codice ATECO è un codice alfanumerico composto da una lettera che identifica il macrosettore e sei cifre, ciascuna corrispondente a un'attività sempre più specifica.

L'analisi si vuole svolgere sulla suddivisione settoriale delle aziende per comprendere se ci siano settori più rischiosi e con una maggior probabilità di riscontrare problematiche economiche.

Inizialmente si svolge una categorizzazione attraverso le prime due cifre dei codici ATECO, utilizzando il criterio ufficiale fornito dalle linee guida dall'ISTAT[31], di seguito riportato.

Sezione	**Divisori (prime 2 cifre)**	**Settore ufficiale**
A	01–03	*Agricoltura, silvicoltura e pesca*
B	05–09	*Attività estrattiva*
C	10–33	*Attività manifatturiere*
D	35	*Fornitura di energia elettrica, gas, vapore e aria condizionata*
E	36–39	*Fornitura di acqua; reti fognarie, attività di trattamento dei rifiuti e risanamento*
F	41–43	*Costruzioni*
G	45–47	*Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli*
H	49–53	*Trasporti e magazzinaggio*
I	55–56	*Servizi di alloggio e ristorazione*
J	58–63	*Informazione e comunicazione*
K	64–66	*Attività finanziarie e assicurative*
L	68	*Attività immobiliari*

[31] ISTAT, Prospetto 5.2, Aggregazione intermedia SNA/ISIC a 38 categorie (a), in *Classificazione delle attività economiche – ATECO 2007*, Metodi e Norme n. 40, Roma, 2009, p. 45.

M	69–75	*Professioni tecniche e scientifiche*
N	77–82	*Attività amministrative e di servizi di supporto*
O	84	*Amministrazione pubblica e difesa; assicurazione sociale obbligatoria*
P	85	*Istruzione*
Q	86–88	*Sanità*
R	89–93	*Attività artistiche, di intrattenimento e divertimento*
S	94–96	*Altre attività di servizi*
T	97–98	*Attività di famiglie e convivenze come datori di lavoro per personale domestico; produzione di beni e servizi indifferenziati per uso proprio da parte di famiglie e convivenze*
U	99	*Attività di organizzazioni e organismi extraterritoriali*

Tabella 3.2: Suddivisione settoriale codici ATECO. Fonte: ISTAT (2009)

Grazie a questa suddivisione possiamo rappresentare graficamente la distribuzione del campione usato per l'analisi e comprendere se lo sviluppo di problematiche legate alla continuità aziendale può essere legato a settori specifici.

Nell'*Appendice E* si riporta la tabella comprensiva delle relative frequenze utilizzate per la costruzione del grafico in Figura 3.2 – *Grafico Bar Plot della distribuzione delle osservazioni per settore economico*. Le frequenze sono state calcolate sul campione in esame rispettivamente per l'insieme di imprese sane e per l'insieme di imprese che hanno sviluppato problematiche di continuità aziendale.

Da una prima analisi del grafico in Figura 3.2 non sembrano emergere tendenze specifiche nello sviluppo di problematiche in base al settore di appartenenza. Tuttavia, si osserva una maggiore propensione allo sviluppo di difficoltà nei settori con un alto numero di imprese, mentre i settori meno popolati tendono a riportare una percentuale più elevata di imprese sane.

In particolare, nei settori con più di 10.000 imprese si rileva una probabilità del 25% che il settore mantenga una prevalenza di imprese sane, contro una probabilità del 75% che presenti una maggioranza di imprese con problematiche di continuità aziendale. Al contrario, nei settori caratterizzati da meno di 10.000 imprese, la probabilità che prevalgano le imprese sane è pari al 71,42%, mentre la probabilità che si sviluppi una maggioranza di imprese con problematiche di continuità aziendale è del 28,57%.

Il limite di 10.000 imprese è stato, anche in questo caso, scelto come indicatore di discontinuità delle tendenze osservate, quindi come punto nel quale si sono osservati dei cambiamenti nelle tendenze dei dati.

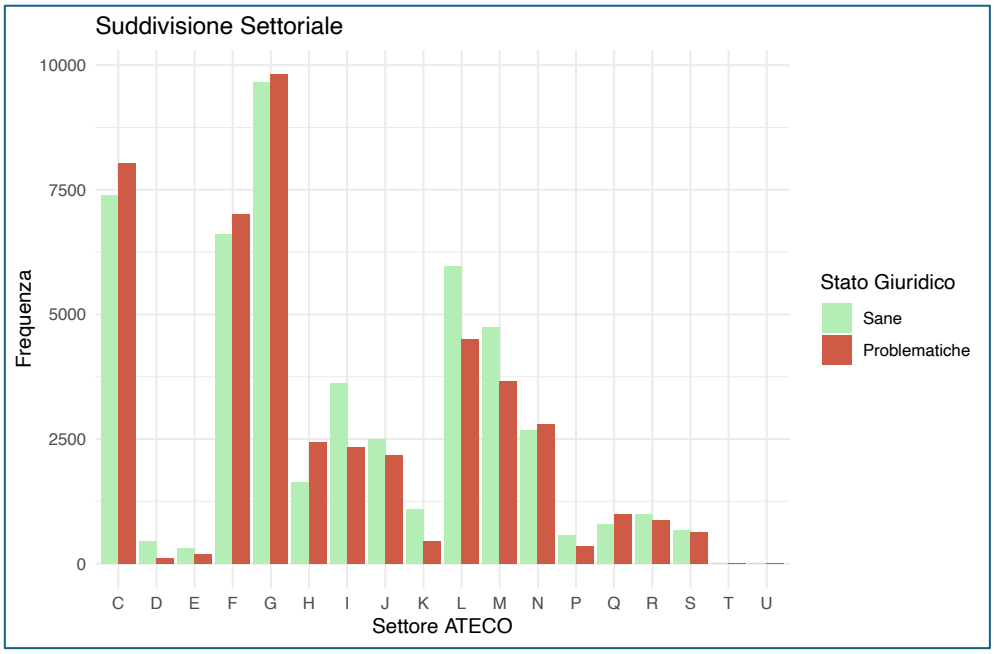


Figura 3.2: Bar Plot della distribuzione delle osservazioni per codice ATECO. Fonte: elaborazione personale su dati AIDA

Queste osservazioni sono coerenti e si confermano con quanto analizzato per le aree geografiche. Nei settori caratterizzati da una maggiore numerosità di imprese, si osserva più frequentemente una prevalenza di aziende con problematiche di continuità. Tuttavia, tale evidenza non consente di affermare un rapporto di causalità diretta tra il settore di appartenenza e la salute dell'impresa. L'elevata eterogeneità nella numerosità delle imprese tra i settori rende complessa un'interpretazione causale delle differenze osservate.

3.3.3 Stato giuridico e Procedure di Cessazione

In questo paragrafo vengono analizzate le due variabili esclusive che caratterizzano il database delle imprese con problematiche di continuità aziendale, per questo motivo i dati analizzati apparterranno solamente ad uno dei due insiemi. Tali variabili descrivono la condizione giuridica e operativa delle imprese.

Per comprendere appieno le dinamiche che caratterizzano ciascuna azienda, l'analisi deve essere condotta in maniera congiunta tra la variabile dello stato giuridico e la variabile della

procedura di cessazione, poiché la loro combinazione consente di individuare le differenti situazioni in cui può trovarsi un'impresa in relazione alla continuità aziendale.

In primo luogo, viene analizzata la variabile dello stato giuridico, al fine di valutare la distribuzione delle imprese del campione rispetto alla loro condizione formale. La distribuzione di frequenza dello stato giuridico è rappresentata mediante un grafico di tipo Bar Plot in Figura 3.3 mentre la relativa tabella di frequenza completa è riportata in *Appendice F*.

Come mostrato dal grafico in Figura 3.3, riportato in seguito, la maggior parte delle imprese risulta cessata, si evidenzia quindi un'ulteriore distinzione per liquidazione (anche per procedura di liquidazione in corso) o per fallimento.

Sono presenti inoltre imprese attive, suddivise ulteriormente in caso di procedure di insolvenza in corso o in stato di insolvenza.

È importante sottolineare che le imprese classificate come "Attive" sono incluse nel database delle imprese con problematiche di continuità aziendale, poiché, pur risultando regolarmente attive, presentano procedure concorsuali aperte o altri elementi di rischio dovuti a problematiche di continuità aziendale.

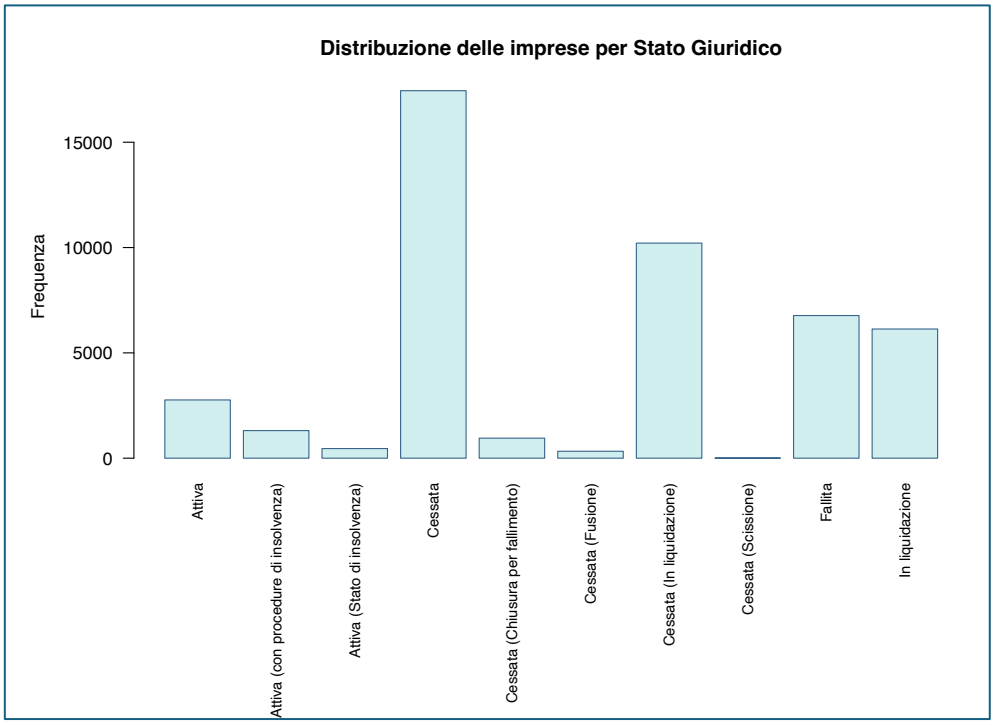


Figura 3.3: Bar Plot della distribuzione delle osservazioni per Stato giuridico. Fonte: elaborazione personale su dati AIDA

Andremo ora ad analizzare le procedure concorsuali o di cessazione relative nello specifico agli stati giuridici più popolati: In liquidazione, Attive, Cessata, Cessata (In liquidazione) e Fallita.

Le tabelle complete delle frequenze dei dati sono riportate in *Appendice F.*

Impresa In liquidazione:

Le imprese categorizzate come "In liquidazione" sono soggette a liquidazione volontaria nel 73,99% dei casi.

Successivamente, è possibile individuare un 15,83% di imprese fallite sottoposte a liquidazione coatta amministrativa, ovvero una procedura concorsuale di natura amministrativa in cui la gestione della liquidazione aziendale viene affidata a un commissario o curatore nominato dall'autorità competente.

Si rilevano inoltre altre procedure di tipo giudiziario, come la liquidazione giudiziale e lo scioglimento per atto dell'autorità, che indicano un intervento diretto dell'autorità giudiziaria in seguito all'incapacità dell'impresa di fronteggiare autonomamente i propri impegni economico-finanziari.

Impresa Attiva:

Per le imprese che risultano attualmente attive si osserva una maggioranza, pari al 48,35%, di imprese che presentano una procedura concorsuale di liquidazione volontaria.

Il 22,17% delle imprese è invece soggetto a concordato preventivo, ovvero come presentato nel primo capitolo, una procedura giudiziale che consente all'impresa in difficoltà di proporre ai creditori un piano di ristrutturazione per evitare la liquidazione o il fallimento, perseguendo la continuità aziendale.

Sono inoltre presenti altre procedure, presentate nel primo capitolo:

- Amministrazione giudiziaria (2,89%): misura disposta dal tribunale in caso di gravi irregolarità gestionali, con la nomina di un amministratore che sostituisce temporaneamente gli organi societari;
- Concordato fallimentare (1,77%): accordo tra il debitore e i creditori che interviene nel corso di una procedura fallimentare già aperta, al fine di chiuderla anticipatamente;

- Accordo di ristrutturazione dei debiti (0,58%): intesa per il riequilibrio della situazione economico-finanziaria;
- Stato di insolvenza (0,29%): condizione in cui l'impresa non è più in grado di adempiere regolarmente alle proprie obbligazioni;
- Amministrazione straordinaria (0,11%): procedura speciale riservata alle grandi imprese in crisi, finalizzata alla conservazione e ristrutturazione del complesso aziendale sotto la vigilanza del Ministero delle Imprese.

Imprese Cessate, Cessate (In Liquidazione) e Fallite:

Per le imprese che hanno cessato la propria attività, nella maggior parte dei casi non viene riportata la motivazione che ha determinato la cessazione, bensì soltanto l'atto burocratico conclusivo.
In particolare, nel 63,03% dei casi la procedura di cessazione indicata è la cancellazione dal registro delle imprese, ossia l'ultimo adempimento amministrativo che sancisce la chiusura definitiva dell'azienda.
Pertanto, in tali situazioni non è possibile risalire con certezza alla causa effettiva della cessazione.

Lo stesso discorso vale per le imprese classificate come *cessate in liquidazione*: in questi casi è possibile individuare soltanto l'ultima procedura depositata, ad esempio "chiusura della liquidazione" se la procedura risulta conclusa, oppure "chiusura per fallimento o liquidazione" qualora l'attività sia cessata a seguito di una procedura concorsuale.

Altre denominazioni, come "scioglimento e messa in liquidazione", indicano invece che la società è stata sciolta e che i beni aziendali sono stati posti in liquidazione per estinguere le passività residue.

Le imprese che risultano *fallite* sono tutte quelle imprese per cui è stata emessa una sentenza di fallimento da parte di un organo giuridico e anche in questo caso ciò che viene riportato come procedura è l'ultima attività depositata, quindi per un 36,36% dei casi si tratta di chiusura del fallimento o ancora cancellazione dal registro delle imprese.

Per questa porzione del campione non è possibile individuare la causa che ha determinato la cessazione dell'attività, poiché tale informazione non è riportata nel database.

Tuttavia, queste imprese verranno comunque utilizzate per l'addestramento del modello come osservazioni negative, in quanto la cessazione dell'attività implica, in ogni caso, la presenza di situazioni problematiche.

3.3.4 Numero dipendenti – ultimo anno disponibile

La variabile rappresentativa del numero di dipendenti nell'ultimo anno disponibile presenta le seguenti statistiche descrittive:

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	0	0	2	10,92	6	12.892
Imprese problematiche	0	0	1	12,38	6	30.995

Tabella 3.3: Statistiche descrittive della variabile "numero di dipendenti". Fonte: elaborazione personale su dati AIDA

Nel caso delle imprese problematiche, i valori medi relativamente bassi possono essere spiegati dal fatto che, nell'ultimo anno disponibile, molte di queste risultano cessate o fallite. Poiché è ragionevole osservare che il momento della chiusura definitiva dell'azienda coincida con l'assenza di personale, è comprensibile che in numerosi casi si registri un numero nullo di dipendenti. Anche per le imprese sane, la distribuzione dei dati mostra una forte concentrazione di valori pari a 0 o 1: 23.781 osservazioni, corrispondenti al 47,87% del totale. Questa elevata incidenza di microimprese o di imprese con personale assente o minimo contribuisce a mantenere la media complessiva molto bassa e una distribuzione fortemente asimmetrica.

In teoria, la variabile "numero di dipendenti" avrebbe potuto fornire informazioni utili sulle dimensioni aziendali. Tuttavia, la presenza di un'elevata quota di valori incoerenti o poco rappresentativi, in particolare per le imprese sane, ne riduce la significatività e l'utilità ai fini dell'analisi.

3.3.5 Indici e Voci di bilancio - ultimo anno disponibile

In questo paragrafo vengono analizzate alcune variabili disponibili esclusivamente per l'ultimo esercizio. Per gli anni precedenti (anno -1, -2 e -3) tali informazioni risultavano troppo incomplete e sono state pertanto eliminate.

Le variabili verranno analizzate grazie al supporto di grafici Box Plot.

Ricavi di Vendita

L'analisi di questa variabile porta a comprendere se il livello dei ricavi di vendita si differenzi tra le imprese che hanno riscontrato problematiche di continuità aziendale e le imprese invece che risultano regolarmente operative.

I dati si riferiscono ai ricavi di vendita registrati nell'ultimo esercizio disponibile nel database, e sono espressi in migliaia di euro.

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	-7	59	255	3.759	953	8.700.014
Imprese problematiche	23	74	244	2.087	779	6.026.236

Tabella 3.4: Statistiche descrittive della variabile "ricavi di vendita". Fonte: elaborazione personale su dati AIDA

Da una prima analisi dei dati, si evince immediatamente che:

- La mediana delle imprese sane (255) è lievemente superiore rispetto a quella delle imprese problematiche (244).
  La media è nettamente più elevata per le imprese sane (3.759) rispetto a quelle problematiche (2.087).
- Il *range* interquartile[32] risulta maggiore per le imprese sane (894) rispetto a quelle problematiche (705), indicando una maggiore dispersione dei ricavi per le imprese sane. Inoltre, il valore massimo degli *outlier* è sensibilmente più elevato nelle imprese sane, come si riflette nella media più alta.
- *Anomalie***:** Si osserva che il valore minimo registrato per le imprese sane è negativo (−7).

Poiché la distribuzione risulta fortemente asimmetrica e caratterizzata da un'elevata varianza, la rappresentazione grafica tramite box plot è stata omessa in quanto risultava poco leggibile, con una compressione della scatola verso la base del grafico.

[32] Si definisce come *range interquartile* la differenza tra il terzo quartile e primo quartile, identifica la distribuzione del 50% dei valori centrali.

L'indicatore di un'elevata varianza è dato dal calcolo della deviazione standard sulla variabile ricavi di tutte le osservazioni dell'intero database (imprese sane e problematiche congiunte), questo calcolo ha portato alla conclusione che in media il valore dei ricavi di vendita si discosta di 57.674,46 migliaia di euro dalla media.

Per ovviare a questo problema, si è applicata una trasformazione logaritmica ai dati, che consente di ridurre la varianza e l'influenza degli *outlier*, rendendo la distribuzione più interpretabile.

Nell'applicazione della trasformazione logaritmica si sono esclusi i valori negativi, in quanto la loro trasformazione avrebbe generato dati *NA*, e di conseguenza un'alterazione dell'andamento centrale, con un'inversione fittizia della tendenza della media. Comportando così un livello della media delle imprese sane, inferiore alla media delle imprese con problematiche di continuità aziendale, rappresentazione non veritiera dei dati originari.

Di seguito si riportano le statistiche descrittive calcolate sui dati trasformati:

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	-6,908	4,590	5,785	5,788	7,006	15,979
Imprese problematiche	3,133	4,308	5,499	5,624	6,658	15,612

Tabella 3.5: Statistiche descrittive della variabile "numero di dipendenti" con trasformazione logaritmica. Fonte: elaborazione personale su dati AIDA

In seguito, si riporta il grafico Box Plot in Figura 3.4 dove i dati relativi alle imprese sane sono rappresentati con la scatola colorata in verde, mentre i dati delle imprese che hanno sviluppato problematiche di continuità aziendale sono rappresentati nella scatola colorata in rosso.

I dati confermano l'aspettativa iniziale: La mediana (logaritmo) delle imprese sane (5,788) è superiore a quella delle imprese problematiche (5,499). Allo stesso modo, la media (logaritmo) delle imprese sane (5,738) è superiore a quella delle imprese problematiche (5,624).

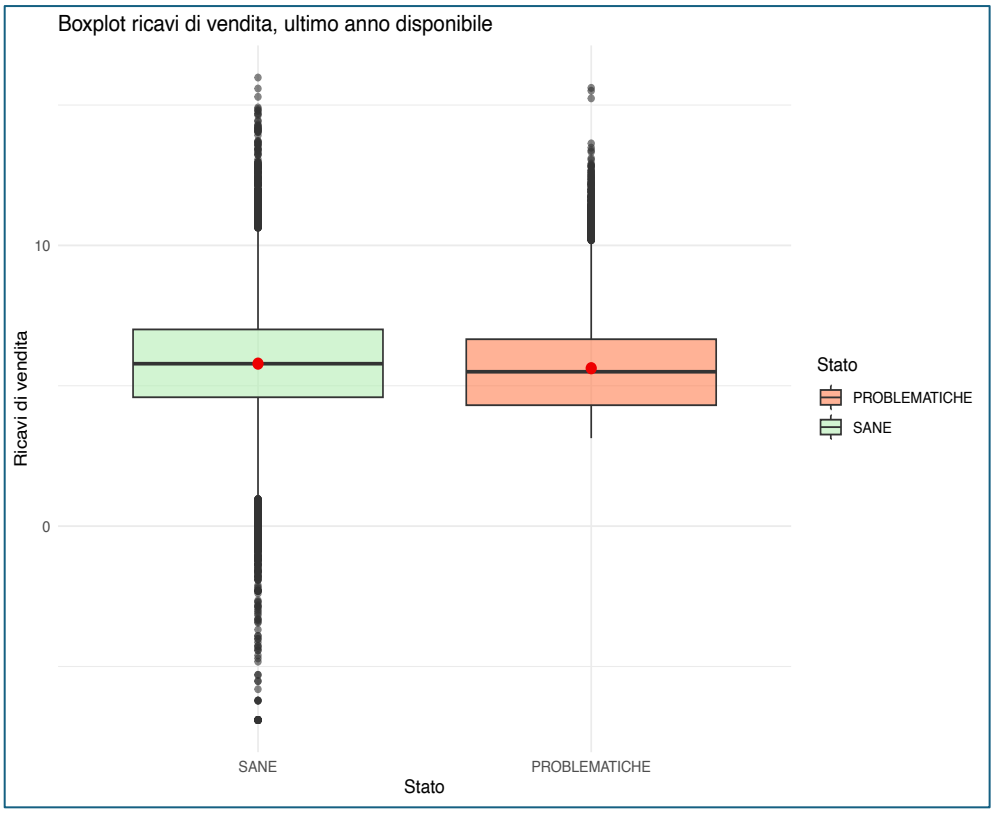


Figura 3.4: Box Plot della variabile "ricavi di vendita" (trasformazione log); confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA

La trasformazione logaritmica ha consentito di ridurre l'influenza dei valori estremi e di evidenziare, anche visivamente, come le imprese sane presentino ricavi mediamente superiori rispetto a quelle con problematiche di continuità aziendale.

Margine di tesoreria

Il margine di tesoreria è utilizzato per valutare la capacità di un'impresa di far fronte ai debiti a breve termine attraverso le proprie disponibilità liquide. Si tratta di un indicatore importante per analizzare la solidità finanziaria nel breve periodo.

Nei dati analizzati, il margine di tesoreria è espresso in migliaia di euro e viene calcolato attraverso la differenza tra liquidità e passività. Il margine di tesoreria deve essere, quindi, positivo affinché la situazione sia stabile.

Si riporta la tabella con le statistiche descrittive relative all'indicatore analizzato:

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	-13.334.546	-102	6.5	-522,1	106,1	470.393,4
Imprese problematiche	-2.922.466	-342,1	- 41,6	-1.353,7	19	251.970,3

Tabella 3.6: Statistiche descrittive della variabile "margine di tesoreria". Fonte: elaborazione personale su dati AIDA

Dal grafico Box Plot in Figura 3.5 si osserva che le imprese sane presentano una distribuzione dei valori prevalentemente intorno allo zero, con una media tuttavia significativamente negativa. Questo fenomeno è dovuto alla presenza di valori estremi molto elevati, che influenzano la media. Per tale motivo è preferibile considerare la mediana, meno sensibile ai valori anomali: come riportato nella tabella dei dati, essa risulta positiva.

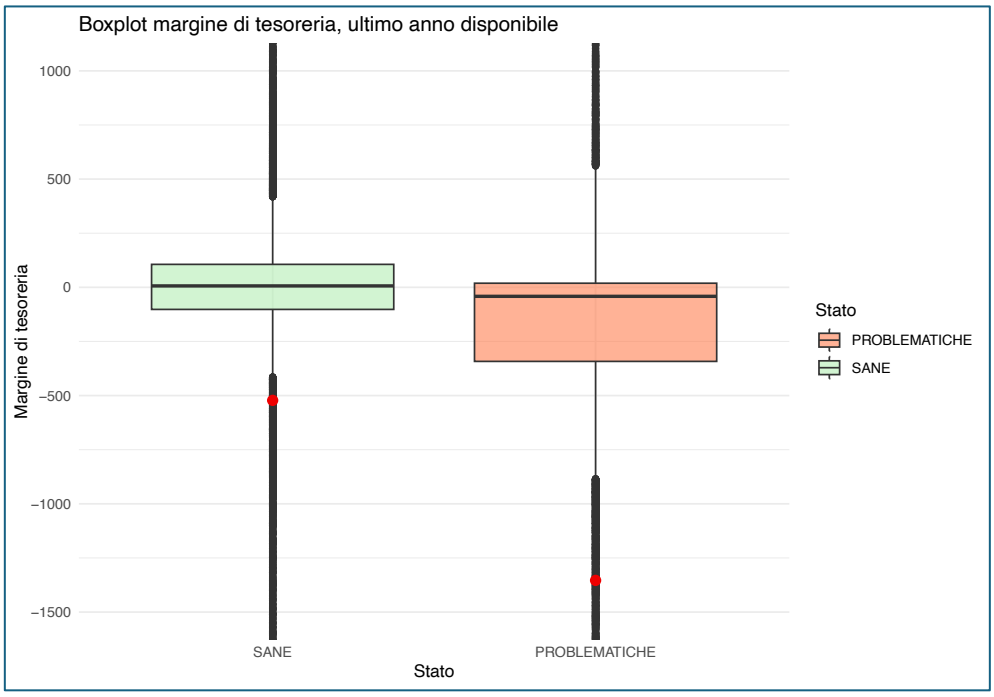


Figura 3.5: Box Plot della variabile "margine di tesoreria"; confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA

Le imprese che presentano problematiche di continuità aziendale mostrano invece, come evidenziato dal grafico, una distribuzione di valori generalmente inferiore rispetto a quella delle imprese sane e prevalentemente negativa, sia per quanto riguarda la media sia per quanto riguarda la mediana.

L'analisi conferma quindi che, quanto più il margine di tesoreria è basso, tanto più ci si avvicina alle tendenze che caratterizzano le imprese che hanno riscontrato problematiche di continuità aziendale.

Si sottolinea che, come per le analisi precedenti, queste osservazioni sono fatte meramente su analisi descrittive, non si può pertanto definire un nesso di causalità certo a livello statistico, indagine che verrà fatta successivamente nel corso dello studio.

Redditività delle vendite - ROS

Il ROS (*Return on Sales*) è l'indicatore che misura il margine residuo dei ricavi dopo la copertura dei costi operativi. Esso consente di comprendere quanto margine operativo rimane per ogni euro di vendita, rappresentando quindi uno strumento utile per monitorare la redditività operativa dell'impresa. Il ROS è sempre espresso in termini percentuali e, in generale, un valore più elevato indica una maggiore solidità aziendale.

Come evidenziato nel grafico riportato in Figura 3.6, le imprese in buona salute mostrano una distribuzione del ROS positiva e più alta rispetto a quelle che hanno presentato problematiche di continuità aziendale.

Ciò dimostra che le aziende sane si distinguono per un margine operativo positivo ed elevato.

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	-49,98	-0,25	4,60	2,34	13,55	29,99
Imprese problematiche	-49,99	-32,05	-4,35	-10,84	4,67	30,00

Tabella 3.7: Statistiche descrittive della variabile "ROS". Fonte: elaborazione personale su dati AIDA

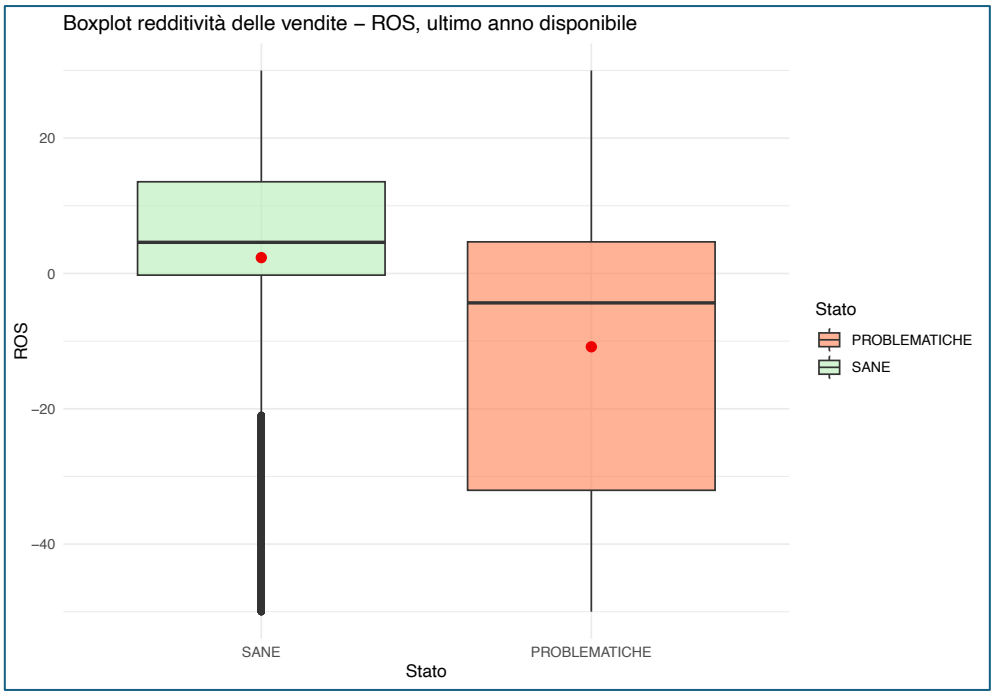


Figura 3.6: Box Plot della variabile "ROS"; confronto imprese sane e problematiche.
Fonte: elaborazione personale su dati AIDA

Posizione finanziaria netta

La posizione finanziaria netta è un indicatore che misura la solidità finanziaria di un'azienda. Viene calcolata come differenza tra passività finanziarie e attività finanziarie, escludendo quindi debiti e crediti di natura commerciale. Si tratta di un indicatore particolarmente rilevante, poiché consente di valutare la capacità di un'azienda di far fronte ai propri impegni finanziari. Valori elevati, se non adeguatamente giustificati, rappresentano potenziali segnali di allarme.

L'indicatore è espresso in migliaia di euro e, come evidenziato dalla formula, un valore più elevato indica una maggiore esposizione debitoria, in quanto i debiti finanziari superano le disponibilità liquide. Questo andamento è visibile anche nel grafico in Figura 3.7, dove la distribuzione delle imprese in difficoltà si concentra su valori più alti rispetto a quella delle imprese sane.

	Min.	1° Quartile	Mediana	Media	3°quartile	Max.
Imprese sane	-465.456	-75,6	-5,8	244,7	55,4	993.259
Imprese problematiche	-256.349,6	-20,4	-0,3	810,5	125,7	872.358,7

Tabella 3.8: Statistiche descrittive della variabile "posizione finanziaria netta". Fonte: elaborazione personale su dati AIDA

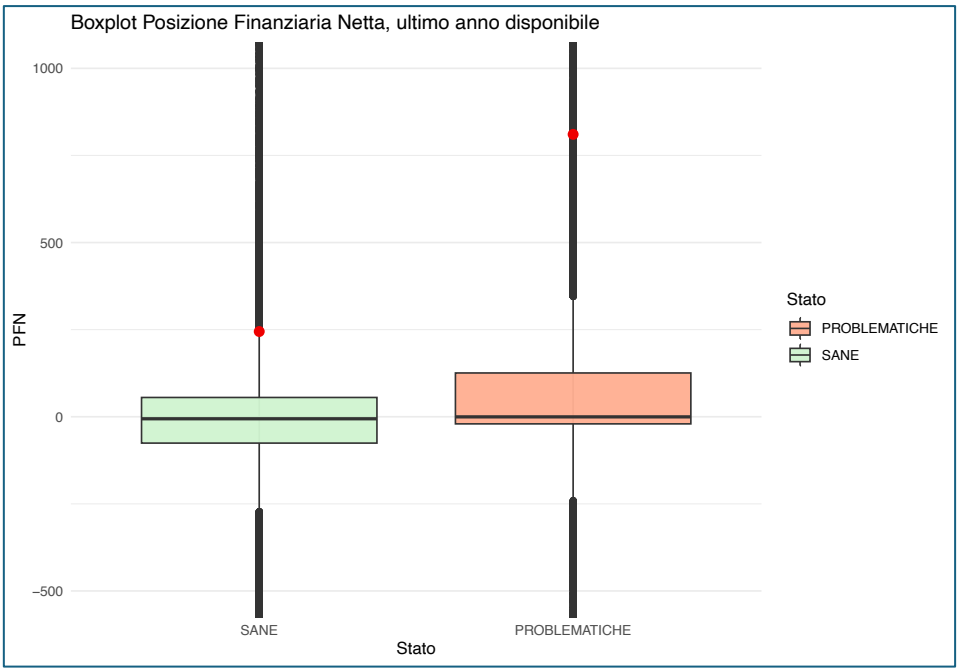


Figura 3.7: Box Plot della variabile "posizione finanziaria netta" confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA

3.3.6 Indici e Voci di bilancio – andamento medio e mediano

Il seguente paragrafo esamina tutte le variabili rappresentative degli indici di bilancio per le quali sono disponibili dati storici, fino a tre anni precedenti all'ultimo anno registrato. L'analisi di queste variabili permette di confrontare l'andamento nel tempo dei due gruppi di aziende, evidenziando eventuali tendenze.

Tutti i dati utilizzati per la realizzazione dei grafici presentati nel paragrafo sono riportati in *Appendice G*.

Indice di liquidità

Questo indice misura la capacità di un'impresa di far fronte agli impieghi necessari per il corretto funzionamento delle sue attività nel breve termine. Viene calcolato attraverso il rapporto tra attività e passività correnti: un valore più elevato indica una maggiore disponibilità di risorse a breve termine rispetto ai debiti di pari scadenza. Di conseguenza, un indice alto riflette una situazione di liquidità più solida e una maggiore stabilità finanziaria dell'azienda.

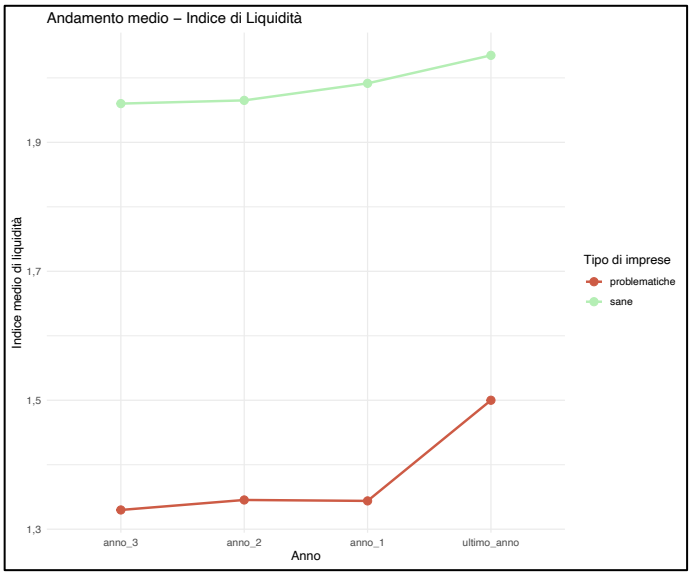


Figura 3.8: Scatter Plot, andamento medio "indice di liquidità". Fonte: elaborazione personale su dati AIDA

Come possiamo osservare nel grafico in Figura 3.8 l'andamento medio delle aziende sane si aggira intorno ad un valore dell'indice pari a 2, valore molto alto che indica un'ottima gestione delle liquidità aziendali. Per le imprese che hanno riscontrato problematiche di continuità aziendale, invece, l'andamento medio dell'indice è molto più basso; infatti, non si aggira su livelli critici ma comunque indica un margine di sicurezza molto ridotto o nullo.

Indice di indebitamento a breve termine

L'indice di indebitamento di breve periodo permette di misurare quanto il capitale proprio permanente permetta di supportare i debiti che richiedono di essere rimborsati nel breve termine; viene calcolato attraverso il rapporto tra passività correnti e patrimonio netto.

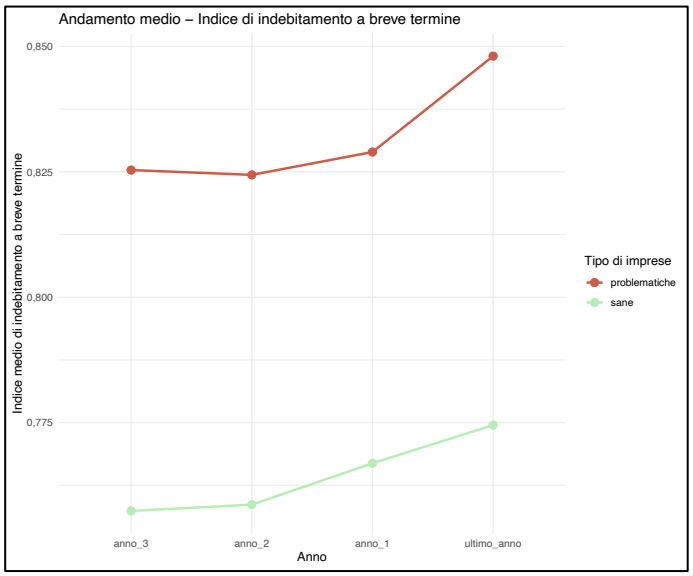


Figura 3.9: Scatter Plot, andamento medio "indice di indebitamento a breve termine ". Fonte: elaborazione personale su dati AIDA

Il livello ottimale è inferiore a 0,5 permettendo ai debiti di essere ampiamente coperti dal patrimonio netto; un valore superiore a 1 invece indica una gestione molto critica. Come possiamo osservare nel grafico le imprese sane hanno un livello di indebitamento più basso mentre le imprese problematiche hanno mediamente un andamento vicino a 1.

Indice di indebitamento a lungo termine

Osservando l'andamento del livello di indebitamento a lungo termine, può inizialmente sembrare controintuitivo che le aziende sane presentino un indebitamento a lungo termine superiore alle aziende che nel tempo hanno riscontrato problematiche di continuità aziendale. Tuttavia, la motivazione può risiedere nel fatto che le imprese più solide godono di maggior fiducia da parte dei creditori e possono quindi accedere più facilmente al debito a lungo termine.

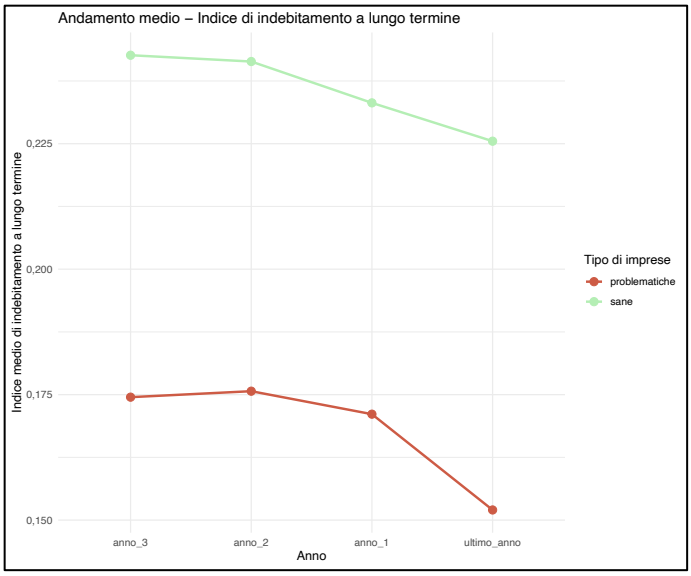


Figura 3.10: Scatter Plot, andamento medio "indice di indebitamento a lungo termine ". Fonte: elaborazione personale su dati AIDA

Al contrario, le aziende meno affidabili, non disponendo della stessa credibilità, non riescono

a beneficiare di tali forme di finanziamento e devono quindi limitarsi a fonti di capitale di più breve durata o di minore entità, come conferma l'analisi del grafico.

Indice di copertura delle immobilizzazioni patrimoniali

Questo indice, attraverso il rapporto tra il patrimonio netto e l'attivo immobilizzato, consente di valutare se gli investimenti durevoli dell'impresa siano stati finanziati con risorse proprie (quindi anch'esse durevoli) oppure con capitali di terzi. Un valore elevato dell'indice indica una situazione positiva, poiché significa che l'attivo immobilizzato è coperto interamente o prevalentemente da capitale proprio.

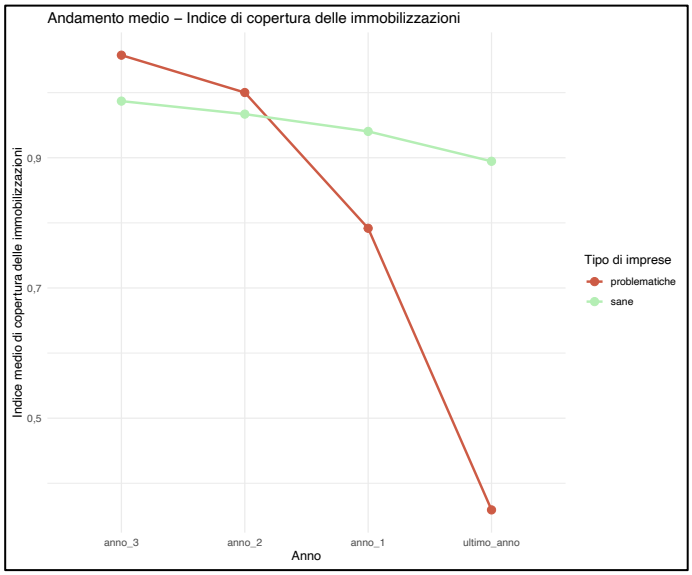


Figura 3.11: Scatter Plot, andamento medio "indice di copertura delle immobilizzazioni". Fonte: elaborazione personale su dati AIDA

Al contrario, un valore basso evidenzia che l'attivo immobilizzato ha un peso elevato rispetto al patrimonio netto, suggerendo che parte degli investimenti durevoli è stata finanziata tramite debito.

Le imprese solide mostrano un andamento stabile dell'indice, mediamente intorno al valore di 1, indicativo di un buon equilibrio tra fonti e impieghi. Le imprese in difficoltà, invece, registrano negli ultimi anni un forte calo di questo indicatore: ciò implica che una quota sempre minore delle immobilizzazioni è finanziata con capitale proprio, con un conseguente aumento del ricorso all'indebitamento.

Rapporto di indebitamento

Il rapporto di indebitamento permette di monitorare la relazione tra il totale dei debiti e il capitale proprio (*equity*). Dal grafico si può osservare che, nelle aziende sane, questo indicatore mantiene un andamento stabile e su livelli relativamente bassi. Al contrario, nelle imprese in difficoltà il rapporto risulta significativamente più elevato e mostra, solo nell'ultimo anno, un brusco calo. Tale dinamica potrebbe essere spiegata da un tentativo estremo di ridurre l'indebitamento, forse attraverso la vendita di attività immobilizzate, nel tentativo di riequilibrare la situazione finanziaria.

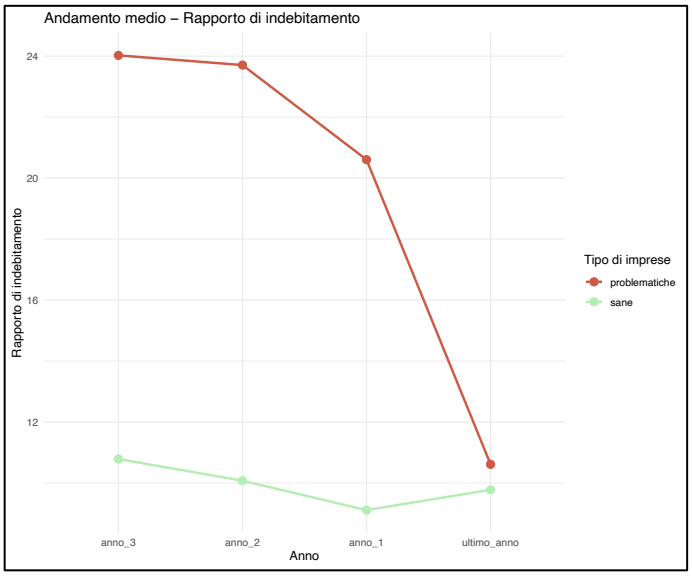


Figura 3.12: Scatter Plot, andamento medio "rapporto di indebitamento". Fonte: elaborazione personale su dati AIDA

Oneri finanziari su fatturato

Gli oneri finanziari rappresentano il costo sostenuto dall'azienda per l'utilizzo di capitale di terzi. Si tratta quindi di un costo fisso che l'impresa deve affrontare indipendentemente dall'andamento operativo. Analizzando il rapporto tra oneri finanziari e fatturato, è possibile valutare il grado di rischio finanziario a carico dell'azienda.

Un rapporto relativamente basso, come nel caso delle aziende sane, indica che gli oneri finanziari sono adeguatamente coperti dal fatturato: le attività operative generano risorse sufficienti a sostenere il costo del debito. Al contrario, nelle imprese in difficoltà, il rapporto si presenta su livelli più elevati, evidenziando

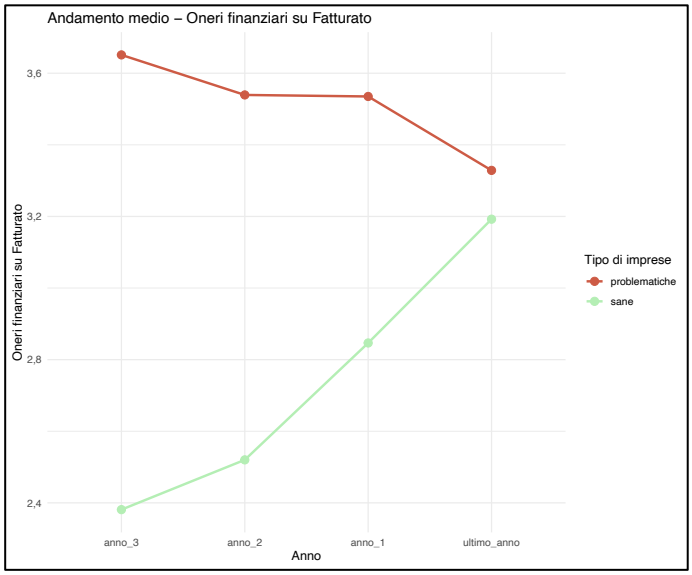


Figura 3.13: Scatter Plot, andamento medio "oneri finanziari su fatturato". Fonte: elaborazione personale su dati AIDA

una possibile criticità nell'equilibrio finanziario, poiché le attività aziendali faticano a coprire il peso degli interessi passivi.

Indice di indipendenza finanziaria

Questo indice misura il rapporto tra il capitale proprio e le fonti di finanziamento esterne.

Il range minimo di sicurezza è generalmente compreso tra il 33% e il 66%, indicando che almeno un terzo dell'azienda è finanziata con mezzi propri. Come si può osservare dal grafico, le imprese sane si collocano all'interno di questo intervallo, mostrando un buon livello di solidità patrimoniale. Al contrario, livelli più bassi di autonomia finanziaria, come quelli delle imprese problematiche, evidenziano una maggiore dipendenza da fonti di finanziamento esterne, quindi un più alto livello di indebitamento, con le conseguenti criticità legate ai costi del debito e alla vulnerabilità finanziaria.

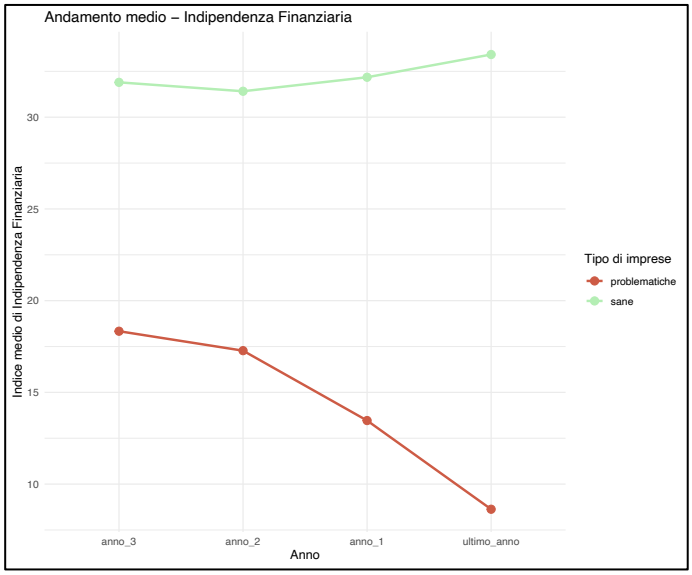


Figura 3.14: Scatter Plot, andamento medio "indice di indipendenza finanziaria". Fonte: elaborazione personale su dati AIDA

Rotazione capitale investito

Questo indice consente di valutare la capacità di un'azienda di generare ricavi di vendita (fatturato) in relazione al totale dei propri investimenti (capitale investito), offrendo così una misura dell'efficienza gestionale.

In linea generale, un valore più elevato di questo indicatore è considerato positivo, poiché segnala una maggiore capacità di generare ricavi rispetto al capitale impiegato.

Nel caso analizzato, tuttavia, il grafico mostra che le imprese sane presentano livelli costanti intorno all'unità, evidenziando una situazione di stabilità operativa.

Le imprese problematiche, invece, registrano valori mediamente più elevati, con una brusca impennata nell'ultimo anno. Sebbene tale andamento possa sembrare positivo in teoria, è opportuno approfondirne le cause, poiché potrebbe derivare non da una reale crescita dei ricavi, ma da una riduzione forzata del capitale investito dovuta al basso merito creditizio e alle difficoltà finanziarie di queste aziende.

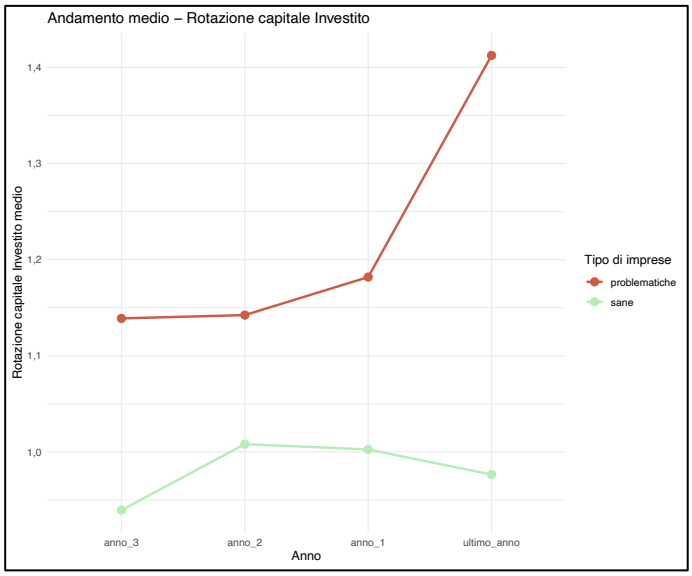


Figura 3.15: Scatter Plot, andamento medio "rotazione del capitale investito". Fonte: elaborazione personale su dati AIDA

EBITDA

L'EBITDA misura il margine operativo lordo e rappresenta la ricchezza che un'azienda è in grado di generare attraverso la propria attività operativa. In questa analisi sono riportati due grafici: il primo (Figura 3.16) mostra l'andamento dell'EBITDA in valore assoluto (in migliaia di euro), mentre il secondo (Figura 3.17) rappresenta l'andamento dell'EBITDA *margin*, ossia l'EBITDA rapportato al fatturato in termini percentuali.

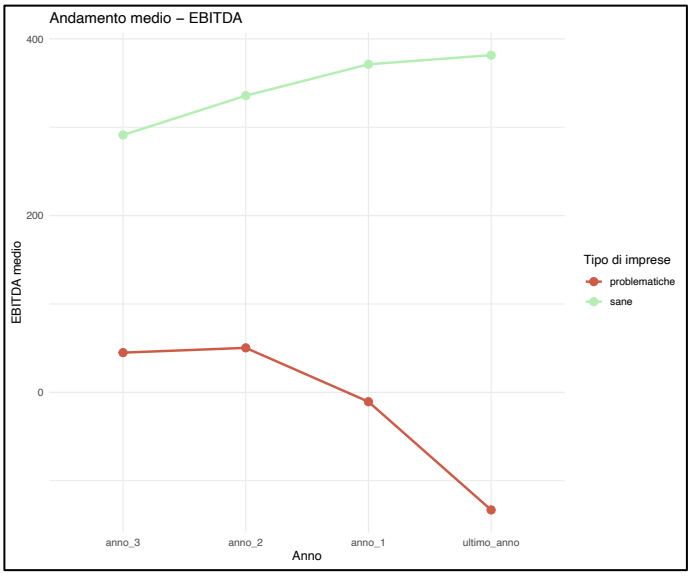


Figura 3.16 : Scatter Plot, andamento medio "EBITDA". Fonte: elaborazione personale su dati AIDA

Dal primo grafico si osserva come le imprese sane presentino un livello di margine operativo nettamente superiore e costante nel tempo, con un progressivo miglioramento della capacità di generare ricchezza operativa. Al contrario, le imprese problematiche mostrano valori molto bassi, prossimi allo zero, e un rapido

crollo nell'ultimo anno, segnale che l'azienda non riesce più a coprire i costi operativi, erodendo completamente il proprio margine.

Per quanto riguarda l'andamento dell'EBITDA *margin*, si è scelto di utilizzare la mediana anziché la media, poiché rappresenta un indicatore più robusto e meno influenzato da valori anomali. Come evidenziato dal grafico, le imprese sane mantengono livelli percentuali positivi e relativamente elevati, indice di una situazione stabile. Le imprese problematiche, invece, mostrano valori bassi e un drastico crollo in negativo nell'ultimo anno, in coerenza con il grafico precedente e indicativo di uno squilibrio strutturale.

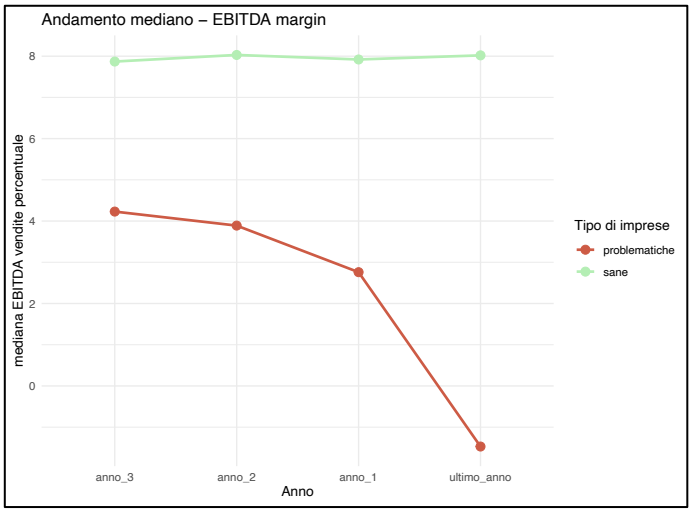


Figura 3.17: Scatter Plot, andamento mediano "EBITDA margin". Fonte: elaborazione personale su dati AIDA

Redditività del capitale proprio - ROE

Il ROE è un indicatore di grande rilievo perché misura la capacità di un'azienda di generare profitto attraverso il rapporto tra utile netto e patrimonio netto. Un valore elevato del ROE indica una maggiore redditività del capitale proprio e, quindi, una migliore capacità dell'impresa di creare valore.

Dal grafico si può osservare che le imprese sane presentano un ROE positivo, segno di una gestione reddititzia e sostenibile. Al contrario, le imprese problematiche mostrano valori nettamente negativi, evidenziando una situazione in cui l'azienda non solo non genera utili, ma distrugge valore, erodendo progressivamente il capitale dei soci.

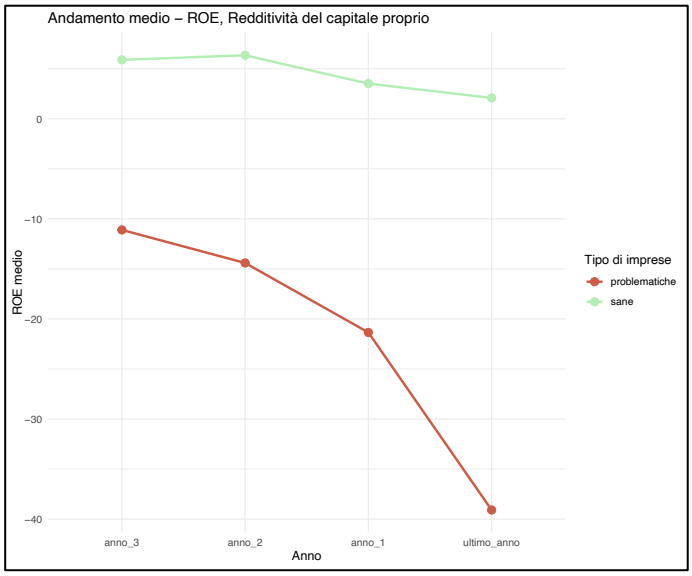


Figura 3.18: Scatter Plot, andamento medio "ROE". Fonte: elaborazione personale su dati AIDA

Capitale circolante netto

Il capitale circolante netto (CCN) consente di valutare se le attività a breve termine sono sufficienti a coprire le passività a breve termine. Si calcola come differenza tra attività correnti e passività correnti.

Un valore positivo indica che le attività a breve superano le passività, segnalando la capacità dell'azienda di far fronte ai propri debiti nel breve periodo.

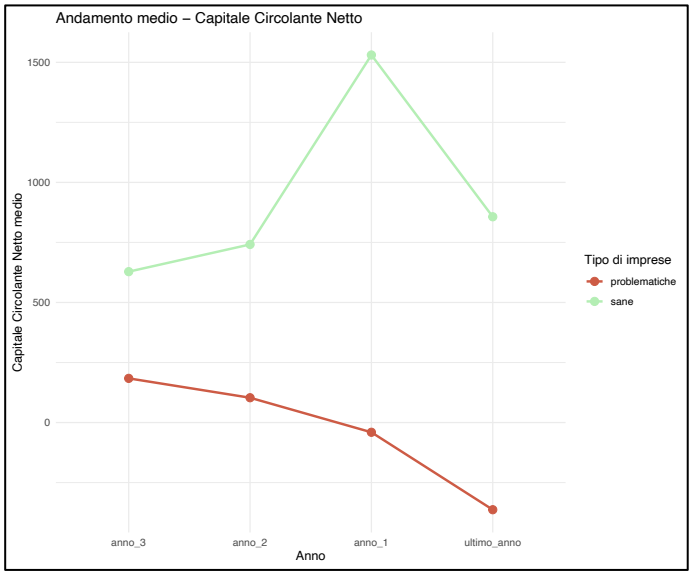


Figura 3.19: Scatter Plot, andamento medio "capitale circolante netto". Fonte: elaborazione personale su dati AIDA

Come si osserva dal grafico in Figura 3.19, le imprese sane mostrano valori ampiamente positivi, indice di una buona solidità e liquidità aziendale. Al contrario, valori negativi del CCN indicano che le passività a breve termine superano le attività correnti, evidenziando difficoltà nel rispettare le scadenze dei debiti. In questo caso, le imprese problematiche si collocano su valori prossimi allo zero o negativi, confermando una situazione di tensione finanziaria e scarsa liquidità.

Totale Attivo e Totale Passivo

I grafici del Totale Attivo e del Totale Passivo mostrano andamenti identici. Ciò conferma la correttezza contabile dei dati analizzati, poiché in ogni bilancio d'esercizio il Totale Attivo (Impieghi) e il Totale Passivo (Fonti di finanziamento) devono sempre coincidere.

Possiamo quindi osservare che le imprese sane hanno un volume d'affari e una capacità di raccolta di capitali nettamente

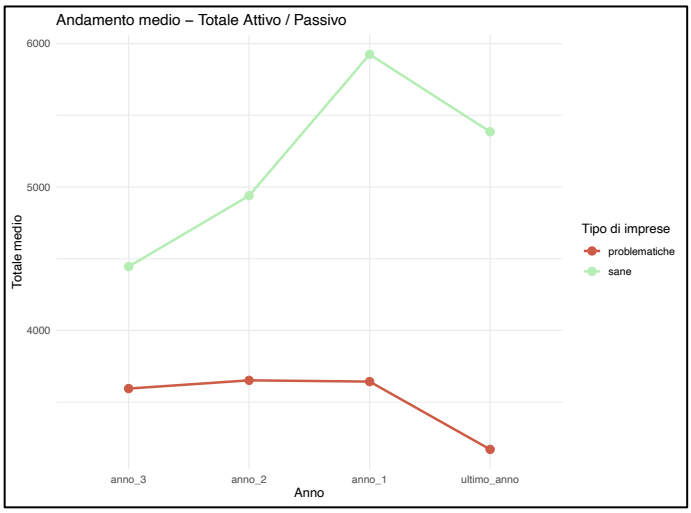


Figura 3.20: Scatter Plot, andamento medio "totale attvio/passivo". Fonte: elaborazione personale su dati AIDA

superiori e in crescita rispetto alle imprese problematiche, che al contrario evidenziano andamenti inferiori e con tendenze in calo.

Totale patrimonio netto

Nel grafico viene rappresentato l'andamento medio del patrimonio netto.

Il trend delle imprese sane è chiaramente positivo e in crescita, il che indica che queste aziende generano utili e incrementano il loro capitale sociale, rafforzando progressivamente la loro solvibilità e indipendenza finanziaria. Al contrario, le imprese problematiche mostrano un andamento decrescente, tendente a valori negativi, probabilmente a causa di perdite d'esercizio continuative che non riescono a compensare.

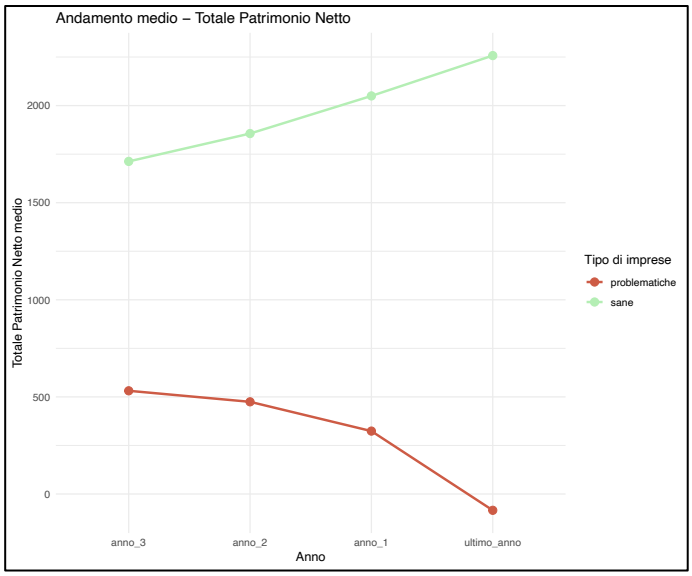


Figura 3.21: Scatter Plot, andamento medio "patrimonio netto". Fonte: elaborazione personale su dati AIDA

Totale Debito

Per la rappresentazione grafica del totale dei debiti in migliaia di euro si è scelto di utilizzare la mediana invece della media, grazie alla sua minore sensibilità ai valori estremi. Questa scelta ci permette di osservare che, come ci si potrebbe aspettare, le imprese problematiche presentano un totale dei debiti più elevato, con una tendenza crescente fino all'ultimo anno, quando si registra un rapido calo. Tale fenomeno, come già evidenziato in precedenza, può essere attribuito alla perdita di merito creditizio e, di conseguenza, alla difficoltà di accedere a nuovi finanziamenti e di mantenere quelli preesistenti.

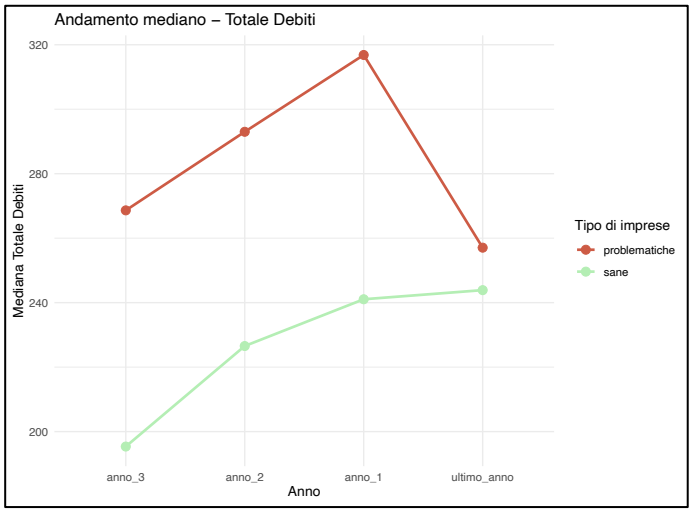


Figura 3.22: Scatter Plot, andamento mediano "totale debito". Fonte: elaborazione personale su dati AIDA

Utile e Perdita di esercizio

L'ultimo valore di bilancio che si vuole analizzare è il livello di utile o perdita espresso in migliaia di euro.

Come indicatore ultimo della stabilità e redditività dell'impresa, possiamo notare che le imprese sane presentano un andamento costante e positivo, mentre le imprese che nel tempo hanno sviluppato problematiche di continuità aziendale presentano un andamento negativo e decrescente, causato da perdite in bilancio.

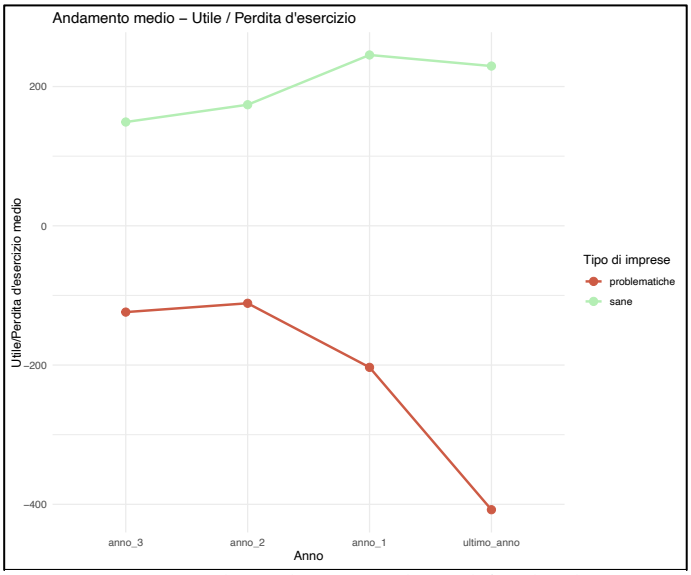


Figura 3.23: Scatter Plot, andamento medio "utile/perdita d'esercizio". Fonte: elaborazione personale su dati AIDA

L'analisi sviluppata sugli indici e sulle principali voci di bilancio conferma i differenti andamenti che caratterizzano il gruppo delle imprese sane rispetto a quello delle imprese che hanno sviluppato problematiche di continuità aziendale. I trend osservati nelle imprese sane risultano coerenti con le teorie economico-finanziarie di solidità e redditività, mentre per le imprese problematiche emergono in modo sistematico indicatori di tensione finanziaria.

Tali evidenze descrittive mettono in luce differenze strutturali tra i due gruppi, differenze che costituiscono il fondamento per le successive analisi predittive, finalizzate a individuare con maggiore accuratezza se un'impresa presenti caratteristiche riconducibili a situazioni di potenziale rischio.

3.4 Correlazioni tra variabili

In questo paragrafo si analizza la correlazione tra le variabili quantitative presenti nel database, più precisamente quelle che rappresentano indici e voci di bilancio. Lo studio della correlazione permette di comprendere la forza e la direzione della relazione lineare tra due variabili.

Prima di procedere con i test specifici, è necessario verificare la distribuzione delle variabili, poiché l'applicazione della correlazione di Pearson richiede che le variabili seguano una distribuzione Gaussiana, ovvero che si presentino con una forma a campana e simmetrica. Sono quindi state effettuate verifiche grafiche mediante istogrammi, per osservare che la forma della distribuzione fosse a campana, e tramite QQ-plot, per valutare l'aderenza alla normalità, quindi quanto i dati deviassero dall'andamento normale.

Successivamente sono stati utilizzati test statistici formali, come il test di Shapiro, che attraverso il p-value consente di determinare se la variabile può essere considerata normalmente distribuita: valori del p-value molto bassi indicano una chiara deviazione dalla normalità.

Da tutte le analisi effettuate risulta che nessuna variabile segue una distribuzione normale, per questo motivo si è applicato il metodo delle trasformazioni.

Si ricorre alle trasformazioni per rendere le distribuzioni delle variabili continue il più normale possibile, cercando di limitare l'influenza di valori estremi e di varianze troppo elevate, questo perché l'applicazione di test parametrici, come la classica correlazione di Pearson, su variabili con una distribuzione non Gaussiana può generare distorsioni nei risultati.

Una delle metodologie per affrontare questo processo è l'applicazione di funzioni all'insieme delle variabili, la più comune è l'applicazione della funzione logaritmica che riduce fortemente l'asimmetria positiva. Questo vuol dire che può essere utilizzata solamente con valori positivi, di conseguenza non è stato possibile utilizzare questo tipo di trasformazione, a causa del grande numero di valori negativi nell'insieme di dati analizzati.

Un'altra metodologia applicabile per avere una maggiore flessibilità è quella chiamata Box-Cox, basata sulla trasformazione dei dati in termini esponenziali *(Box, G. E. P., & Cox, D. R. 1964)*. Secondo questa metodologia, l'algoritmo cerca autonomamente l'esponente più adatto da applicare al dato per renderlo il più normale possibile; tuttavia, anche questa applicazione richiede che i dati siano strettamente positivi.

In un contesto reale dove sono presenti dati economici-finanziari concreti è comune lavorare in presenza di molti dati negati e prossimi allo zero. Per questo motivo si è scelto di utilizzare il metodo della trasformazione Yeo–Johnson, che consiste in un'estensione della metodologia Box-Cox e che consente di trattare anche variabili con segno misto. La Yeo–Johnson applica una trasformazione di potenza diversa per valori positivi e negativi, stimando automaticamente il parametro che ottimizza la normalizzazione della variabile *(Yeo, I. K., & Johnson, R. A. 2000)*.

Nemmeno attraverso questo test è stato possibile migliorare la forma delle distribuzioni delle variabili per renderle più adatte a metodi statistici che richiedessero assunzioni di normalità. Per questo motivo la correlazione tra le variabili verrà d'ora in avanti calcolata attraverso test non parametrici, nello specifico verrà utilizzata la correlazione di Spearman.

La correlazione di Spearman misura l'associazione non parametrica tra due variabili quantitative e a differenza della correlazione di Pearson non richiede normalità, permettendo quindi di misurare quanto due variabili aumentano o diminuiscono insieme in modo monotono, non necessariamente lineare *(Spearman, C., 1904; Siegel, S., & Castellan, N. J. 1988)*.

Questo tipo di correlazione funziona attraverso la suddivisione in ranghi: vengono convertiti i valori originali delle osservazioni in ranghi, successivamente si calcola la correlazione di Pearson sui ranghi generati.

Formalmente:

$$\rho = cor(rank(X_i), rank\ (Y_i))$$

Il processo di creazione dei ranghi consiste nell'ordinare tutte le osservazioni in ordine crescente e nell'assegnare a ciascuna di esse un numero intero che sostituisce il valore originario. L'osservazione più piccola riceve il rango 1, mentre alle successive vengono attribuiti ranghi progressivi in base alla loro posizione nell'ordinamento.

Nel caso in cui il valore di alcune osservazioni si ripetesse, ad esse viene attribuito un valore mediato. Questo vuol dire che se ad esempio sono la seconda e la terza osservazione ad avere un valore originario uguale, invece di attribuirgli i ranghi 2 e 3 si svolge il seguente calcolo: (2+3) / 2 = 2,5; alla seconda e alla terza osservazione verrà quindi attribuito il rango 2,5.

Successivamente sui ranghi viene calcolata la correlazione di Pearson, calcolata come il rapporto tra la covarianza e la varianza standardizzata, secondo la formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- la *covarianza*, calcolata attraverso il prodotto degli scarti della media, misura la variazione congiunta di due variabili, come le variabili si muovono a vicenda;
- La *varianza*, calcolata con i quadrati degli scarti della media, calcola quanto i dati di un insieme si discostino dalla loro media.

Grazie al calcolo della correlazione di Spearman è possibile comprendere come due variabili si influenzino reciprocamente e quanto siano correlate tra loro, anche se la loro distribuzione non è normale. Il coefficiente di correlazione di Spearman assume valori compresi tra -1 e +1: -1 indica la massima correlazione negativa, cioè quando un aumento di una variabile comporta una diminuzione proporzionale dell'altra; +1 indica la massima correlazione positiva, ossia quando le due variabili si influenzano nella stessa direzione in maniera proporzionale.

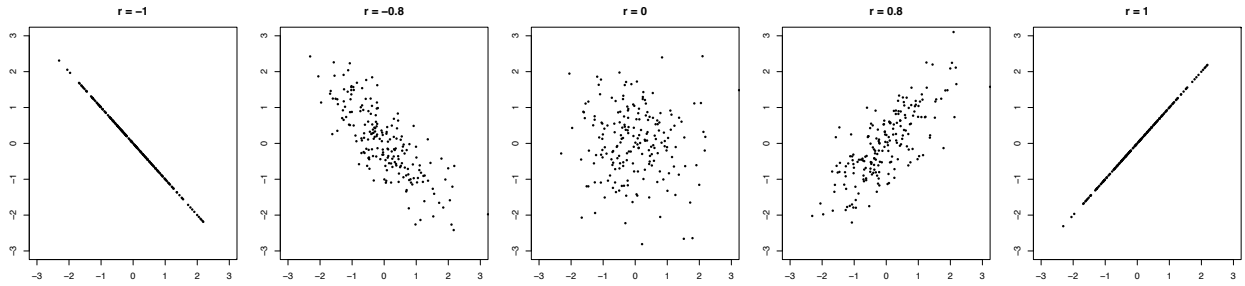


Figura 3.24: esempi grafici per livello di correlazione. Fonte: elaborazione personale attraverso software R

Lo studio della correlazione è inoltre fondamentale per identificare la multicollinearità. Quando due variabili presentano un coefficiente di correlazione molto elevato (in questa analisi considereremo superiore a 0,8), esse tendono a fornire informazioni ridondanti. In linea generale per evitare distorsioni nei modelli causate da ripetizioni eccessive, per le coppie di variabili con correlazione maggiore di 0,8 si terrà in considerazione solo una delle due variabili.

Di seguito sono riportate le matrici di correlazione relative ai due database: il grafico in Figura 3.25 per le imprese sane e il grafico in Figura 3.26 per le imprese con problematiche di continuità aziendale.

Le matrici di correlazione rappresentano graficamente le relazioni tra le variabili, consentendo di individuare quelle con i livelli di correlazione più elevati. Nei grafici, le correlazioni positive sono indicate con tonalità di blu, mentre quelle negative con sfumature di rosso. Inoltre, l'inclinazione delle ellissi aiuta a interpretare in modo intuitivo la direzione della relazione tra le variabili.
Ellissi più sottili e dai colori più intensi indicano correlazioni più forti.

In *Appendice H* vengono riportate le matrici contenenti tutti i coefficienti di correlazione, dalle quali possiamo apprendere che le variabili a maggior rischio di multicollinearità, quindi di ripetizione di informazioni per le imprese sane, sono:

i_indebitamento_breve	~	i_indebitamento_lungo	-0,9997
EBITDA_margin	~	ROS	0,8825
tot_debiti	~	tot_passivo_attivo	0,8630
EBITDA	~	utile_perdita	0,8068

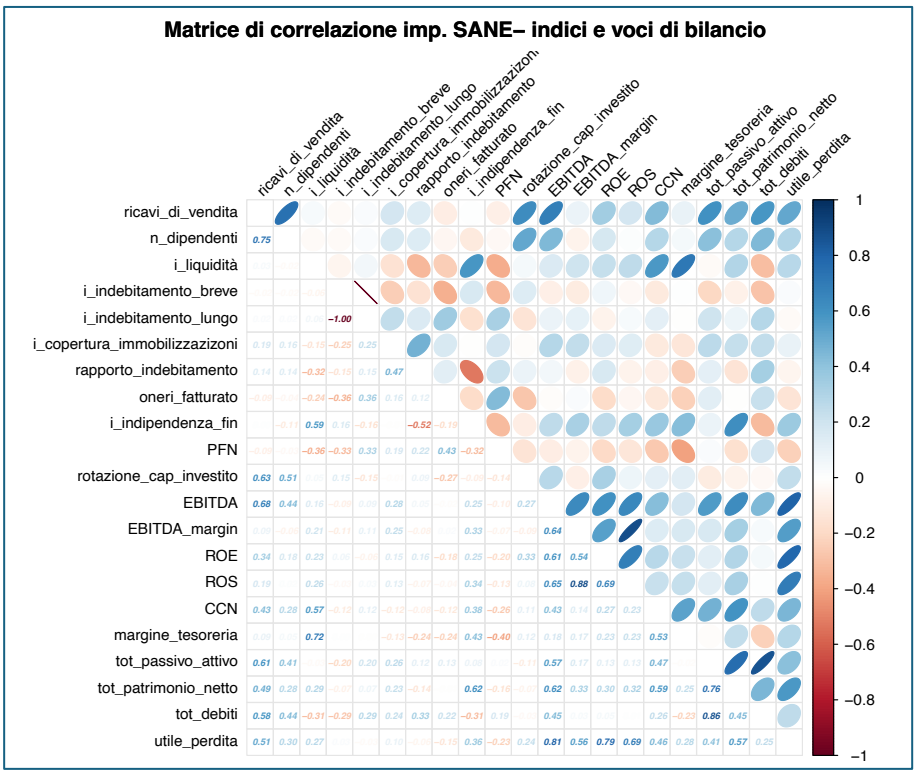


Figura 3.25: Matrice di correlazione tra indici e voci di bilancio di imprese sane. Fonte: elaborazione personale su dati AIDA

Nelle imprese caratterizzate da una struttura finanziaria sana si riscontra un'elevata correlazione negativa tra l'indice di indebitamento a breve termine e quello a lungo termine. Tale relazione

deriva dal loro andamento speculare che si può osservare anche nell'analisi descrittiva fatta in precedenza: un incremento del ricorso al debito a breve termine si associa generalmente a una contrazione dell'indebitamento a più lunga scadenza. Questo fenomeno riflette il processo attraverso il quale l'impresa che incrementa la quota di debito con rimborso immediato, diminuisce contestualmente la necessità, e probabilmente anche la possibilità, di contrarre ulteriori debiti con orizzonte temporale più esteso. Data l'origine di questa elevata multicollinearità, nelle analisi successive si considererà solo l'indice di indebitamento a lungo termine.

La forte relazione positiva tra EBITDA *margin* e ROS è del tutto normale. Come evidenziato nell'analisi descrittiva, entrambi gli indicatori misurano la redditività operativa rapportata ai ricavi: l'EBITDA *margin* esprime la marginalità prima di ammortamenti e accantonamenti, mentre il ROS rappresenta la redditività operativa netta sulle vendite. Poiché il denominatore di entrambi i rapporti si basa sulla stessa grandezza (i ricavi), è naturale che si muovano nella stessa direzione e mostrino una correlazione così elevata.

La correlazione tra totale debiti e totale passivo, così come quella tra EBITDA e utile/perdita, riflette invece l'esistenza di relazioni strutturali tra le voci di bilancio. Il totale del passivo include infatti il totale dei debiti, per cui un aumento di questi ultimi determina quasi automaticamente un incremento del passivo complessivo. Analogamente, l'EBITDA rappresenta una fase intermedia del conto economico da cui deriva, attraverso ammortamenti, accantonamenti, oneri finanziari e imposte, il risultato netto. È quindi fisiologico che variazioni del margine operativo lordo si riflettano, con elevata coerenza, sull'utile finale.

Come mostrato dalla matrice di correlazione in Figura 3.26, anche nelle imprese che presentano problematiche di continuità aziendale emergono le stesse relazioni forti già osservate nel campione delle imprese sane, con coefficienti di segno e intensità molto simili. Questo conferma che alcuni legami strutturali tra gli indici economico-finanziari permangono indipendentemente dalla condizione di stabilità.

i_indebitamento_breve	~	i_indebitamento_lungo	-0,9990
EBITDA_margin	~	ROS	0,9460
tot_debiti	~	tot_passivo_attivo	0,8524
EBITDA	~	utile_perdita	0,8371

Nel caso delle imprese problematiche si rilevano inoltre alcune ulteriori correlazioni particolarmente elevate ($\rho > 0{,}80$), nello specifico:

ROS	~	utile_perdita	0,8120
EBITDA	~	ROS	0,8531
EBITDA	~	EBITDA_margin	0,8817
i_indipendenza_fin	~	tot_patrimonio_netto	0,8333

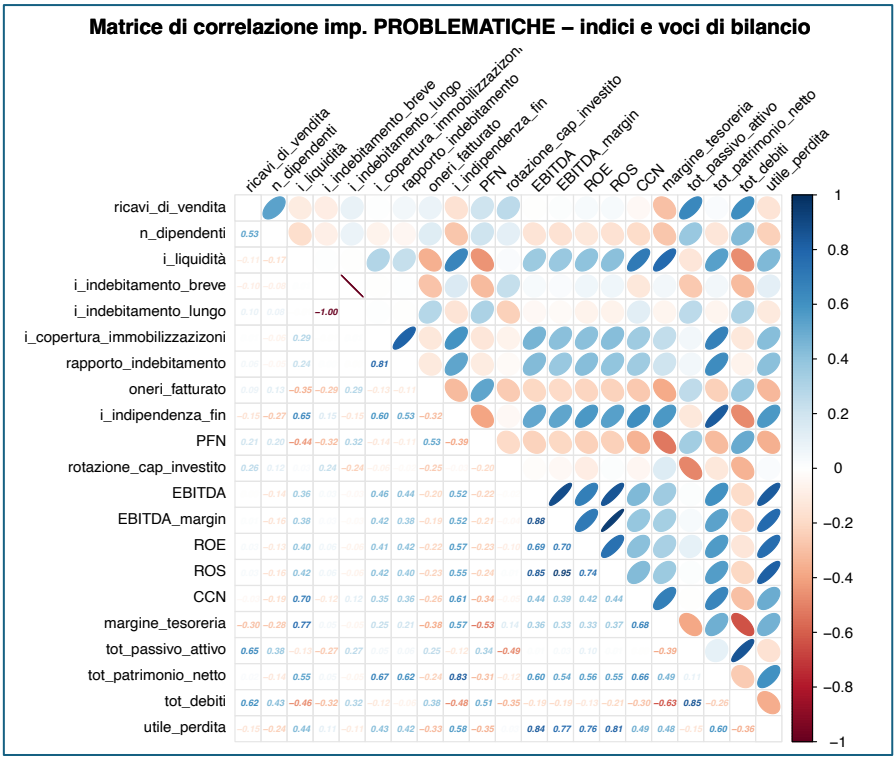


Figura 3.26:Matrice di correlazione tra indici e voci di bilancio di imprese che hanno riscontrato problematiche di continuità aziendale. Fonte: elaborazione personale su dati AIDA

Anche in questo gruppo, analogamente a quanto osservato per le imprese sane, le correlazioni più forti riguardano indicatori che descrivono aspetti economici e finanziari strettamente connessi.

Ad esempio, la relazione tra ROS e utile netto riflette il fatto che il ROS misura la redditività operativa mentre l'utile rappresenta il risultato finale dell'esercizio: una marginalità operativa elevata tende naturalmente a tradursi in un risultato netto migliore. Allo stesso modo, la forte associazione tra EBITDA, EBITDA margin e ROS evidenzia la coerenza interna degli indicatori legati alla gestione caratteristica.

Oltre all'individuazione delle coppie di variabili caratterizzate da correlazioni molto elevate, che si traducono in potenziali fonti di multicollinearità, queste analisi consentono di mettere in evidenza relazioni particolarmente significative che saranno utili nelle fasi successive dello studio, sia nell'interpretazione delle dinamiche economico-finanziarie sia nella selezione delle variabili più rilevanti per i modelli.

3.4.1 Valutazione comparativa delle medie nei due gruppi aziendali

In questa fase dell'analisi si procede al confronto tra il gruppo delle imprese sane e quello delle imprese problematiche per comprendere se ci siano delle effettive differenze tra i valori osservati nei due gruppi.

Il t-test per campioni indipendenti, utilizzato inizialmente, appartiene alla famiglia dei test parametrici e presuppone la normalità delle variabili analizzate. Poiché nessuna delle variabili soddisfa tale assunzione, al fine di rendere l'analisi più robusta, il t-test è stato affiancato dal corrispondente test non parametrico di Mann–Whitney (Wilcoxon rank-sum), che non richiede l'ipotesi di normalità.

Operativamente, il t-test confronta le medie dei due gruppi, mentre il test di Mann–Whitney valuta le differenze tra le distribuzioni dei gruppi attraverso il confronto dei ranghi attribuiti ai valori originali. In questo modo è possibile verificare se i gruppi differiscono in maniera significativa sia in termini di media sia in termini di posizione nella distribuzione, superando le criticità legate alla non normalità e alla presenza di eventuali asimmetrie o outlier.

Entrambi i test permettono di stabilire se una determinata variabile assume valori significativamente diversi nei due gruppi, tenendo conto della variabilità interna.

Il risultato è sintetizzato dal p-value: valori molto piccoli indicano che, sotto l'ipotesi che le due popolazioni siano uguali, sarebbe estremamente improbabile osservare differenze così marcate. In tali casi la variabile può essere considerata un potenziale indicatore utile per evidenziare segnali di instabilità aziendale.

Al contrario, un p-value elevato suggerisce che la differenza osservata potrebbe essere attribuibile al caso, e che quindi la variabile non rappresenta un buon discriminante tra imprese sane e imprese problematiche.

Il test di Mann–Whitney fornisce un p-value non parametrico e rappresenta una verifica dei risultati ottenuti tramite il t-test. Poiché le variabili analizzate non presentano distribuzioni normali, nei casi di discordanza tra i due test è stato considerato prevalente il risultato del test di Mann–Whitney, essendo metodologicamente più robusto in presenza di distribuzioni non Gaussiane e asimmetrie.

I risultati completi relativi a tutte le variabili sono riportati in *Appendice G*. L'analisi è stata condotta adottando un livello di significatività pari a 0,05, soglia comunemente utilizzata nella pratica statistica.

Di seguito vengono presentate le variabili che mostrano i p-value più bassi e più alti, con l'obiettivo di evidenziare rispettivamente gli indicatori più informativi e quelli meno utili nel distinguere tra imprese sane e problematiche.

Le variabili che presentano differenze fortemente significative per entrambi i test, che possono pertanto essere considerate potenziali indicatori di segnali di instabilità economico-finanziaria, includono:

- l'indice di liquidità, fino al terzo anno precedente;
- l'indice di indebitamento, sia a breve che a lungo termine, relativo all'ultimo anno;
- l'indice di indipendenza finanziaria, fino al terzo anno precedente;
- la rotazione del capitale investito (ultimo anno);
- la redditività del capitale proprio (ROE), fino al terzo anno precedente.

Analizzando anche la direzione delle relazioni, emerge che per variabili come l'indice di liquidità la media delle imprese sane risulta sistematicamente più elevata rispetto a quella delle imprese problematiche. Ciò conferma quanto emerso nell'analisi descrittiva, secondo cui una maggiore liquidità costituisce un indicatore di solidità finanziaria. Diversamente, nel caso della rotazione del capitale investito, le imprese problematiche presentano valori medi superiori.

Tale comportamento, già evidenziato nella fase descrittiva, può essere interpretato come effetto di una contrazione del capitale investito dovuta a difficoltà economico-finanziarie.

Per le variabili che non presentano differenze statisticamente significative tra i due gruppi (p-value superiore a 0,05), si osservano alcune discrepanze tra i risultati del t-test e quelli del test di Mann–Whitney. In tali situazioni, data la non normalità delle distribuzioni, si è scelto di

considerare prevalenti i risultati del Mann–Whitney, più adatto alla natura dei dati analizzati. Le variabili non significative secondo il test non parametrico risultano quindi le seguenti:

- indice di copertura delle immobilizzazioni patrimoniali del terzo anno antecedente,
- il totale attivo e passivo del secondo anno.

In questi casi, la direzione della differenza media non risulta interpretabile in modo coerente, poiché la relazione tra i gruppi appare casuale, come confermato dai valori di p-value elevati.

In sintesi, l'analisi condotta attraverso il t-test conferma l'esistenza di differenze strutturali tra imprese sane e problematiche per la maggior parte delle variabili. Le variabili economico-finanziarie che mostrano una significatività elevata rappresentano indicatori coerenti e affidabili dello stato di salute aziendale, mentre quelle non significative non consentono di discriminare tra i due gruppi. Questi risultati costituiscono una base solida per le successive fasi dell'analisi e per la selezione delle variabili maggiormente rilevanti nei modelli previsionali.

4 Le reti Bayesiane per lo studio della continuità aziendale

In questo capitolo viene descritto il processo di costruzione del modello di rete Bayesiana finalizzato alla previsione della continuità aziendale.

Dopo aver definito nel primo capitolo il quadro teorico di riferimento e aver approfondito nel secondo capitolo gli aspetti metodologici delle reti Bayesiane, e aver successivamente analizzato nel terzo capitolo la struttura del database e le caratteristiche delle variabili, in questa sezione si passa alla fase operativa di modellazione.

L'obiettivo del presente elaborato è sviluppare un modello capace di fare previsioni sulle probabilità che un'impresa sviluppi o meno problematiche di continuità aziendale sulla base degli indicatori economico-finanziari disponibili, attraverso l'identificazione di possibili scenari.

Tale classificazione si fonda proprio sui meccanismi di inferenza probabilistica presentati nel Capitolo 2: ogni variabile contribuisce ad aggiornare la probabilità associata allo stato di continuità, secondo la logica Bayesiana di calcolo della probabilità condizionata.

La costruzione del modello parte con una selezione delle variabili in coerenza con quanto emerso nel Capitolo 3. Sono state eliminate le variabili caratterizzate da elevata multicollinearità, così da evitare ridondanze informative che avrebbero potuto alterare i processi successivi.

In particolare, sono state rimosse le variabili relative all'indebitamento a breve fortemente correlate con quelle sull'indebitamento a medio-lungo termine, oltre alla voce "totale attivo/passivo" che risulta eccessivamente sovrapposta al totale dei debiti. Questa ulteriore semplificazione ha portato alla riduzione del numero di variabili coinvolte nell'analisi pari a 19, l'elenco delle variabili considerate viene riportato in *Appendice L*.

Successivamente è stata svolta la fase di apprendimento della rete Bayesiana che rappresenta le relazioni di dipendenza e indipendenza tra le variabili. Come discusso nel Capitolo 2, il DAG costituisce la struttura portante del modello Bayesiano, poiché definisce quali variabili influenzano direttamente le altre.

Il modello è stato sviluppato adottando un approccio multilivello, articolato su quattro esercizi contabili, indicati, dal più recente al meno recente, come: "ultimo anno", "anno_1", "anno_2", "anno_3". In questa sezione vengono descritte le strutture ottenute considerando esclusivamente le relazioni di dipendenza e indipendenza riscontrate nel database delle imprese sane, per ciascun livello di analisi. Successivamente, tali strutture fisiologiche vengono confrontate con le reti costruite sulle imprese problematiche, al fine di identificare in modo concreto le differenze che caratterizzano i due gruppi di imprese.

Successivamente si è identificato un modello utile al fine previsto. Per fare ciò si è utilizzato l'intero database, raggruppando insieme sia le imprese sane sia quelle problematiche. Al database è stata poi aggiunta una variabile dicotomica, considerata come target, rappresentativa della continuità aziendale. Tale scelta è stata motivata dal fatto che, una volta introdotto il nodo *continuità*, era necessario che il modello apprendesse correttamente le relazioni che caratterizzano le imprese con criticità, così da permettere alla rete di distinguere tra situazioni fisiologiche e situazioni problematiche.

Successivamente è stata condotta un'analisi specifica volta a definire la collocazione più appropriata del nodo relativo alla continuità aziendale all'interno del DAG. Alla luce delle evidenze empiriche raccolte, si è concluso che la configurazione ottimale fosse quella in cui il nodo *continuità* risultasse collegato esclusivamente alle variabili economico-finanziarie dell'ultimo esercizio, assumendo il ruolo di nodo figlio: esso riceve quindi relazioni dai nodi genitori, ma non esercita alcuna influenza in uscita verso altre variabili.

Il modello è stato quindi poi stimato a partire dai dati con l'indicazione di tali vincoli.

Infine, è stata sviluppata una fase inferenziale, tramite simulazioni di scenari su casi reali. L'obiettivo è stato quello di osservare il comportamento della rete al variare delle condizioni economico-finanziarie e comprendere la natura delle risposte probabilistiche generate dal modello.

4.1 Apprendimento della rete

Per apprendere il modello di rete Bayesiana coerente con le relazioni statistiche presenti nei dati, si è scelto di stimare le indipendenze condizionali tra le variabili mediante l'algoritmo *pc.stable* implementato in R. Tale algoritmo si basa sull'applicazione sistematica di test di indipendenza condizionata e consente di individuare esclusivamente le relazioni statisticamente significative.

Come approfondito nel capitolo metodologico (Capitolo 2), *pc.stable* identifica e orienta gli archi attraverso l'analisi sistematica delle indipendenze condizionali e delle v-structure. Nel contesto applicativo di questo studio, tali meccanismi permettono di ricostruire una struttura coerente con le relazioni statistiche presenti nei dati.

È attraverso queste configurazioni strutturali che l'algoritmo *pc.stable* determina l'orientamento degli archi in una rete Bayesiana.

L'algoritmo *pc.stable* si distingue dagli altri metodi di apprendimento strutturale perché non si basa sull'uso dei Markov Blanket. Come illustrato nel secondo capitolo, il Markov Blanket rappresenta l'insieme dei nodi che rende una variabile indipendente dal resto della rete; molti algoritmi (come Grow-Shrink, IAMB e le loro varianti) fondano l'intero processo di costruzione del DAG proprio sull'individuazione di tali insiemi.
Pc.stable, invece, adotta un approccio differente: parte da un grafo completamente connesso ed elimina progressivamente gli archi non supportati dai test di indipendenza condizionale. Parallelamente identifica e orienta le v-structure, garantendo che il grafo risultante sia aciclico e coerente con le indipendenze osservate nei dati (*Colombo, D., & Maathuis, M. H., 2014)*.

Dal punto di vista teorico, *pc.stable* rappresenta la prima implementazione operativa dell'algoritmo IC (Inductive Causation) proposto da T. S. Verma e J. Pearl *(Verma, T. S., & Pearl, J. 1991)*. L'algoritmo IC fornisce la base concettuale per la costruzione della struttura causale di un grafo a partire dalle relazioni di indipendenza condizionale e introduce formalmente il ruolo centrale delle v-structure come meccanismo di orientamento. Tuttavia, nella sua formulazione originaria, l'algoritmo IC non era applicabile a casistiche reali a causa dell'elevata complessità computazionale.

Come evidenziato da G. Scutarie e R. Nagarajan *(G. Scutari, R. Denis, 2014; R. Nagarajan, M. Scutari, S. Lebre, 2013)*, l'algoritmo *pc.stable* supera tali limiti traducendo la teoria in una procedura efficiente a livello computazionale, rendendo in questo modo possibile ricostruire strutture di dipendenza anche a partire da grandi insiemi di dati.

L'algoritmo *pc.stable* può stimare le relazioni di indipendenza condizionata utilizzando test parametrici, come il t-test, che presuppongono la normalità delle variabili, oppure test non parametrici, i quali non richiedono alcuna assunzione sulla forma della distribuzione. I test non parametrici basano infatti la loro logica sulle permutazioni: le variabili vengono "mescolate" ripetutamente in modo da renderle sicuramente indipendenti, e per ciascuna permutazione si calcola una misura di dipendenza tra le variabili rimescolate. In questo modo si ottiene una distribuzione empirica che rappresenta il livello di dipendenza atteso se le variabili fossero realmente indipendenti.

Una volta stimata questa distribuzione di riferimento, il valore osservato nei dati reali viene confrontato con i valori derivati dalle permutazioni. Se il valore reale risulta significativamente più elevato rispetto alla distribuzione generata, si conclude che tra le variabili esiste una relazione di dipendenza; se invece il valore osservato rientra nell'intervallo prodotto dalle permutazioni, le variabili possono essere considerate indipendenti.

Si è scelto di utilizzare esclusivamente test non parametrici per la stima delle indipendenze condizionali in coerenza con quanto emerso nel Capitolo 3, dove i test di normalità e le analisi grafiche avevano mostrato distribuzioni fortemente asimmetriche.

I test non parametrici più utilizzati per la stima dell'indipendenza condizionale sono i seguenti:

- *mc-cor*: utilizza la correlazione parziale tra due variabili condizionata su una terza, misurando il legame lineare residuo dopo aver rimosso l'influenza della variabile condizionante;
- *mc-zf*: rappresenta la versione non parametrica del test di Fisher's Z, basata sulla trasformazione Z applicata alla correlazione parziale;
- *mc-mi-g*: calcola la *mutual information*, fornendo un indice di dipendenza più generale e non limitato alla sola componente lineare.

In questo studio è stato adottato il test *mc-cor*, scelta motivata dalla natura continua delle variabili e dall'elevata numerosità campionaria: in queste condizioni, infatti, *mc-cor* garantisce

prestazioni più stabili ed efficienti. Per alleggerire il processo computazionale sono state inoltre limitate a 500 le permutazioni utilizzate e il test è stato applicato su un campione casuale di 10.000 osservazioni.

In virtù di questi elementi, l'adozione dell'algoritmo *pc.stable* combinato con test non parametrici di indipendenza condizionata rappresenta, nel contesto di questa ricerca, la soluzione metodologicamente più adeguata all'apprendimento della rete Bayesiana con indicatori economico-finanziari analizzati.

Inoltre, data la natura continua delle variabili considerate, si farà riferimento a modelli di Gaussian Bayesian Network come descritti nel Capitolo 2.

4.1.1 Rete Bayesiana ultimo anno

L'apprendimento della rete Bayesiana relativa all'ultimo anno contabile ("ultimo_anno") tramite l'algoritmo *pc.stable* ha evidenziato due "archi non diretti". Poiché il modello ha rilevato l'esistenza di una relazione tra le variabili senza determinarne la direzione.

Nello specifico le relazioni coinvolte erano le seguenti:

- Numero di dipendenti ↔ EBITDA
- EBITDA *margin* percentuale ↔ ROS

L'analisi delle dipendenze tramite v-structure non ha evidenziato relazioni sufficientemente definite da consentire l'orientamento automatico degli archi. Per questa ragione, la direzione delle connessioni è stata stabilita sulla base delle logiche economico-finanziarie che regolano le relazioni tra gli indicatori considerati.

Per il primo collegamento, si è scelto di orientare il nodo "numero di dipendenti" come genitore del nodo "EBITDA". Tale scelta deriva dal ruolo del personale come input fondamentale del processo produttivo: un organico più numeroso consente un livello di attività potenzialmente maggiore, che si riflette in una capacità produttiva più elevata. L'aumento della produttività tende, a sua volta, a generare margini operativi più consistenti, influenzando direttamente l'EBITDA. La direzione dell'arco riflette quindi una logica causale riconosciuta nella teoria economico-aziendale.

Per quanto riguarda la seconda relazione, quella tra EBITDA *margin* e ROS, l'orientamento è stato definito considerando la natura degli indicatori. L'EBITDA *margin* misura la marginalità operativa lorda rapportata alle vendite; il ROS, invece, rappresenta una misura più a valle, poiché calcola il risultato operativo dopo ammortamenti e accantonamenti, sempre in rapporto ai ricavi. In termini economici, è quindi più coerente affermare che la marginalità operativa influenza il livello di utile operativo, e non il contrario. Senza adeguati margini lordi, infatti, non può formarsi un utile operativo positivo. Per questo motivo l'arco è stato orientato da EBITDA margin verso ROS.

Una volta fissate manualmente queste relazioni si può visualizzare la struttura della rete Bayesiana degli indicatori di bilancio dell'ultimo anno disponibile come segue:

```
> dag.s_prova1
  Bayesian network learned via Constraint-based methods

  model: [ricavi_delle_vendite_...          19
  arcs:                                     28
    undirected arcs:                        0
    directed arcs:                          28
  average markov blanket size:              4.53
  average neighbourhood size:               2.95
  average branching factor:                 1.47

  learning algorithm:              PC (Stable)
  conditional independence test:     Pearson's Correlation
(MC)
  alpha threshold:                          0.05
  permutations:                             500
  tests used in the learning procedure:     2811
```

L'algoritmo *pc.stable*, tramite test di indipendenza condizionata basati sulla correlazione di Pearson con p-value stimato via permutazioni (metodo Monte Carlo), ha identificato 28 archi tra i 19 nodi considerati.

Oltre alla struttura del grafo, è possibile analizzare una serie di indicatori utili per valutare il livello di connessione della rete.

Tra questi, l'indicatore *Average Markov Blanket size* rappresenta la dimensione media del Markov Blanket di ciascuna variabile, cioè l'insieme minimo di nodi che rende un nodo indipendente dal resto della rete. Nel modello stimato, ogni variabile presenta in media tra i

quattro e i cinque nodi nel proprio Markov Blanket, valore relativamente elevato che evidenzia una rete fortemente interconnessa, caratteristica tipica dei dati economico-finanziari.

L'indicatore *Average Neighbourhood size* misura invece il valore medio del vicinato (neighbourhood size) e indica il numero medio di archi incidenti su ciascun nodo. Risulta intorno a 3, valore inferiore alla Markov Blanket size perché quest'ultima include anche i genitori dei figli (co-genitori).

L'indicatore *Average Branching factor* esprime il numero medio di variabili influenzate da ciascun nodo, cioè il numero di figli. Nel modello ottenuto questo valore è prossimo a due, suggerendo che ogni variabile tende a trasmettere la propria influenza a uno o due nodi successivi.

La rete Bayesiana stimata per l'ultimo anno contabile disponibile è rappresentata in Figura 4.1.

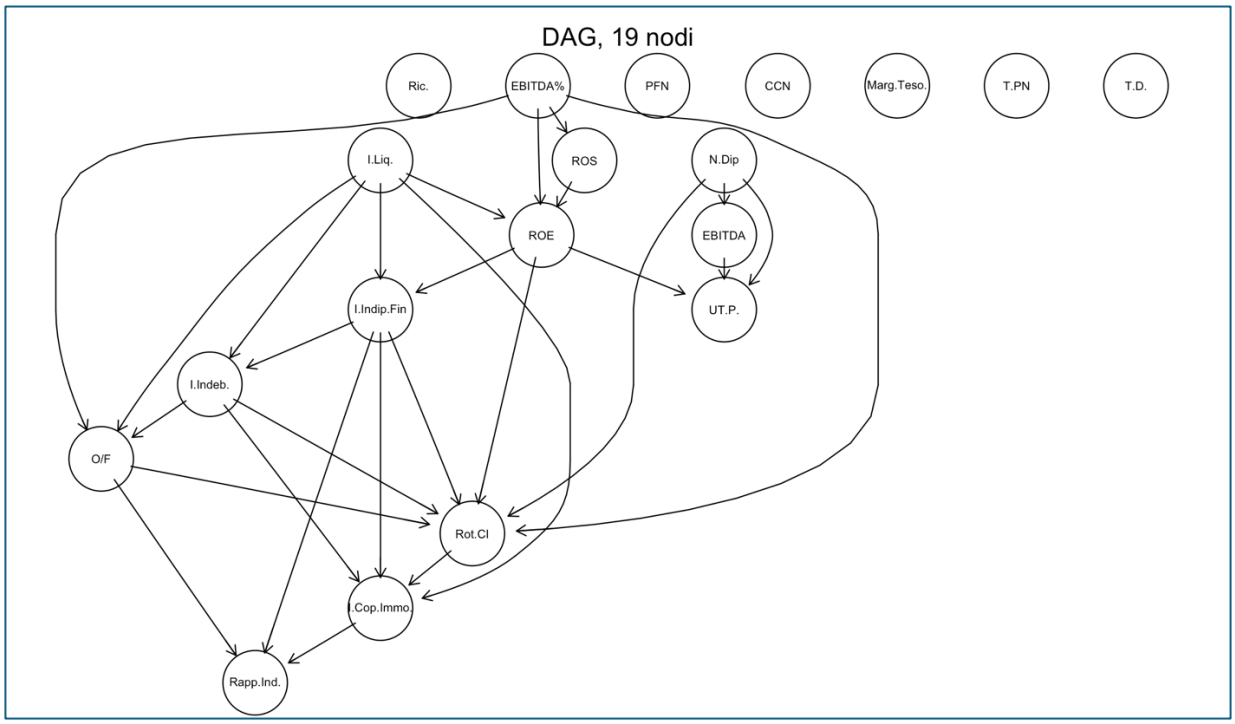


Figura 4.1: Dag a 19 nodi. Fonte: Elaborazione personale su dati AIDA

4.1.2 La Rete Bayesiana nei diversi anni contabili

La fase successiva dell'analisi riguarda la costruzione della struttura della rete che riflette le relazioni di dipendenza e indipendenza tra le variabili, considerando l'evoluzione storica dei dati.

Il processo è stato sviluppato definendo, per ciascun esercizio di riferimento, la rete delle relazioni tra gli indici di bilancio del singolo esercizio.

La scelta di non replicare le relazioni individuate nell'ultimo anno su tutti gli altri livelli, ma di stimare per ogni anno una struttura autonoma, deriva dal presupposto teorico secondo il quale le relazioni tra le variabili economico-finanziarie non sono necessariamente stabili nel tempo, ma possono essere influenzate dal contesto macroeconomico, dalle condizioni di mercato e dal ciclo di vita delle imprese, di conseguenza è coerente pensare che si modifichino di anno in anno.

Per questo motivo, per ognuno degli anni analizzati è stata stimata una rete indipendente tramite l'algoritmo *pc.stable*, così come fatto per l'ultimo anno, in modo da consentire alla struttura del grafo di riflettere correttamente le specificità del periodo storico a cui ciascun set di indici appartiene.

In questa fase è emerso che per l'ultimo anno disponibile erano presenti 19 nodi, ovvero 19 variabili, mentre per gli anni precedenti se ne presentavano solamente 14, a causa della maggiore difficoltà a reperire i dati antecedenti nel tempo.

Di conseguenza si è condotta un'analisi per comprendere se i cinque nodi aggiuntivi dell'ultimo anno comportassero distorsioni nelle relazioni con gli altri indici. Per condurre questa analisi si sono confrontate le relazioni tra i nodi comuni nella rete precedentemente costruita con una nuova rete costruita solamente con i nodi comuni agli altri anni.

Da questa analisi è emerso che non differivano le relazioni tra i nodi, bensì differivano le direzioni in quanto nelle due reti si presentavano come opposte. Nella rete costruita solamente sulle 14 variabili emerge inoltre una relazione che probabilmente nell'altra rete risultava indiretta.

Si riportano in seguito le relazioni interessate,

per la rete costruita con tutte le 19 variabili:

indice_di_liquidita_ultimo_anno_disp	→	indice_di_indebitam_a_lungo_percent_ ultimo_anno_disp
indice_di_liquidita_ultimo_anno_disp	→	oneri_finanz_su_fatt_percent_percent_ ultimo_anno_disp
indice_di_liquidita_ultimo_anno_disp	→	redditivita_del_capitale_proprio_roe_ percent_percent_ultimo_anno_disp
indice_di_copertura_delle_immob_patr imoniale_percent_ultimo_anno_disp	→	rapporto_di_indebitamento_ ultimo_anno_disp
indice_di_indip_finanz_percent_percen t_ultimo_anno_disp	→	indice_di_indebitam_a_lungo_ percent_ultimo_anno_disp

Per la rete costruita solamente sulle 14 variabili comuni:

indice_di_indebitam_a_lungo_percent_ ultimo_anno_disp	→	indice_di_liquidita_ultimo_anno_disp
indice_di_indebitam_a_lungo_percent_ ultimo_anno_disp	→	indice_di_indip_finanz_percent_percen t_ultimo_anno_disp
rapporto_di_indebitamento_ ultimo_anno_disp	→	indice_di_copertura_delle_immob_patri moniale_percent_ultimo_anno_disp
Oneri_finanz_su_fatt_percent_percent_ ultimo_anno_disp	→	indice_di_liquidita_ultimo_anno_disp
ebitda_vendite_percent_percent_ ultimo_anno_disp	→	indice_di_copertura_delle_immob_patri moniale_percent_ultimo_anno_disp
redditivita_del_capitale_proprio_roe_ percent_percent_ultimo_anno_disp	→	indice_di_liquidita_ultimo_anno_disp

Da un'analisi condotta secondo logiche economico-finanziarie, le relazioni della rete costruita con le sole 14 variabili comuni risultano più naturali e più plausibili, di conseguenza d'ora in avanti per l'ultimo anno disponibile, si andranno a togliere le variabili aggiuntive considerando solo quelle comuni agli ultimi anni.

La variazione nella direzione degli archi tra il modello a 19 variabili e quello ridotto a 14 non implica necessariamente una contraddizione nelle relazioni statistiche individuate, bensì la presenza di variabili aggiuntive può modificare le condizioni di indipendenza condizionale e quindi influenzare l'orientamento risultante degli archi pur mantenendo invariata la dipendenza

tra le variabili. L'assenza dei cinque nodi aggiuntivi comporta quindi una riorganizzazione locale delle direzioni, che non altera il significato economico delle relazioni individuate.

La definizione del DAG articolato nei diversi anni fornisce dunque il quadro complessivo delle dipendenze tra gli indici nel tempo. La sezione successiva esamina in modo approfondito il modificarsi delle relazioni identificate.

4.1.3 Analisi orizzontale e verticale: confronto tra le reti

In questo paragrafo si analizza in dettaglio come le relazioni di dipendenza e indipendenza tra gli indicatori si modifichino nel corso dell'evoluzione storica dei dati, cercando di individuare i pattern evolutivi. Viene inoltre sviluppato un confronto tra le strutture rilevate nelle reti costruite sulle imprese fisiologicamente sane e quelle ottenute considerando esclusivamente le imprese problematiche, con l'obiettivo di cogliere sia le dinamiche temporali delle relazioni sia i cambiamenti strutturali che si manifestano quando gli indicatori economico-finanziari iniziano a riflettere condizioni di instabilità aziendale.

Andamento temporale delle imprese fisiologicamente sane:

Per comprendere come si modificano nel tempo le relazioni di dipendenza e di indipendenza tra gli indicatori economico-finanziari, è necessario analizzare in modo sistematico gli archi delle reti Bayesiane.

In questa sezione vengono esaminate esclusivamente le reti stimate sulle imprese sane, allo scopo di individuare le relazioni fisiologiche, ossia quelle che caratterizzano il funzionamento ordinario dell'impresa. La loro evoluzione rappresenta un punto di riferimento per comprendere, nel seguito, come tali meccanismi possano deteriorarsi nelle imprese problematiche.

Si riporta una tabella riassuntiva (*Tabella 4.1*) per comprendere in maniera schematica come è costruita la rete:

Livello della rete	Descrizione	Numero di nodi nel livello	Numero di archi nel livello
1	Indici di bilancio relativi all'ultimo anno disponibile	14	24
2	Indici di bilancio relativi all'anno n-1	14	24
3	Indici di bilancio relativi all'anno n-2	14	26
4	Indici di bilancio relativi all'anno n-3	14	27
Totale		56	101

Tabella 4.1: Struttura della Rete Bayesiana multilivello, imprese sane. Fonte: elaborazione personale su dati AIDA

Il primo passo consiste nell'identificare le dipendenze che compaiono in tutti i livelli temporali analizzati, poiché esse costituiscono la struttura stabile della rete, ovvero quell'insieme di relazioni che non subisce modifiche nel tempo. Le relazioni persistenti risultano essere:

`indice_di_liquidita` → `indice_di_copertura_delle_immob_ patrimoniale_ percent`

`indice_di_indebitam_a_lungo_percent` → `indice_di_copertura_delle_immob_patrimoniale_percent`

`indice_di_indebitam_a_lungo_percent` → `rotaz_cap_investito_volte`

`oneri_finanz_su_fatt_percent_percent` → `rotaz_cap_investito_volte`

`ebitda_vendite_percent_percent` → `rotaz_cap_investito_volte`

`ebitda_vendite_percent_percent` → `redditivita_del_capitale_proprio_roe_percent_percent`

Questi archi rappresentano meccanismi di dipendenza fisiologica consolidati nel tempo e descrivono il comportamento strutturale tipico delle imprese sane, nel quale si mantiene un equilibrio tra liquidità, redditività, struttura patrimoniale ed efficienza operativa. Le relazioni individuate riflettono infatti aspetti fondamentali della solidità aziendale: la dipendenza tra liquidità o livello di indebitamento e la copertura dell'attivo immobilizzato evidenzia l'importanza, per un'impresa sana, di sostenere gli investimenti di lungo periodo con risorse adeguate. Analogamente, la marginalità operativa, misurata tramite l'EBITDA,

influenza in modo determinante la redditività del capitale, confermando il ruolo centrale della performance operativa nella generazione di valore.

Successivamente si analizzano le restanti relazioni, osservando i cambiamenti che si verificano tra gli archi nei diversi anni. Si rileva che alcune dipendenze, presenti nei periodi precedenti, scompaiono nell'ultimo anno; nello specifico si tratta dei seguenti archi:

- Rapporto di indebitamento → indice di indipendenza finanziaria
- Oneri finanziari su fatturato → indice di indebitamento a lungo
- Oneri finanziari su fatturato → redditività del capitale proprio, ROE

Allo stesso modo, vi sono relazioni che compaiono nel terzo anno antecedente, poi scompaiono, per ricomparire nuovamente in un periodo successivo. È il caso, ad esempio, della relazione tra l'indice di indebitamento a lungo termine e l'indice di liquidità: tale relazione si è manifestata nel terzo esercizio antecedente a quello considerato, per poi sparire nei due esercizi successivi e riapparire nell'esercizio più recente analizzato.

Questa variabilità riflette il fatto che gli indicatori economico-finanziari sono sensibili anche a condizioni di contesto esterne alla struttura aziendale.
Dinamiche macroeconomiche, cambiamenti strutturali nel settore di appartenenza, interventi di politiche economiche, variazioni nei tassi d'interesse, nonché più in generale situazioni di instabilità sociale o economica, possono modificare l'intensità e la natura delle relazioni tra gli indicatori interni all'impresa.

Di conseguenza, l'algoritmo di apprendimento può ritenere non più necessaria una relazione che in anni precedenti appariva diretta. Questo accade perché, al mutare delle condizioni esterne e interne, altre variabili diventano più rilevanti a livello informativo e permettono di spiegare quella stessa dipendenza in modo indiretto. In tali casi, l'arco viene rimosso poiché risulta ridondante dal punto di vista probabilistico.

La presenza alternata di alcune relazioni nel tempo va dunque interpretata come un adattamento della struttura della rete alle modificazioni del contesto economico e del sistema di interdipendenze tra variabili, più che come un segnale di cambiamento della solidità delle imprese analizzate.

Si riporta in *Appendice I* la legenda esplicativa degli acronimi utilizzati per i nodi.

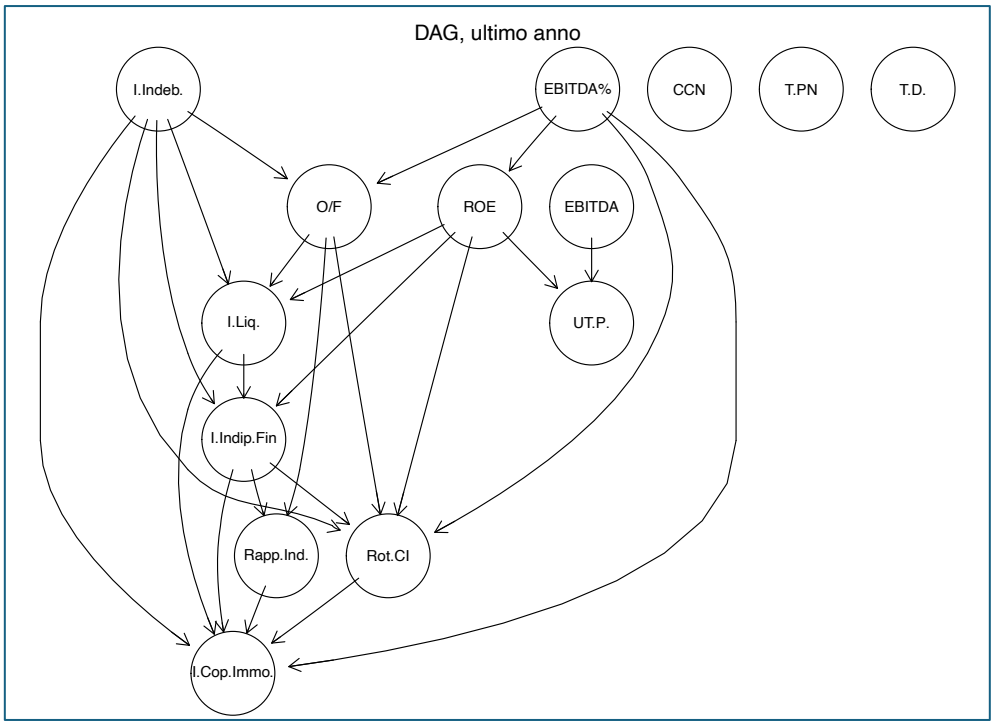


Figura 4.2: DAG imprese sane – ultimo anno disponibile. Fonte: elaborazione personale su dati AIDA

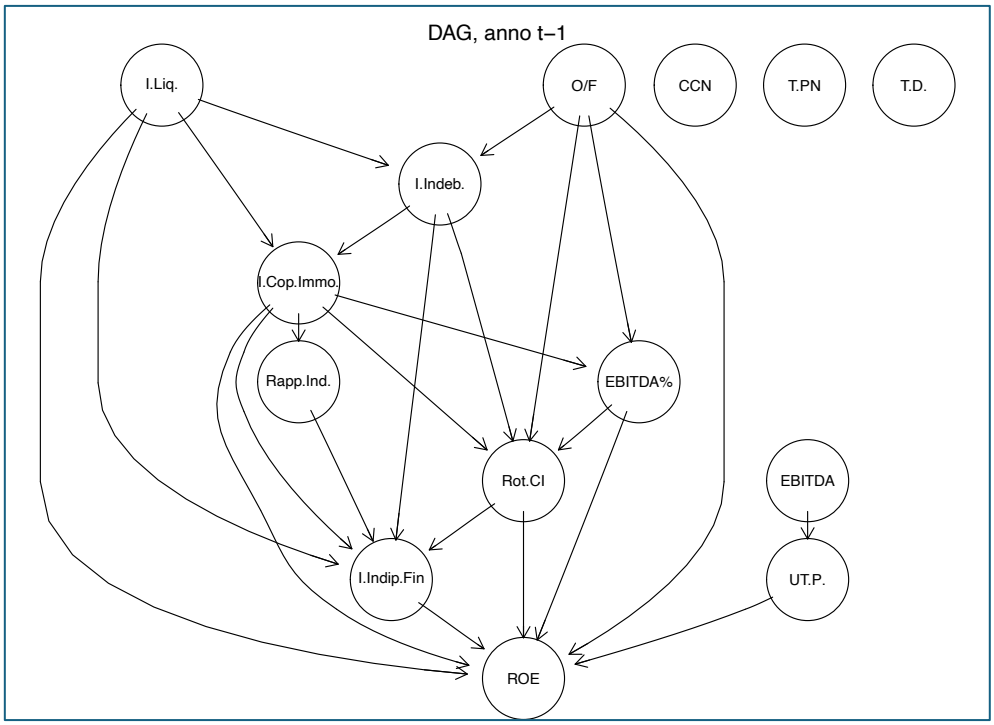


Figura 4.3:imprese sane – anno t-1. Fonte: elaborazione personale su dati AIDA

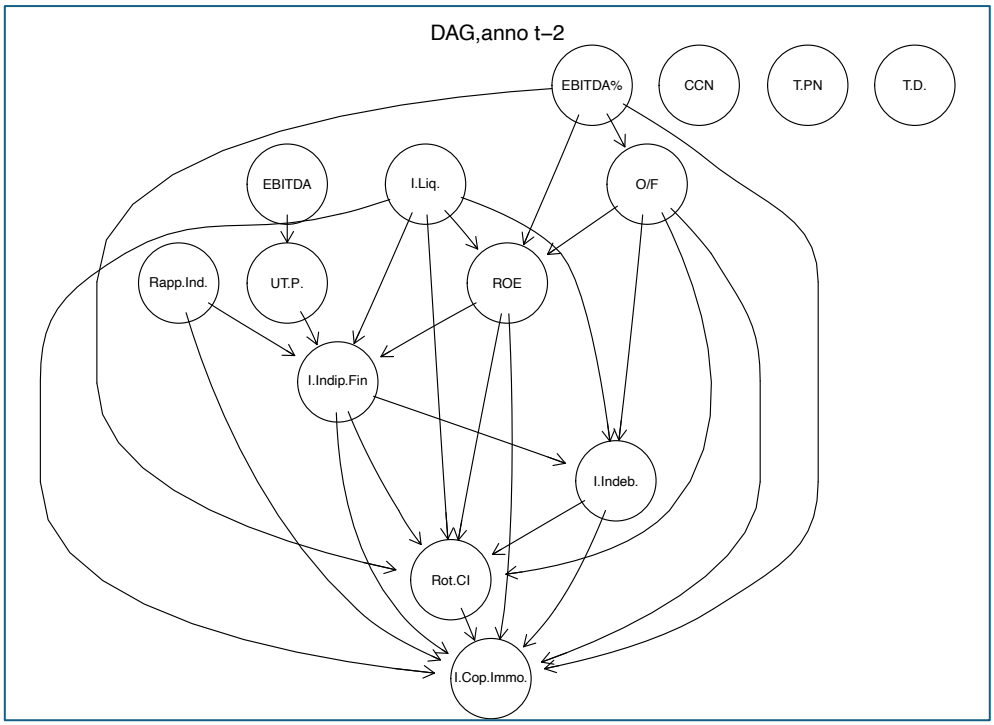


Figura 4.4: imprese sane – anno t-2. Fonte: elaborazione personale su dati AIDA

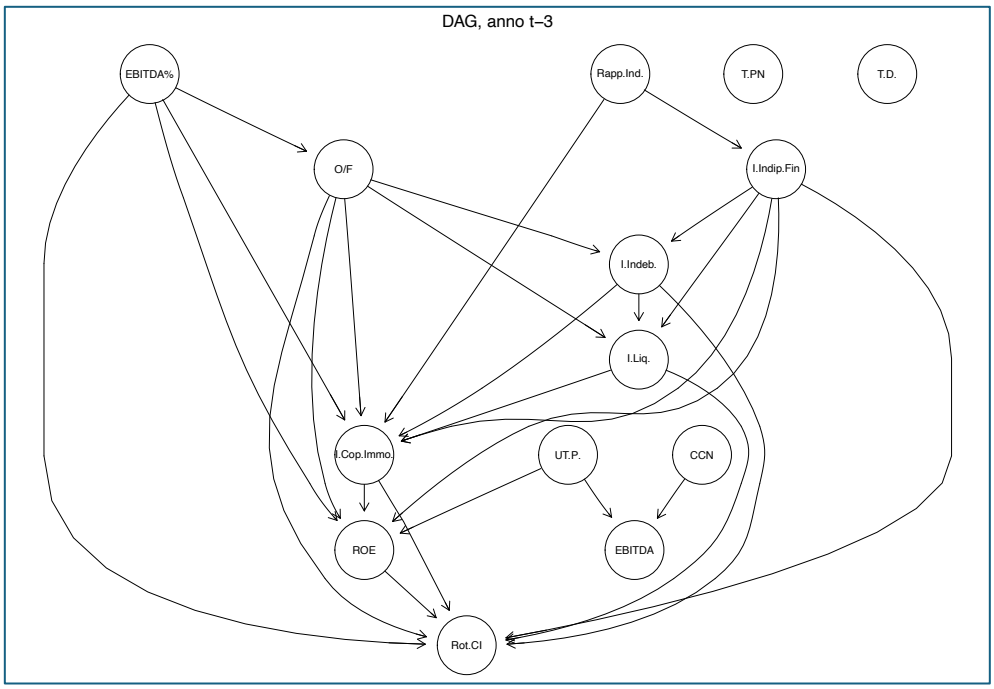


Figura 4.5: imprese sane – anno t-3. Fonte: elaborazione personale su dati AIDA

Andamento temporale delle imprese che hanno sviluppato squilibri:

Per quanto riguarda le imprese che hanno sviluppato squilibri economico-finanziari la struttura della rete Bayesiana si presenta come segue:

Livello della rete	Descrizione	Numero di nodi nel livello	Numero di archi nel livello
1	Indici di bilancio relativi all'ultimo anno disponibile	14	29
2	Indici di bilancio relativi all'anno n-1	14	26
3	Indici di bilancio relativi all'anno n-2	14	30
4	Indici di bilancio relativi all'anno n-3	14	22
Totale		56	107

Tabella 4.2: Struttura della Rete Bayesiana multilivello, imprese problematiche. Fonte: elaborazione personale su dati AIDA.

Analizzando lo sviluppo delle relazioni tra queste imprese si evidenzia come le relazioni permanenti presenti in tutti gli esercizi analizzati siano in misura significativamente inferiore; infatti, si evidenziano solo le seguenti relazioni:

```
indice_di_liquidita          → indice_di_indip_finanz_percet
                               _percent
indice_di_liquidita          → redditivita_del_capitale_
                               proprio_roe_percent_percent
rotaz_cap_investito_volte    → ebitda_vendite_percent_
                               percent
```

Queste relazioni più che identificare una sorta di solidità identificano relazioni di dipendenza, individuando quali sono le variabili che più dipendono tra loro in un modo che non si modifica nel tempo.

Le relazioni importanti da sottolineare sono quelle che si verificano solamente nell'ultimo esercizio analizzato e che non si sono mai presentate negli esercizi precedenti. Sono particolarmente significative perché identificano quelle che sono le relazioni che individuano segnali di squilibri e situazioni di crisi, tra queste individuiamo:

```
indice_di_copertura_delle_immob_patrimoniale_percent → indice_di_indip_finanz_percent_percent
indice_di_copertura_delle_immob_patrimoniale_percent → rotaz_cap_investito_volte
oneri_finanz_su_fatt_percent → capitale_circolante_netto_migl_eur
indice_di_indip_finanz_percent_percent → indice_di_indebitam_a_lungo_percent
indice_di_indip_finanz_percent_percent → ebitda_vendite_percent_percent
rotaz_cap_investito_volte → indice_di_indebitam_a_lungo_percent
ebitda_migl_eur → redditivita_del_capitale_proprio_roe_percent_percent
ebitda_migl_eur → capitale_circolante_netto_migl_eur
redditivita_del_capitale_proprio_roe_percent_percent → rotaz_cap_investito_volte
redditivita_del_capitale_proprio_roe_percent_percent → ebitda_vendite_percent_percent
capitale_circolante_netto_migl_eur → indice_di_indebitam_a_lungo_percent
capitale_circolante_netto_migl_eur → utile_perdita_di_esercizio_migl_eur
utile_perdita_di_esercizio_migl_eur → ebitda_vendite_percent_percent
```

Queste relazioni patologiche mettono in evidenza le fragilità tipiche delle imprese in stato di difficoltà. In particolare, si osserva come l'attivo immobilizzato finisca per creare una dipendenza diretta sia con l'indipendenza finanziaria sia con la struttura patrimoniale, segnalando un evidente squilibrio nella proporzione tra investimenti e attività.

Dall'analisi emerge inoltre che gli oneri finanziari incidono sul capitale circolante netto, andando di fatto ad assorbire risorse operative e riducendo la capacità dell'impresa di sostenere i fabbisogni di gestione corrente.

Un ulteriore elemento critico è rappresentato dalla presenza di relazioni con verso invertito rispetto alla rete delle imprese sane, che rivelano dinamiche non fisiologiche. Tra queste risultano particolarmente significative:

- ROE → rotazione del capitale investito
- ROE → EBITDA vendite

Tali inversioni indicano che la redditività non è più il risultato dell'efficienza operativa, ma ne diventa il determinante anomalo, riflettendo una condizione di deterioramento della coerenza economico-finanziaria complessiva.

Un ulteriore segnale di criticità emerge dalla relazione tra utile e EBITDA. Nelle imprese in equilibrio, l'utile rappresenta l'esito finale delle dinamiche operative e finanziarie, collocandosi quindi a valle della catena delle relazioni. Nel caso delle imprese problematiche, invece, l'utile diventa una variabile in grado di influenzare direttamente la marginalità operativa: ciò indica che nei risultati d'esercizio non si riflette soltanto la redditività corrente, ma anche componenti negative accumulate nel tempo.

La presenza di tali passività, tipiche di perdite accumulate negli esercizi precedenti, finisce per comprimere progressivamente l'EBITDA, erodendo la capacità dell'impresa di generare margini operativi e segnalando un deterioramento strutturale del modello economico.

Un'ulteriore evidenza rilevante riguarda la comparsa di archi che collegano aree aziendali normalmente distinte. Questo intreccio anomalo tra dimensioni patrimoniali, finanziarie e operative segnala la perdita di specializzazione funzionale degli indicatori e riflette un progressivo deterioramento della coerenza interna del modello di business.

Tali configurazioni, non presenti nelle imprese sane, costituiscono veri e propri segnali anticipatori di squilibrio: mostrano infatti che l'impresa sta entrando in una fase in cui le determinanti economico-finanziarie non seguono più schemi ordinati e prevedibili, ma reagiscono secondo dinamiche non lineari e difficilmente controllabili.

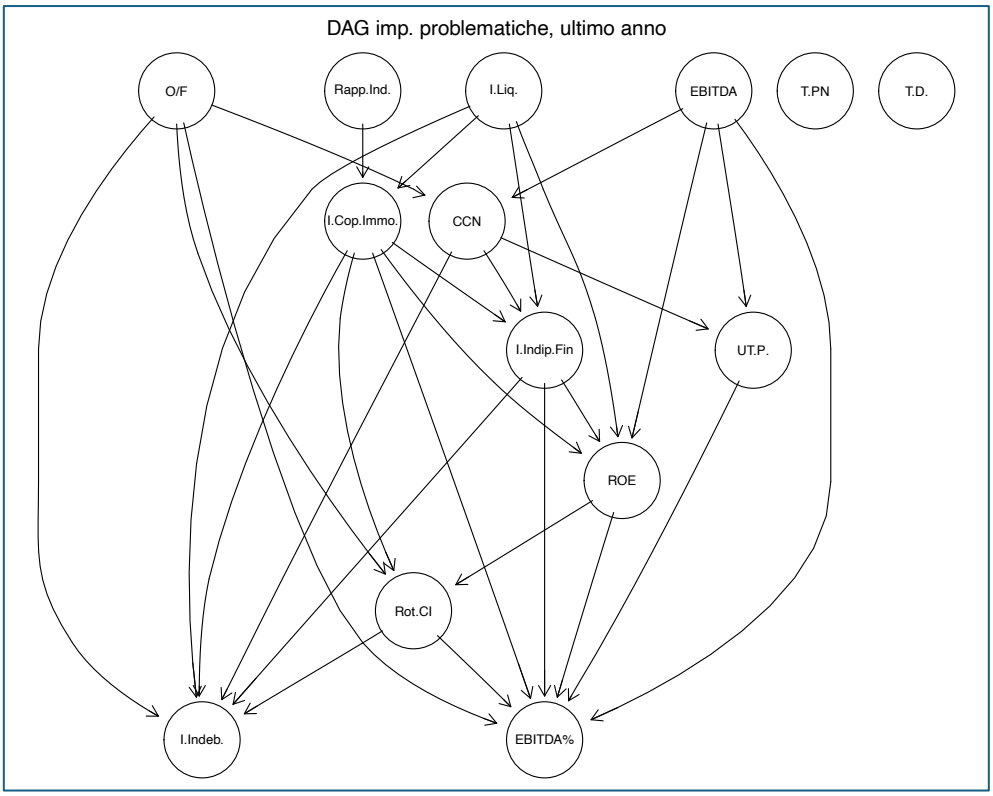


Figura 4.6: DAG imprese problematiche – ultimo anno disp. Fonte: elaborazione personale su dati AIDA

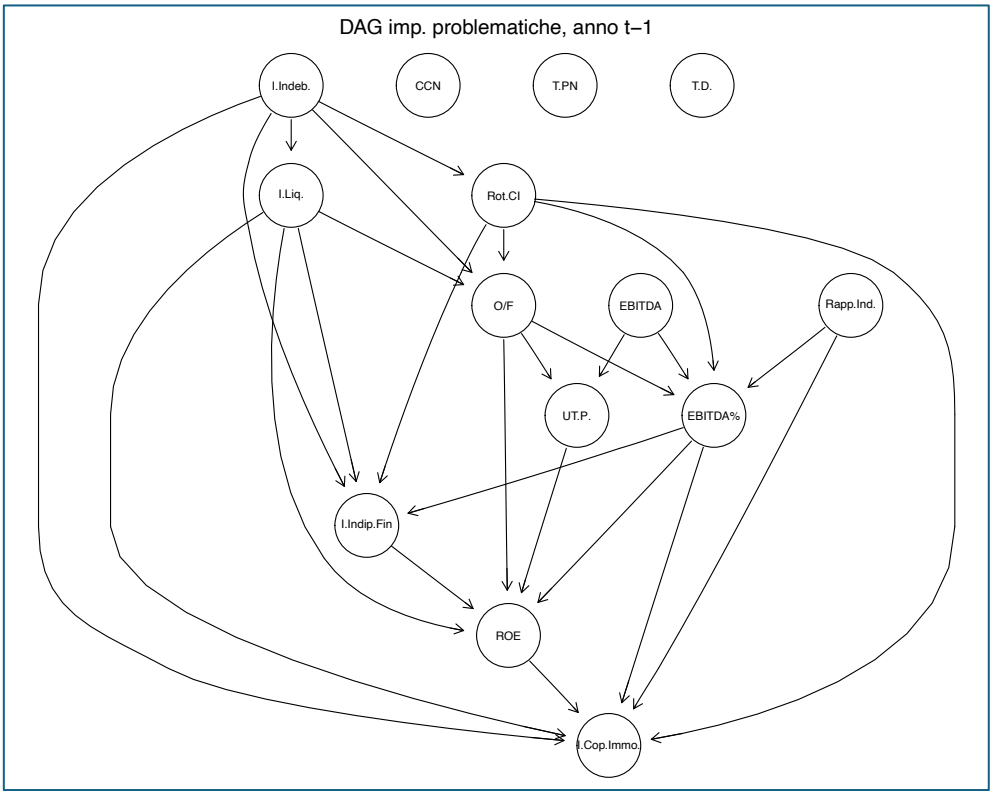


Figura 4.7: DAG imprese problematiche – anno t-1. Fonte: elaborazione personale su dati AIDA

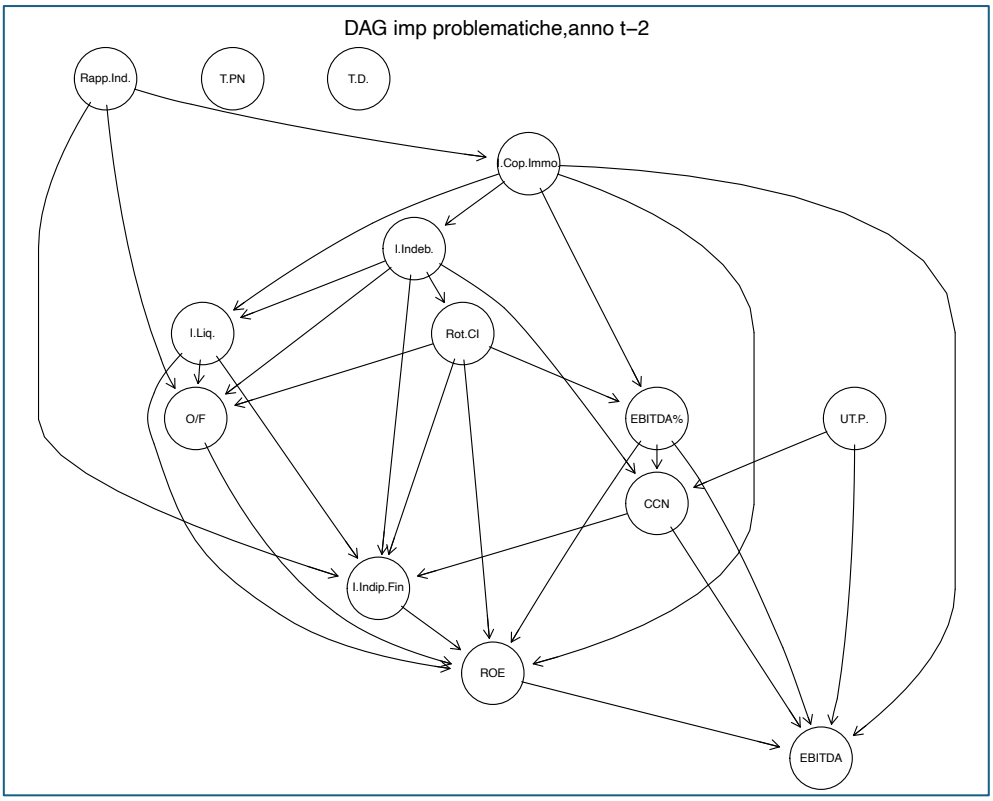


Figura 4.8:DAG imprese problematiche – anno t-3. Fonte: elaborazione personale su dati AIDA

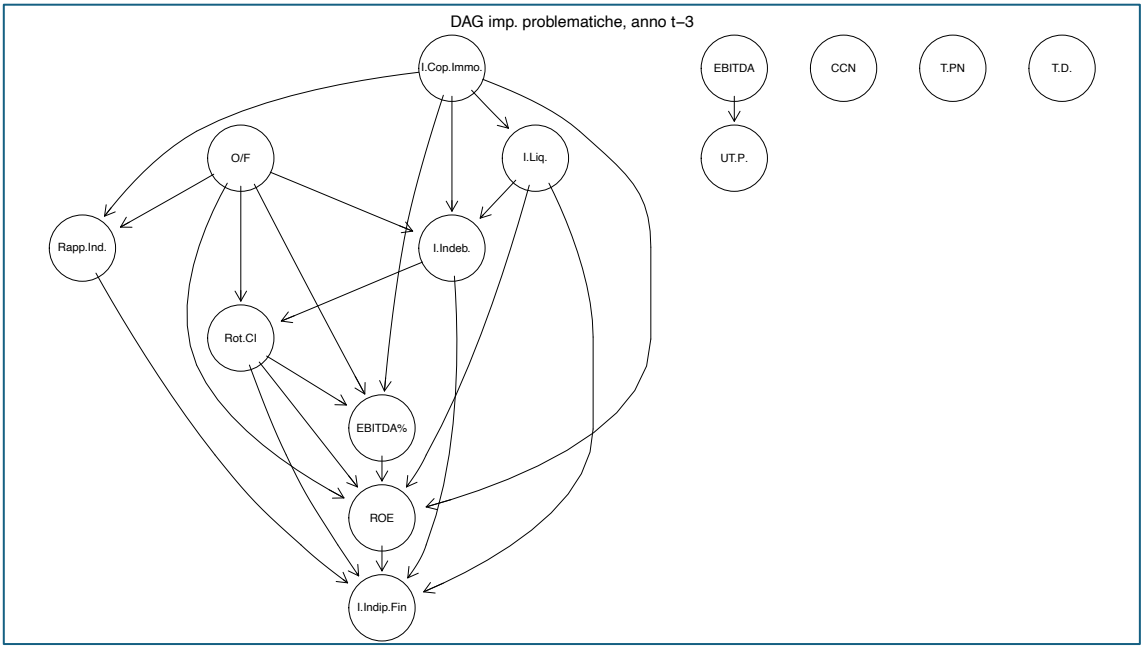


Figura 4.9: DAG imprese problematiche – anno t-3. Fonte: elaborazione personale su dati AIDA

Confronto tra rete fisiologica e rete patologica nell'ultimo anno:

In questa sezione si confrontano le relazioni presenti nelle reti delle imprese sane con quelle rilevate nelle imprese che hanno sviluppato problematiche di continuità aziendale, utilizzando l'ultimo esercizio disponibile. Tale anno è considerato il più rappresentativo poiché offre l'immagine più recente e incisiva delle dinamiche che caratterizzano la salute economico-finanziaria dell'impresa.

Dall'analisi emergono alcune relazioni che compaiono in entrambe le reti, indipendentemente dallo stato di continuità aziendale. Tali dipendenze, che risultano dunque comuni e non influenzate dalla condizione di crisi, possono essere interpretate come relazioni strutturali, ovvero meccanismi fondamentali e ricorrenti nel funzionamento delle organizzazioni. Le relazioni condivise sono le seguenti:

`indice_di_liquidita_ultimo_anno_disp`	→	`indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp`
`indice_di_liquidita_ultimo_anno_disp`	→	`indice_di_indip_finanz_percent_percent_ultimo_anno_disp`
`rapporto_di_indebitamento_ultimo_anno_disp`	→	`indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp`
`oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp`	→	`rotaz_cap_investito_volte_ultimo_anno_disp`
`ebitda_migl_eur_ultimo_anno_disp`	→	`utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp`
`redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp`	→	`rotaz_cap_investito_volte_ultimo_anno_disp`

La persistenza di tali relazioni in entrambi i modelli indica che alcuni legami economico-finanziari rimangono validi a prescindere dalla presenza o meno di segnali di crisi. In particolare, il fatto che l'indice di liquidità influenzi sia la copertura dell'attivo immobilizzato sia l'indipendenza finanziaria evidenzia che, in qualsiasi condizione, la liquidità svolge un ruolo cruciale nel sostenere gli investimenti durevoli e nel contribuire alla stabilità della struttura finanziaria.

Si confermano inoltre relazioni strutturali come quella tra EBITDA e utile/perdita di esercizio, che riflette il normale processo di formazione del reddito, indipendentemente dalla solidità o dalla fragilità dell'impresa.

Si prosegue con l'analisi delle relazioni maggiormente significative per la distinzione tra i due gruppi di imprese. I nodi considerati derivano dall'analisi svolta nel terzo capitolo, finalizzata a individuare quali variabili risultassero maggiormente discriminanti tra il gruppo delle imprese sane e quello delle imprese che presentano problematiche di continuità. Tale individuazione è stata condotta attraverso l'applicazione del *t.test* e del test non parametrico di *Mann–Whitney*, che hanno permesso di evidenziare in modo statisticamente significativo le variabili più efficaci nel differenziare i due insiemi.

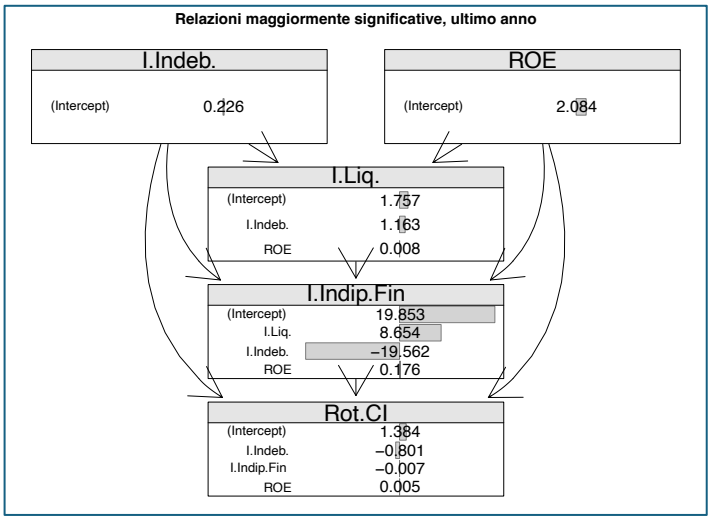


Figura 4.10: Rete Bayesiana parametrizzata delle imprese sane - ultimo anno disponibile.
Fonte: elaborazione personale su dati AIDA

Per la costruzione dei grafici presentati in questa sezione è stato necessario stimare le probabilità condizionate per ciascuna variabile del database. A tal fine è stato utilizzato l'algoritmo `bn.fit` di R, adottando il metodo "mle-g" per la stima dei parametri in presenza di variabili continue.

Il calcolo dei parametri per tutti i nodi della rete ha permesso di generare i grafici riportati nelle Figure 4.9 e 4.10, nei quali sono rappresentate le equazioni di regressione stimate per le variabili ritenute maggiormente significative nella discriminazione tra i due gruppi. Tali grafici rendono immediatamente visibile la natura e l'intensità delle relazioni

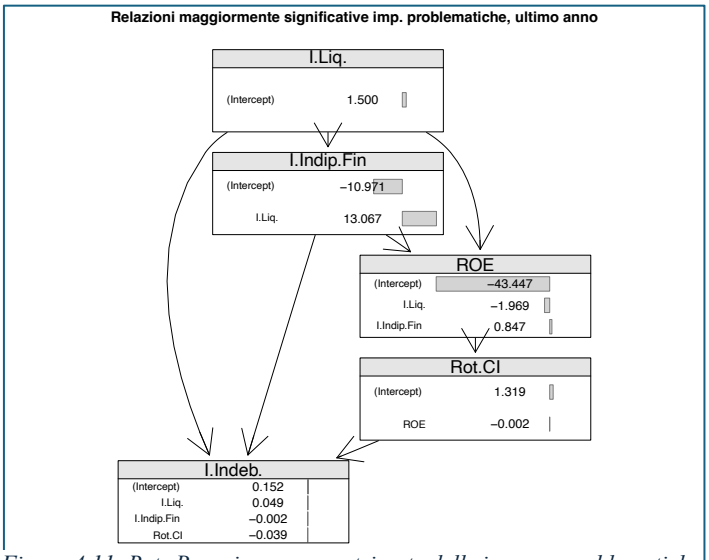


Figura 4.11: Rete Bayesiana parametrizzata delle imprese problematiche - ultimo anno disponibile.
Fonte: elaborazione personale su dati AIDA

identificate all'interno della struttura Bayesiana, mostrando per ciascun nodo il contributo dei rispettivi genitori.

È opportuno sottolineare che queste rappresentazioni grafiche illustrano solo un sottoinsieme selezionato di variabili, scelto in quanto maggiormente rappresentativo delle differenze tra imprese sane e imprese con problematiche di continuità aziendale. Le analisi che seguono non possono quindi sostituire integralmente l'esame dell'intera rete Bayesiana, poiché una selezione parziale comporta inevitabilmente la perdita di alcune relazioni informative presenti nel modello completo.

Analizzando l'indice di liquidità, indicatore che si è confermato centrale in tutte le analisi svolte, emerge che nelle imprese sane esso risulta influenzato da ulteriori variabili quali l'indice di indebitamento e il ROE. La relazione stimata può essere espressa attraverso la seguente equazione di regressione:

$$Indice\ di\ liquidità = 1{,}757 + 1{,}163 \cdot I.Indebitamento + 0{,}008 \cdot ROE$$

Questa equazione mostra che, in un contesto di imprese fisiologicamente sane, l'indice di liquidità dipende dai propri nodi genitori. L'intercetta indica che, in assenza degli effetti delle variabili esplicative (quindi se i valori dei nodi genitori sono pari a zero), l'indice di liquidità assumerebbe un valore pari a 1,757. Inoltre, un aumento di un punto dell'indice di indebitamento determina un incremento di 1,163 punti nell'indice di liquidità, suggerendo che le imprese finanziariamente solide adottano meccanismi di prudenza, accrescendo la propria liquidità al crescere del debito. Anche il ROE esercita un effetto positivo, seppur marginale: un incremento di un punto nel ROE comporta un aumento di 0,008 punti nell'indice di liquidità.

Diversamente, per le imprese problematiche l'indice di liquidità non risulta influenzato da nessuna delle variabili considerate, presentando unicamente un'intercetta pari a 1,5

In entrambi i gruppi si osserva una relazione dall'indice di liquidità verso l'indice di indipendenza finanziaria. Si tratta di una delle relazioni strutturali individuate nel corso dell'analisi, ovvero un legame che permane indipendentemente dallo stato di salute dell'impresa. Pur mantenendo la stessa direzione, l'intensità dell'effetto risulta tuttavia molto diversa tra imprese sane e imprese problematiche, mettendo in evidenza dinamiche finanziarie differenti nei due contesti.

Per le imprese sane, l'equazione di regressione dell'indice di indipendenza finanziaria è:

$$I.Indip.Fin. = 19{,}853 + 8{,}654 \cdot I.Liquidità - 19{,}562 \cdot I.Indebitamento + 0{,}176 \cdot ROE$$

La relazione evidenzia un forte effetto positivo della liquidità: un aumento di un punto dell'indice di liquidità comporta un incremento di 8,654 punti dell'indipendenza finanziaria, riflettendo come una maggiore disponibilità di risorse liquide rafforzi la capacità dell'impresa di autofinanziarsi. L'indice di indebitamento mostra invece un effetto marcatamente negativo: un aumento di un punto riduce l'indipendenza finanziaria di 19,562 punti, indicando come il ricorso al capitale di terzi comprometta la solidità patrimoniale. Anche il ROE esercita un effetto positivo, sebbene marginale: un incremento di un punto comporta un aumento di 0,176 dell'indipendenza finanziaria, poiché una maggiore redditività consente all'impresa di dipendere meno da finanziamenti esterni. Nel complesso, questa equazione descrive un sistema finanziariamente equilibrato, in cui liquidità, indebitamento e redditività concorrono in maniera coerente alla determinazione dell'indipendenza finanziaria.

Per le imprese problematiche, l'equazione assume invece una forma molto più semplice:

$$I.Indip.Fin. = -10{,}971 + 13{,}067 \cdot I.Liquidità$$

In questo caso, l'intercetta negativa indica che, in assenza di liquidità, l'indice di indipendenza finanziaria sarebbe già inferiore a zero, segnalando una struttura patrimoniale squilibrata sin dall'origine e caratterizzata da una prevalenza di capitale di terzi. Ciò contrasta nettamente con quanto osservato nelle imprese sane, che presentano invece un'intercetta positiva coerente con la normale configurazione iniziale di un'impresa finanziata mediante mezzi propri. Inoltre, nelle imprese problematiche l'indipendenza finanziaria risulta determinata esclusivamente dal livello di liquidità, con un coefficiente particolarmente elevato. L'assenza di relazioni con l'indebitamento e con la redditività suggerisce un quadro molto più instabile, in cui la capacità di mantenere autonomia finanziaria dipende quasi unicamente dalle disponibilità liquide, evidenziando un'elevata precarietà strutturale.

In sintesi, mentre nelle imprese sane l'indipendenza finanziaria deriva dall'equilibrio tra liquidità, indebitamento e redditività, nelle imprese problematiche essa è legata quasi esclusivamente alla liquidità, riflettendo una struttura finanziaria più fragile e meno articolata.

In conclusione, il confronto tra le reti Bayesiane delle imprese sane e di quelle problematiche rivela che emergono differenze significative nella complessità e nell'intensità dei meccanismi che governano gli equilibri economico-finanziari.

Queste evidenze confermano come le reti Bayesiane non solo descrivono le relazioni tra gli indicatori, ma sono in grado di mettere in luce il deterioramento progressivo dei legami. Tale deterioramento caratterizza il passaggio da una condizione di normalità a una situazione di crisi, e le reti Bayesiane forniscono così una chiave interpretativa utile per comprendere i meccanismi alla base del declino della struttura aziendale e dell'avvicinarsi ad una situazione di squilibrio strutturale.

4.1.4 Il nodo continuità: la rete previsionale

Una volta comprese le differenze nelle relazioni che caratterizzano le imprese sane rispetto a quelle problematiche, in questo paragrafo si procede alla costruzione della rete definitiva, che costituirà la base per lo sviluppo del modello previsionale presentato nella sezione successiva.

Per ottenere una rete previsionale in grado di stimare correttamente lo stato di salute di un'impresa, è necessario che il database utilizzato per l'apprendimento contenga osservazioni appartenenti a entrambi i livelli della variabile target, ossia la variabile *continuità*, che distingue tra imprese sane e imprese che hanno sviluppato problematiche di continuità aziendale. La presenza di entrambe le condizioni è fondamentale affinché il modello possa riconoscere i pattern tipici delle imprese solide così come quelli associati a situazioni di squilibrio.

La costruzione della rete previsionale definitiva è stata condotta utilizzando esclusivamente i dati dell'ultimo esercizio disponibile. Tale scelta è motivata dal fatto che gli indicatori più prossimi alla perdita di continuità risultano maggiormente informativi rispetto agli anni precedenti e, come evidenziato dall'analisi descrittiva del Capitolo 3, i principali segnali di deterioramento emergono proprio nell'ultimo esercizio antecedente la cessazione dell'attività.

Per definire correttamente la struttura della rete è stato necessario individuare anche il posizionamento più appropriato del nodo che rappresenta la continuità aziendale. In linea teorica, tale nodo potrebbe essere modellato come genitore di tutte le variabili, seguendo la logica secondo cui uno stato di salute sano o problematico determinerebbe l'andamento degli indicatori economico-finanziari. In alternativa, il nodo continuità può essere trattato come figlio, assumendo che siano gli indici di bilancio a generare configurazioni informative

da cui emerge lo stato finale dell'impresa, che può concretizzarsi in una condizione di normalità operativa oppure sfociare in procedure volte alla salvaguardia della continuità aziendale.

Per affrontare questa scelta strutturale si è deciso di costruire due reti distinte: una con il nodo continuità posto come figlio e l'altra come genitore, al fine di valutare quale configurazione risultasse più rappresentativa e funzionale rispetto al fenomeno analizzato.

La prima rete è stata costruita imponendo una *blacklist* che vietasse al nodo continuità di essere genitore di qualsiasi altra variabile. Tale vincolo riflette l'ipotesi modellistica secondo cui la continuità aziendale costituisce un *outcome* finale e non può esercitare dipendenze causali sugli indicatori economico-finanziari. Di conseguenza, nella struttura appresa il nodo continuità può comparire unicamente come nodo figlio.

In questa configurazione la variabile continuità è stata trattata come continua, codificata con valore 1 per le imprese sane e 0 per quelle problematiche. Tale scelta è necessaria perché, nelle reti ibride di tipo *Conditional Gaussian*, un nodo continuo non può essere genitore di un nodo discreto, mentre può correttamente assumerne il ruolo di figlio.

La seconda rete è stata invece costruita con l'obiettivo opposto: forzare il nodo continuità a essere esclusivamente un nodo genitore. In questo caso la variabile è rimasta categoriale a due livelli, poiché nella modellizzazione *Conditional Gaussian* i nodi discreti possono essere genitori di nodi continui senza violare le assunzioni del modello. Anche in questo caso è stata predisposta una *blacklist*, ma con la funzione inversa: impedire che il nodo continuità fosse figlio di altri nodi.

La struttura è stata quindi appresa mediante il metodo "mle-cg", appropriato per database che contengono congiuntamente variabili continue e categoriali.

Successivamente, si è reso necessario comprendere quale delle due reti fosse maggiormente appropriata ai fini della ricerca. Per questo motivo sono stati sviluppati due diversi test di confronto, finalizzati a valutare rispettivamente la capacità descrittiva e la capacità previsionale.

Il primo test effettuato riguarda il calcolo degli *score* di struttura, funzioni matematiche utilizzate per confrontare reti Bayesiane alternative. Gli score misurano la capacità della rete di rappresentare correttamente il database, penalizzando al tempo stesso la complessità della struttura. A parità di condizioni, il modello preferibile è quello con valore di score più elevato,

poiché rappresenta il miglior compromesso tra accuratezza e complessità. Gli score ottenuti sono i seguenti:

- Score del DAG con nodo continuità come figlio: - 8.494.879
- Score del DAG con nodo continuità come genitore: - 8.364.684

Sebbene entrambi i valori risultino negativi, poiché −8.364.684 è meno negativo (cioè più elevato) rispetto a −8.494.879, secondo questo criterio la rete che meglio rappresenta i dati è quella con il nodo continuità posto come genitore.

Il secondo test è stato progettato per comprendere quale delle due reti offrisse migliori prestazioni in termini di previsione della continuità aziendale.

Per prima cosa, è stato eliminato il nodo continuità dai database utilizzati per il test, in modo da consentire al modello di stimarne il valore attraverso la funzione *predict*, senza conoscere l'effettiva risposta. In questo modo, si sono potuti confrontare i valori osservati nel database originario con i valori predetti dal modello. La funzione utilizzata per questo confronto è il Mean Squared Error (MSE), che misura l'errore quadratico medio tra valori reali e valori stimati.
Un MSE più basso indica una migliore accuratezza predittiva.

I risultati ottenuti sono:

- MSE del DAG con nodo continuità come figlio: 0, 1974965
- MSE del DAG con nodo continuità come genitore: 0, 249715

I valori mostrano chiaramente che il modello con continuità come figlio è quello che prevede con maggiore precisione lo stato dell'impresa, generando un errore medio inferiore.

A questo punto, la scelta della rete da adottare è stata guidata dall'obiettivo principale della ricerca. Lo scopo dello studio è infatti quello di sviluppare un modello capace di stimare, sulla base degli indicatori economico-finanziari disponibili, la probabilità che un'impresa manifesti uno stato di squilibrio oppure mantenga una condizione di continuità operativa. In quest'ottica, la caratteristica fondamentale richiesta alla rete è la capacità predittiva nei confronti della variabile continuità.

Poiché il test basato sull'MSE ha evidenziato che la rete con il nodo continuità posto come figlio fornisce previsioni più accurate, essa risulta la configurazione strutturale più coerente con le finalità dello studio.

Di conseguenza, la struttura del DAG utilizzata per la stima del modello sarà una rete Bayesiana continua in cui il nodo continuità è collocato a valle della rete, esclusivamente come nodo figlio, caratterizzato da soli archi in ingresso e privo di archi in uscita.

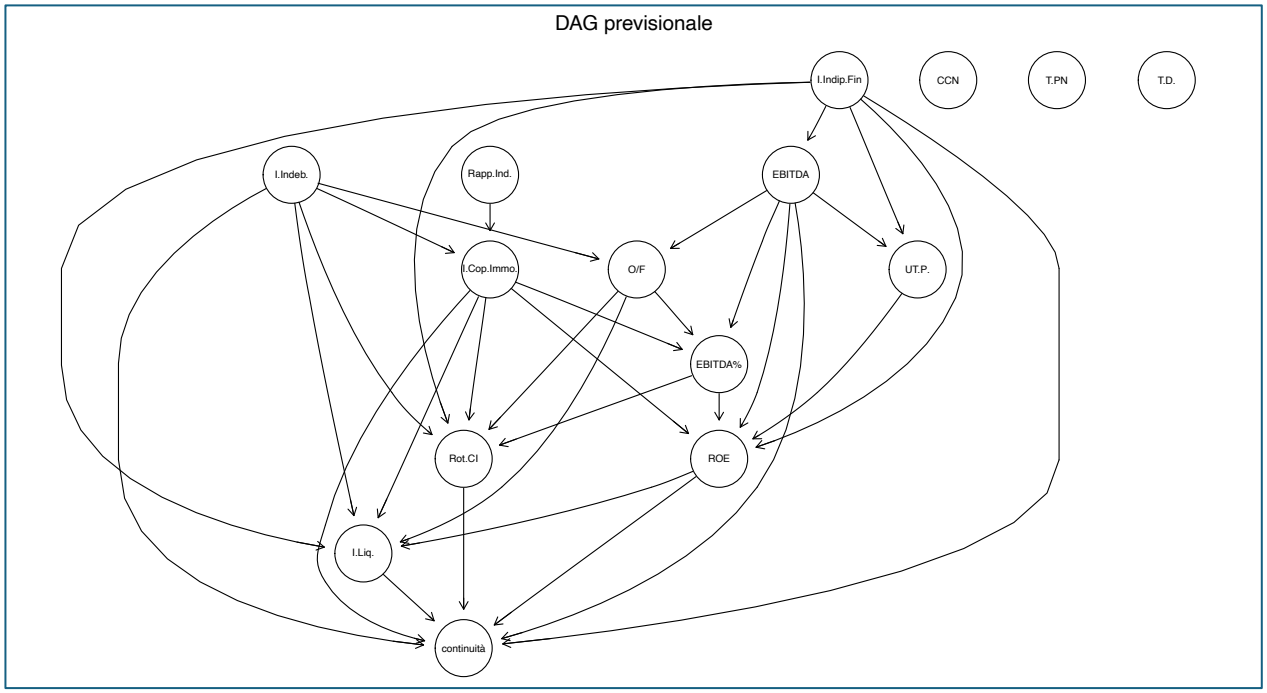


Figura 4.12: Rete Bayesiana previsionale. Fonte: elaborazione personale su dati AIDA

La rete così definita rappresenta pertanto la configurazione definitiva adottata per il modello previsionale. Le reti analizzate in precedenza, costruite separatamente sulle sole imprese sane o sulle sole imprese problematiche, sono state utilizzate esclusivamente a fini descrittivi, con lo scopo di evidenziare rispettivamente le dinamiche fisiologiche e patologiche dei due gruppi. La struttura finale, invece, è stimata sull'intero insieme delle imprese, includendo sia realtà in equilibrio sia realtà che hanno sperimentato condizioni di crisi.

Questa scelta consente di ottenere un modello pienamente funzionale all'obiettivo della ricerca: prevedere, sulla base degli indicatori di bilancio, la probabilità che un'impresa manifesti uno stato di squilibrio oppure mantenga condizioni di continuità operativa. Per tali ragioni, la rete con il nodo continuità posto come figlio costituisce la struttura più appropriata e viene assunta come base per la successiva fase di stima e valutazione del modello.

4.2 Costruzione del modello

In questo paragrafo viene illustrata la fase di costruzione del modello che verrà successivamente utilizzato per l'inferenza statistica.

L'obiettivo è definire una struttura capace di apprendere, durante la fase di addestramento, i pattern e le relazioni presenti nel database di riferimento, così da riconoscerli anche quando il modello verrà applicato a nuove osservazioni.

Una volta completato l'apprendimento, il modello sarà in grado di valutare i dati di un'impresa non ancora osservata e stabilire se le sue caratteristiche corrispondano a quelle tipiche di un'impresa sana oppure di un'impresa che, secondo quanto emerso dall'addestramento, presenta maggiori probabilità di sviluppare problematiche di continuità aziendale. Il modello sarà quindi in grado di generare una probabilità che permetterà di capire la salute dell'impresa analizzata.

4.2.1 Matrice di confusione e affidabilità del modello

Una volta completata la costruzione di tutti gli elementi necessari alla fase inferenziale, è fondamentale valutare la solidità e l'affidabilità del modello Bayesiano generato. A tal fine, è stata costruita una *matrice di confusione* tramite la funzione `confusionMatrix()` di R.

Per costruire la matrice di confusione è stato necessario considerare due insiemi differenti di dati: un *dataset* **training** per l'addestramento del modello e un *dataset* **test** per verificarne l'effettiva affidabilità.

Come *dataset* training è stato utilizzato l'insieme completo delle variabili, inclusa quella che definisce la continuità aziendale, in modo che il modello potesse apprendere quali trend associare a un'impresa sana e quali, invece, a situazioni di difficoltà economico-finanziaria.

Una volta addestrato il modello su questo database, si è proceduto alla fase di test dell'affidabilità utilizzando un *dataset* test privo della variabile "continuità aziendale". Questo per verificare la capacità del modello di associare autonomamente le tendenze dei dati di bilancio a un'impresa sana o problematica. Confrontando le risposte ottenute sul *dataset* test con i valori reali, è possibile osservare quando il modello risulta affidabile.

La matrice mette quindi a confronto i valori reali della variabile "continuità aziendale", presenti nel database utilizzato per l'addestramento, con i valori predetti dal modello. Per ottenere tali previsioni, la variabile "continuità" è stata preventivamente rimossa dal *dataset*: in questo modo il modello è stato costretto a decidere autonomamente l'appartenenza di ciascuna osservazione al gruppo delle imprese "sane" o "problematiche".

Attraverso il confronto tra classi reali e classi predette è quindi possibile quantificare il numero di classificazioni corrette e il numero di errori e valutare quanto il modello sia affidabile nella classificazione dei dati.

In particolare, l'analisi della matrice di confusione ha evidenziato quanto segue:

- *Falsi negativi: 18.710*
  Si tratta di imprese che nella realtà presentano problematiche di continuità aziendale, ma che il modello ha erroneamente classificato come sane. Questo tipo di errore è particolarmente critico, poiché implica che l'impresa in difficoltà non venga identificata come tale dal modello.
- *Falsi positivi: 8.907*
  In questo caso il modello ha generato un "falso allarme", classificando come problematica un'impresa che in realtà è sana. Sebbene rappresentino un errore, i falsi positivi hanno un impatto meno grave rispetto ai falsi negativi dal punto di vista applicativo.

Il fatto che i falsi negativi risultino più numerosi dei falsi positivi non rappresenta un segnale del tutto favorevole. Un'elevata incidenza di falsi negativi indica infatti una tendenza del modello a sottostimare le situazioni di crisi, per comprendere se questo indicatore può compromettere la criticità predittiva del modello rispetto al principale obiettivo dello studio si prosegue l'analisi di ulteriori indicatori di affidabilità.

```
> confusionMatrix(
     factor(prev.db.continuo, levels = c("prob","sana")),
     factor(etich.nuove.continuo, levels = c("prob","s…)))
     Confusion Matrix and Statistics
```

		Reference	
		Prob	Sana
Prediction	Prob	27.669	8.907
	Sana	18.710	40.768

```
Accuracy                 :    0,7125
95% CI                   :    (07096; 07153)
No Information Rate      :    0,5172
P-Value [Acc > NIR]      :    < 0,00000000000000022

Kappa                    :    0,4202

Mcnemar's Test P-Value   :    < 0,00000000000000022

Sensitivity              :    0,5966
Specificity              :    0,8207
Pos Pred Value           :    0,7565
Neg Pred Value           :    0,6854
Prevalence               :    0,4828
Detection Rate           :    0,2881
Detection Prevalence     :    0,3808
Balanced Accuracy        :    0,7086

'Positive' Class         :    prob
```

- *Accurancy (0,7125):*

  Identifica il fatto che il modello classifica correttamente il 71,25% dei casi;

- *95% Confidence Interval (0,7096; 0,7153):*

  L'intervallo di confidenza al 95% fornisce un'indicazione della stabilità dell'accuratezza: se il modello venisse applicato più volte a dati simili, l'accuratezza ricadrebbe con elevata probabilità all'interno di questo intervallo ristretto. La ridotta ampiezza delle due estremità conferma la robustezza della stima;

- *No information rate (0,5172)*:

  Questo indicatore definisce la percentuale di osservazioni che verrebbero classificate come corrette se si classificassero tutte le osservazioni come la classe più probabile (sane);

- *P-Value [Acc > NIR]* (<2,2 e^{-16}):

  Il p-value associato al confronto tra accuracy e NIR valuta se tale differenza sia statisticamente significativa. Un valore estremamente vicino a zero indica che l'accuratezza del modello è significativamente superiore a quella ottenibile tramite NIR;

- *Kappa (0,4202):*

  L'indice di Cohen's Kappa misura se l'accuratezza è casuale o se c'è potere predittivo, secondo la scala interpretativa di J.R. Landis e G.G. Koch[33], un valore superiore a 0,40 si colloca nel livello moderato, confermando che il modello possiede una capacità predittiva limitata;

- *Mcnemar's Test P-Value* (<2,2 e^{-16}):

  Il test di McNemar verifica la presenza di asimmetria negli errori di classificazione fra le due classi. Il p-value molto basso indica che il modello sbaglia più spesso in una direzione rispetto all'altra. In effetti, come visto nella matrice di confusione, i falsi negativi superano i falsi positivi, suggerendo una tendenza a non riconoscere pienamente le imprese problematiche;

- *Sensitivity* (0,5966): su 100 imprese realmente problematiche il modello ne classifica correttamente circa 60;

- *Specificity* (0,8207): su 100 imprese realmente sane il modello ne classifica correttamente circa 82;

- *Positive Prediction Value* (0,7565): Tra tutte le imprese predette come problematiche il 75,65% lo sono davvero;

---

[33] La classificazione dell'indice *Kappa* segue la scala proposta da J.R. Landis e G.G. Koch (1977), secondo cui i valori si interpretano come: lieve (0,00–0,20), debole (0,21–0,40), moderato (0,41–0,60), sostanziale (0,61–0,80) e quasi perfetto (0,81–1,00).

Landis, J.R., & Koch, G.G. (1977). *The Measurement of Observer Agreement for Categorical Data*. Biometrics, 33(1), 159–174.

- *Negative Prediction Value* (0,6854): Tra tutte le imprese predette come sane il 68,54% lo sono davvero;

- *Prevalence* (0,4828): rappresenta la proporzione della classe positiva;

- *Detection Rate* (0,2881): indica quanta parte dell'intero database viene classificata correttamente come problematica;

- *Detection Prevalence* (0,3808): quante imprese sono problematiche secondo il modello;

- *Balanced Accuracy* (0,7086): indica la media tra *sensitivity* e *specificity*; essendo l'indicatore elevato il modello si presenta come bilanciato.

Nel complesso, il modello evidenzia performance soddisfacenti e un buon livello di attendibilità nella classificazione delle imprese. L'accuratezza risulta elevata, allo stesso modo, l'indice di Kappa e la balanced accuracy confermano la presenza di una reale capacità predittiva, evidenziando che il modello non opera in modo casuale ma coglie effettivamente le differenze informative tra imprese sane e problematiche. I valori di sensitivity e specificity risultano complessivamente equilibrati, indicando una buona capacità sia di individuare i casi di continuità sia di rilevare situazioni potenzialmente critiche.

Permane tuttavia un margine di miglioramento nella riduzione dei falsi negativi, poiché il modello tende a non identificare completamente tutte le imprese problematiche. Tale aspetto rappresenta un possibile ambito di sviluppo futuro, soprattutto in applicazioni operative in cui la mancata rilevazione di un rischio potrebbe risultare particolarmente rilevante.

4.3 Inferenza

In questa sezione dell'elaborato vengono descritti gli utilizzi operativi del modello costruito, in particolare l'inferenza previsionale per la simulazione di scenari.

Dopo aver definito la struttura della rete e stimato i relativi parametri, l'inferenza permette di aggiornare le distribuzioni di probabilità dei nodi alla luce di nuove evidenze, ovvero di nuovi valori assegnati ai nodi della rete. Questo processo consente di valutare scenari alternativi, formulare previsioni e analizzare l'impatto informativo degli indicatori economico-finanziari sullo stato di continuità aziendale.

L'obiettivo principale del modello è rispondere alla domanda:

"Qual è la probabilità che un'impresa non presenti problematiche di continuità aziendale, dati questi specifici indici economico-finanziari?"

La rete restituisce quindi una stima probabilistica che quantifica il livello di vicinanza o lontananza dell'impresa rispetto a condizioni di squilibrio, fornendo così un indicatore sintetico di rischio.

Oltre all'analisi puntuale su singoli casi, il modello può essere utilizzato anche per una valutazione temporale della salute dell'impresa. Stimando annualmente la probabilità di continuità e osservandone l'evoluzione, è possibile evidenziare tendenze significative: una progressiva riduzione dell'indicatore potrebbe rappresentare un segnale di allarme, suggerendo un deterioramento strutturale della stabilità economico-finanziaria. Tale impostazione consente di affiancare alla fotografia statica una vera e propria analisi dinamica dei rischi.

Le attività inferenziali sono state sviluppate tramite due funzioni del software R, appartenenti al pacchetto `bnlearn`: *cpquery* e *predict*.

Il primo consente di eseguire inferenza probabilistica, stimando la probabilità condizionata di un evento dato un insieme di osservazioni. È quindi particolarmente utile nella simulazione di scenari ipotetici.

Il secondo algoritmo, invece, permette di effettuare classificazione, attribuendo a un'impresa la classe più probabile del nodo target (ad esempio "sana" o "problematica") sulla base degli

indicatori disponibili. Questo approccio viene utilizzato quando l'obiettivo è ottenere una previsione discreta.

È tuttavia necessaria una precisazione: nel database di addestramento la variabile *continuità* è una variabile continua. Di conseguenza, il risultato restituito dal modello non è una classe immediata, bensì un valore compreso nell'intervallo [-0,2523; 1,9352], con valore medio pari a 0,72.

Pertanto, più il valore predetto risulta elevato e superiore alla soglia media di 0,72, più l'impresa viene classificata come sana; al contrario, valori significativamente inferiori a tale soglia vengono interpretati come indicativi di una condizione problematica.

L'integrazione dei due strumenti consente quindi sia di produrre stime probabilistiche continue (tramite *cpquery*), sia di generare classificazioni discrete (tramite *predict*), offrendo così una rappresentazione completa e articolata del rischio di continuità aziendale.

Per tutte le procedure inferenziali sono state inoltre scaricate nuove osservazioni dalla banca dati AIDA, utilizzate come casi di test per la valutazione del modello.

4.3.1 Analisi di scenario

Per l'analisi per scenari si è scelto un campione di 47 imprese sane (impresa A, B, C) e un campione di 34 imprese che hanno riscontrato problematiche di continuità aziendale di diverso tipo (impresa D, E, F), tra questi nuovi campioni sono state scelte sei imprese caratteristiche degli scenari più utili ai fini dell'analisi.

Tutti gli scenari analizzati si riportano in *Appendice M*, qui di seguito si analizzano i più significativi.

Scenario 1 – impresa A

Il primo scenario che si vuole analizzare riguarda un'impresa con i seguenti valori di bilancio:

indice_di_liquidita_ultimo_anno_disp	4,49
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	0,60
rapporto_di_indebitamento_ultimo_anno_disp	1,26
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp	0,00
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	1,23
indice_di_indip_finanz_percent_percent_ultimo_anno_disp	79,28
rotaz_cap_investito_volte_ultimo_anno_disp	0,14
ebitda_vendite_percent_percent_ultimo_anno_disp	14,28
ebitda_migl_eur_ultimo_anno_disp	45,17
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp	72,71
capitale_circolante_netto_migl_eur_ultimo_anno_disp	517,19
totale_patrimonio_netto_migl_eur_ultimo_anno_disp	1.464,44
totale_debiti_migl_eur_ultimo_anno_disp	373,11
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	1.064,72

Interpretando singolarmente questi indicatori, emerge un quadro complessivamente molto positivo: la liquidità è ampiamente superiore ai livelli di sicurezza, l'indebitamento è moderato e con un impatto marginale sul fatturato, l'impresa è fortemente patrimonializzata e quindi scarsamente dipendente da finanziamenti terzi, la redditività è notevole e i flussi operativi risultano solidi.

Nel complesso, un'analisi preliminare porterebbe a considerare l'impresa in buono stato di salute economico-finanziaria.

Procedendo con l'inferenza probabilistica sulla Rete Bayesiana appresa e usando come evidenza insieme di questi valori, si ottiene una probabilità che l'impresa sia sana pari all'86,01%, un valore decisamente elevato che conferma la presenza di condizioni economiche favorevoli. La stima ottenuta tramite algoritmo *predict* restituisce un valore pari a 0,9809726, che ricade nella classe delle imprese sane.

Sulla base di tali risultati, il modello classifica quindi l'impresa come sana, evidenziando una probabilità residua del 13,92% di incorrere in problematiche di continuità aziendale.

Scenario 2 – impresa B

Ora si vogliono analizzare degli scenari anomali, il primo dei quali presenta la seguente struttura:

indice_di_liquidita_ultimo_anno_disp	0,90
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp"	0,30
rapporto_di_indebitamento_ultimo_anno_disp"	1,66
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp"	0,01
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp"	8,45
indice_di_indip_finanz_percent_percent_ultimo_anno_disp"	60,37
rotaz_cap_investito_volte_ultimo_anno_disp"	0,10
ebitda_vendite_percent_percent_ultimo_anno_disp"	52,94
ebitda_migl_eur_ultimo_anno_disp"	6.627,49
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp"	29,57
capitale_circolante_netto_migl_eur_ultimo_anno_disp"	-3.135,53
totale_patrimonio_netto_migl_eur_ultimo_anno_disp"	71.103,59
totale_debiti_migl_eur_ultimo_anno_disp"	45.405,37
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"	21.025,80

L'impresa mostra un indice di liquidità pari a 0,90, valore che indica un'incapacità della liquidità disponibile di coprire integralmente gli impegni a breve. L'indicatore di indebitamento a lungo termine risulta molto contenuto (0,30), mentre il rapporto di indebitamento complessivo è pari a 1,66. Tale configurazione lascia intuire che la componente principale dei debiti sia concentrata nel breve periodo, considerato anche il livello assoluto dei debiti, pari a 45 milioni di euro, in relazione a un patrimonio netto di 71 milioni di euro. La struttura finanziaria così configurata genera un impatto significativo sul conto economico, come evidenziato dagli oneri finanziari sul fatturato pari all'8,45%, un valore decisamente elevato.

Sul piano operativo, l'azienda presenta una rotazione del capitale investito molto bassa (0,10), mentre la marginalità risulta sorprendentemente elevata, con un EBITDA/Vendite del 52,94% e un EBITDA assoluto pari a 6 milioni di euro. Allo stesso modo, la redditività del capitale proprio, espressa da un ROE del 29,57%, risulta molto positiva, così come l'utile di esercizio, pari a 21 milioni di euro. Viceversa, il capitale circolante netto assume un valore negativo, pari a – 3 milioni di euro, confermando l'evidente squilibrio finanziario di breve periodo.

Proprio per la presenza congiunta di indicatori critici è stata scelta questa impresa come caso anomalo da confrontare con lo scenario precedente, che invece presentava una configurazione strutturalmente più equilibrata.

Nonostante tali squilibri, il modello classifica l'impresa come sana, con una probabilità di fallimento pari al 19,87%. Questo risultato può essere spiegato considerando che l'azienda presenta un livello di indipendenza finanziaria molto elevato, pari al 60,37%, accompagnato da una redditività del capitale significativa e da un utile coerente con tali performance. Come evidenziato nell'analisi condotta nel terzo capitolo, questi indicatori risultano essere particolarmente influenti nella determinazione della condizione di salute dell'impresa all'interno del modello, attribuendo loro un peso maggiore rispetto agli squilibri di natura strutturale. La spiegazione plausibile è quindi che, nonostante la situazione finanziaria presenti segnali di tensione, i risultati economici estremamente positivi permettono all'azienda di mantenere complessivamente una posizione ritenuta sostenibile sia dal modello, sia nella realtà.

Scenario 3 – impresa C

Si propone un altro scenario ambiguo dove la situazione è la seguente:

indice_di_liquidita_ultimo_anno_disp	1,39
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp"	0,19
rapporto_di_indebitamento_ultimo_anno_disp"	2,45
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp"	0,34
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp"	0,40
indice_di_indip_finanz_percent_percent_ultimo_anno_disp"	40,79
rotaz_cap_investito_volte_ultimo_anno_disp"	3,71
ebitda_vendite_percent_percent_ultimo_anno_disp"	9,17
ebitda_migl_eur_ultimo_anno_disp"	199,15
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp"	45,48
capitale_circolante_netto_migl_eur_ultimo_anno_disp"	206,71
totale_patrimonio_netto_migl_eur_ultimo_anno_disp"	237,91
totale_debiti_migl_eur_ultimo_anno_disp"	294,73
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"	108,21

Questo scenario è stato scelto come caso ambiguo, in quanto l'azienda analizzata presenta uno stato giuridico sano e non risulta coinvolta in procedure volte alla tutela della continuità aziendale; tuttavia, il modello la identifica come un'impresa problematica, attribuendole una probabilità del 40,28% di incorrere in squilibri economico finanziarie e una probabilità del 59,81% di essere sana. Sebbene il modello ritenga quindi più probabile una condizione di equilibrio rispetto a una situazione di crisi, l'algoritmo *predict* restituisce un valore pari a 0,61 che, pur essendo di poco inferiore alla soglia centrale, ricade nel range associato alle imprese problematiche.

Questo risultato può essere interpretato come un potenziale segnale d'allarme, soprattutto alla luce della natura non estremamente marcata, ma comunque presente, degli elementi di rischio.

Dal punto di vista strutturale, l'impresa mostra un indice di liquidità (1,39) adeguato a coprire gli impegni a breve termine, ma un profilo di indebitamento complessivamente elevato, con un ammontare di debiti (294 mila euro) che supera il patrimonio netto (237 mila euro). Nonostante ciò, l'incidenza degli oneri finanziari sul fatturato (0,40%) è molto contenuta, suggerendo condizioni favorevoli di accesso al credito o un costo del debito particolarmente basso. L'indipendenza finanziaria (40,79%) non risulta elevata, ma nemmeno tale da destare particolare preoccupazione. Anche la redditività del capitale proprio (45,48%) si colloca su livelli complessivamente buoni, sostenuta da indicatori economici che confermano un'attività operativa efficiente.

Nel complesso, il quadro aziendale appare abbastanza stabile, pur presentando alcune criticità legate alla struttura del debito. Proprio questi elementi sembrano essere individuati dal modello come fattori di rischio, contribuendo a ridurre in modo significativo la probabilità che l'impresa venga classificata come pienamente sana.

Tale combinazione di aspetti positivi e criticità strutturali giustifica quindi la natura ambigua dello scenario e la valutazione prudenziale prodotta dal modello.

Scenario 4 – impresa D

Si riporta ora lo scenario di un'impresa fallita che nell'ultimo anno disponibile presentava le seguenti caratteristiche:

indice_di_liquidita_ultimo_anno_disp	0,16
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	0,04
rapporto_di_indebitamento_ultimo_anno_disp	-6,21
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp	-4,8
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	0
indice_di_indip_finanz_percent_percent_ultimo_anno_disp	-16,1
rotaz_cap_investito_volte_ultimo_anno_disp	0,02
ebitda_vendite_percent_percent_ultimo_anno_disp	71,6
ebitda_migl_eur_ultimo_anno_disp	130,058
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp	-11,85
capitale_circolante_netto_migl_eur_ultimo_anno_disp	-5.847,795
totale_patrimonio_netto_migl_eur_ultimo_anno_disp	-1.047,662

totale_debiti_migl_eur_ultimo_anno_disp	7.276,692
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	124,175
stato giuridico	Fallita

Come si può immediatamente osservare, gli indicatori critici sono numerosi.

L'indice di liquidità risulta estremamente basso, al punto da non consentire la copertura delle obbligazioni a breve termine, evidenziando una tensione finanziaria ormai strutturale.

I valori negativi del rapporto di indebitamento (-6,21) e dell'indice di indipendenza finanziaria (-16,1%) indicano la presenza di un patrimonio netto ampiamente negativo (-1 milione di euro), circostanza confermata dall'apposito indicatore patrimoniale, e la conseguente assenza di mezzi propri sufficienti a sostenere l'attività. Tale condizione rappresenta uno degli elementi più chiari di insolvenza tecnica dell'impresa.

Anche il ROE (-11,85%) negativo contribuisce al quadro di criticità, segnalando una sistematica distruzione di valore da parte della gestione.

L'analisi dettagliata degli indici del bilancio, nel loro complesso, descrive dunque una situazione fortemente compromessa. Coerentemente, l'inferenza restituita dal modello Bayesiano assegna una probabilità che l'impresa sia sana pari al 41,94%, contro una probabilità di 57,98% che emergano problematiche strutturali. Utilizzando inoltre l'algoritmo *predict*, l'indicatore stimato risulta 0,41, valore che rientra nella categoria delle imprese problematiche, confermando ulteriormente la valutazione prodotta dal modello probabilistico.

Scenario 5 – impresa E

Ora si analizza uno scenario che potrebbe sembrare controintuitivo:

indice_di_liquidita_ultimo_anno_disp	1,30
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	0,00
rapporto_di_indebitamento_ultimo_anno_disp	2,27
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp	0,41
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	0,06
indice_di_indip_finanz_percent_percent_ultimo_anno_disp	44,08
rotaz_cap_investito_volte_ultimo_anno_disp	3,09
ebitda_vendite_percent_percent_ultimo_anno_disp	7,49
ebitda_migl_eur_ultimo_anno_disp	395,44
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp	27,56

capitale_circolante_netto_migl_eur_ultimo_anno_disp	465,29
totale_patrimonio_netto_migl_eur_ultimo_anno_disp	750,13
totale_debiti_migl_eur_ultimo_anno_disp	911,19
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	206,77
stato giuridico	Cessata

L'impresa presenta un indice di liquidità pari a 1,30, valore positivo e sufficiente a far fronte alle obbligazioni a breve termine. Anche gli indici reddituali risultano favorevoli: l'EBITDA è pari a 395 mila euro, la marginalità operativa raggiunge il 7,49% e il ROE si attesta al 27,56%, configurando una buona redditività complessiva che si traduce in un utile d'esercizio positivo.

La componente da monitorare riguarda invece l'area dell'indebitamento. Sebbene il rapporto di indebitamento pari a 2,27 risulti nel complesso sostenibile, il totale dei debiti (911 mila euro) supera il valore del patrimonio netto (750 mila euro), determinando un indice di indipendenza finanziaria (44,08%) inferiore alla media delle imprese sane. Tale configurazione non indica una crisi immediata, ma segnala un livello di leva finanziaria che richiede attenzione, poiché più elevato rispetto alle strutture patrimoniali tipicamente osservate nelle imprese considerate pienamente solide.

Nel complesso, la struttura economico-finanziaria non si presenta critica; tuttavia, la presenza di alcuni valori inferiori alla media e una situazione di indebitamento prossima ai limiti della sostenibilità contribuiscono a orientare la valutazione del modello Bayesiano verso una classificazione più prudenziale.

L'algoritmo *predict* restituisce infatti un valore pari a 0,57, che colloca l'impresa nella classe delle problematiche. Parallelamente, l'inferenza Bayesiana stima una probabilità di 56,46% che l'impresa sia sana e una probabilità del 43,21% che risulti problematica.

Ciò avviene perché, pur non essendo presenti segnali di squilibrio profondo, la rete individua alcune aree di vulnerabilità che la rendono statisticamente più simile ai profili borderline rispetto alle imprese completamente sane.

Scenario 6 – impresa F

L'ultimo scenario che si vuole riportare riguarda un'impresa con la seguente struttura:

indice_di_liquidita_ultimo_anno_disp	0,06
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	0,00
rapporto_di_indebitamento_ultimo_anno_disp	4,24
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp	0,02
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	1,27
indice_di_indip_finanz_percent_percent_ultimo_anno_disp	23,58
rotaz_cap_investito_volte_ultimo_anno_disp	0,78
ebitda_vendite_percent_percent_ultimo_anno_disp	31,78
ebitda_migl_eur_ultimo_anno_disp	334,36
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp	60,89
capitale_circolante_netto_migl_eur_ultimo_anno_disp	223,92
totale_patrimonio_netto_migl_eur_ultimo_anno_disp	309,99
totale_debiti_migl_eur_ultimo_anno_disp	988,03
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	188,75
stato giuridico	Fallita

Questo scenario presenta diversi indicatori di crisi particolarmente gravi, pienamente coerenti con lo stato giuridico finale di impresa fallita.

L'indice di liquidità risulta estremamente basso (0,06), evidenziando l'incapacità dell'organizzazione di far fronte alle obbligazioni di breve periodo. Anche il rapporto di indebitamento (4,42) è nettamente superiore alla media delle imprese sane, configurando una struttura finanziaria eccessivamente basata sul capitale di terzi. Tale quadro è ulteriormente aggravato dalla presenza di oneri finanziari (1,27%) significativi rispetto al fatturato e da un indice di indipendenza finanziaria (23,58%) molto ridotto.

Nonostante queste evidenze di squilibrio, l'impresa presenta una redditività sorprendentemente elevata, caratterizzata da margini positivi (EBITDA/vendite 31,78%) e da un utile d'esercizio consistente (188 mila euro). Questa particolare configurazione induce il modello Bayesiano ad attribuire alla struttura un valore *predict* (0,7237) superiore alla soglia media, collocandola quindi, seppur di poco, nella categoria delle imprese considerate sane. In coerenza con tale risultato, la rete calcola una probabilità del 69,30% che l'azienda sia sana.

Tale attribuzione deriva dal fatto che il modello, associa livelli elevati di redditività ai profili tipici delle imprese sane attribuendogli un peso maggiore rispetto alla criticità della situazione

di indebitamento. Tuttavia, in questo caso la redditività non è strutturale, ma verosimilmente derivante da componenti straordinarie che non compensano gli squilibri profondi evidenziati dagli altri indicatori. È probabilmente la presenza di tali anomalie ad aver comportato il successivo fallimento dell'impresa, nonostante l'apparente redditività straordinaria.

4.3.2 Scenario evolutivo

In questa sezione si propone un utilizzo alternativo del modello sviluppato, finalizzato al monitoraggio della probabilità che un'impresa sia sana nel corso degli anni e all'analisi dei relativi trend evolutivi. L'obiettivo è verificare la capacità del modello di cogliere tempestivamente eventuali segnali di deterioramento della continuità aziendale, così da consentire l'adozione di adeguate misure preventive quando emerge un andamento particolarmente negativo.

L'analisi è stata condotta su un ulteriore campione di imprese che, alla data attuale, risultano cessate o presentano problematiche di continuità aziendale. Per ciascuna di esse sono stati estratti i dati di bilancio relativi ai tre esercizi precedenti l'anno di cessazione.

Sulla base di tali informazioni sono state calcolate, anno per anno, le probabilità che ogni impresa fosse in condizioni di salute, così da identificare eventuali tendenze decrescenti.

Parallelamente, sono stati ottenuti anche i valori stimati tramite l'algoritmo *predict*, allo scopo di valutare come il modello classifichi la condizione delle imprese incluse nel campione.

Tutti i risultati completi sono riportati nell'*Appendice I*, comprendenti la probabilità di salute, l'indicatore di previsione del modello, la classe attribuita e la procedura di cessazione relativa a ciascuna impresa e ad ogni anno analizzato. In appendice verranno inoltre riportati gli indicatori di bilancio delle imprese oggetto della seguente analisi.

Il campione considerato è composto da 25 imprese; in questa sezione, tuttavia, vengono presentati solo cinque casi esemplificativi, scelti per rappresentare le principali dinamiche emerse dall'analisi complessiva.

Come si osserva a un primo sguardo nel grafico di figura 4.12, i trend evidenziati dalle imprese analizzate risultano molto eterogenei: alcune presentano un rapido deterioramento, altre

mostrano andamenti altalenanti, mentre per alcune il trend sembrerebbe addirittura positivo fino a ridosso della cessazione.

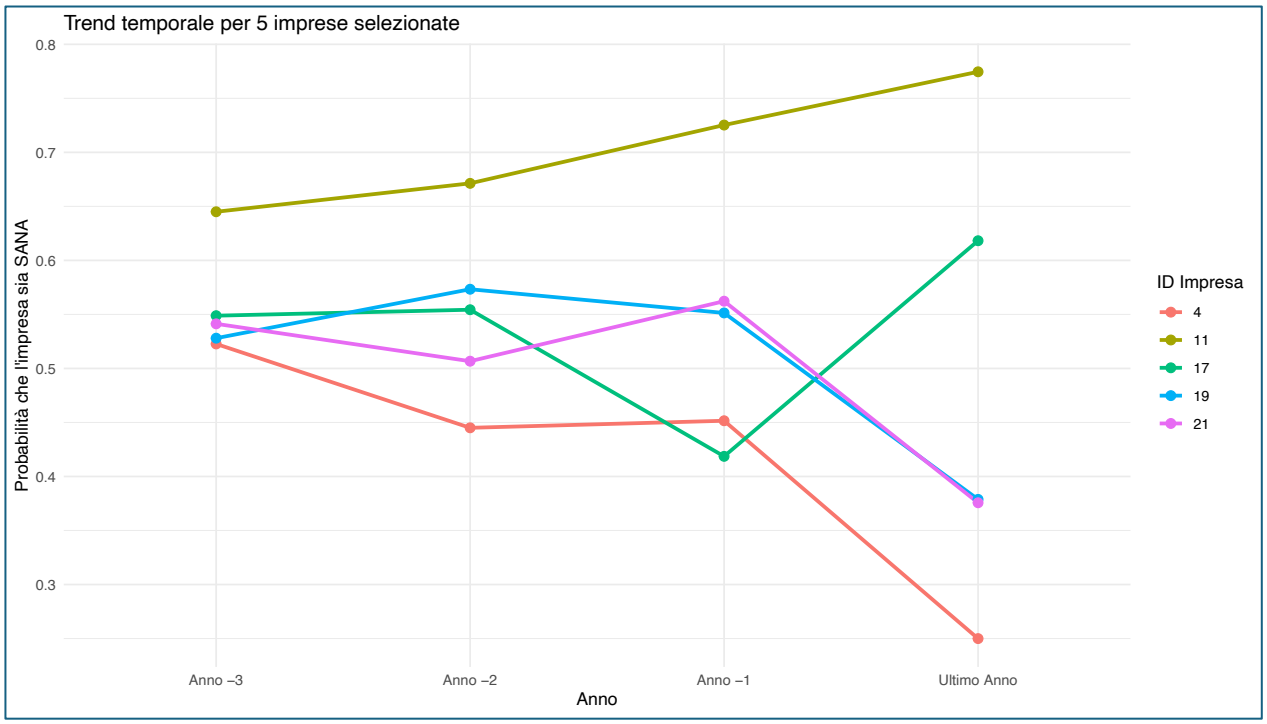


Figura 4.13: Scatterplot andamento delle probabilità di salute di un campione di imprese nei 3 anni antecedenti la cessazione. Fonte: elaborazione personale su dati AIDA

Tra le imprese rappresentate, l'impresa numero 4 costituisce il caso più tipico di cessazione determinata da problematiche economico-finanziarie. Già tre anni prima della cessazione essa mostrava una probabilità di salute relativamente bassa, venendo classificata dal modello come "problematica" con una probabilità di salute del 52,27%. Nel triennio successivo si osserva un progressivo peggioramento: la probabilità di continuità passa dal 44,51% (anno -2) al 45,16% (anno -1), fino a raggiungere un livello critico del 25% nell'ultimo esercizio, culminando nella cessazione in liquidazione.

L'analisi degli indici di bilancio conferma pienamente questo deterioramento. Per quanto riguarda la struttura finanziaria, il livello dei debiti rimane persistentemente elevato (124.885 → 72.848 → 107.665 → 59.253), soprattutto se confrontato con un patrimonio netto decisamente contenuto (16.837 → 17.046 → 15.149 → 12.617). Questo squilibrio determina un'indipendenza finanziaria estremamente bassa, con valori prossimi alle soglie critiche (8,90% → 13,52% → 10,64% → 14,27%).

Anche l'incidenza degli oneri finanziari sul fatturato mostra un peggioramento significativo, passando dall'1,57% (anno -3) al 5,18% (ultimo anno), segnalando un crescente peso del costo del debito sulla gestione operativa. Tale incremento rappresenta un chiaro indicatore di vulnerabilità finanziaria.

Sul versante reddituale, il trend dell'EBITDA/vendite evidenzia una perdita progressiva di marginalità (11,15% → 5,99% → 5,20% → -12,13%), che si riflette in una redditività del capitale via via più debole e infine fortemente negativa (7,74% → 1,23% → -12,62% → -149,45%). Il livello di redditività dell'ultimo anno segnala un vero e proprio processo di distruzione del valore.

L'utile d'esercizio conferma questo quadro: anch'esso presenta una dinamica fortemente decrescente, con una transizione da un piccolo utile a perdite ingenti nell'anno finale (1.304 → 0,21 → -1.912 → -18.856).

Nel complesso, l'analisi congiunta degli indici economico-finanziari e delle probabilità stimate dal modello mette in luce un chiaro e progressivo deterioramento della salute aziendale, che precede e spiega la cessazione dell'attività.

Per quanto riguarda le imprese 19 e 21, anch'esse vengono costantemente classificate dal modello come problematiche in tutte le annualità analizzate. Pur mostrando trend meno drammatici rispetto all'impresa 4, entrambe presentano andamenti altalenanti che si mantengono comunque su valori mediamente bassi, ben al di sotto delle probabilità osservate nelle imprese considerate sane. Un elemento comune e particolarmente significativo è il marcato peggioramento registrato nell'ultimo anno, che porta entrambe le imprese a livelli prossimi alle soglie critiche e precede direttamente la cessazione delle attività.

Dal confronto della struttura economico-finanziaria emergono differenze rilevanti: l'impresa 19 presenta un rapporto di indebitamento nettamente più elevato e critico rispetto all'impresa 21, evidenziando una maggiore esposizione finanziaria. Nonostante ciò, entrambe condividono una redditività del capitale proprio fortemente negativa, indicativa di una sistematica distruzione del valore. Tale condizione si manifesta chiaramente nei risultati d'esercizio, che mostrano perdite significative nell'ultimo anno per entrambe le imprese.

La combinazione di instabilità gestionale, redditività insufficiente e peggioramento finale definisce un quadro di fragilità strutturale che rende insostenibile la prosecuzione dell'attività aziendale, culminando nella cessazione in entrambi i casi.

Per quanto riguarda l'impresa 17, il grafico in figura 4.12 mostra un andamento differente rispetto ai casi precedentemente analizzati: nell'ultimo anno disponibile si osserva infatti un miglioramento delle probabilità di continuità, in contrasto con il rapido deterioramento evidenziato dalle altre imprese del campione.

Questo apparente miglioramento trova spiegazione nell'analisi degli indici reddituali. In particolare, l'impresa registra nell'ultimo esercizio un aumento straordinario della marginalità, con un indicatore EBITDA/Vendite che passa dal 6,97% nel quarto anno antecedente la cessazione al 58,40% nell'ultimo anno. In modo coerente, anche la redditività del capitale proprio migliora significativamente, passando dal 15,62% fino al 31,90%, dopo aver raggiunto valori fortemente negativi nel secondo anno antecedente (-37,36%). Questi trend trovano riscontro anche nell'andamento dell'utile d'esercizio, che segue un percorso altalenante ma si chiude con un risultato positivo nell'ultimo anno (1.851 → 5.134 → -4.633 → 7.269).

Tuttavia, nonostante la ripresa sul piano reddituale, l'impresa presenta gravi squilibri finanziari che ne compromettono la stabilità complessiva. L'indice di liquidità risulta infatti costantemente insufficiente (0,74 → 0,88 → 0,84 → 0,89), evidenziando l'incapacità di far fronte agli impegni di breve periodo. Allo stesso tempo, il rapporto di indebitamento rimane elevato e instabile (6,89 → 5,38 → 7,15 → 4,77), mentre l'indipendenza finanziaria si mantiene su livelli decisamente critici (14,51% → 18,58% → 13,89% → 20,98%).

Tra tutti gli indicatori analizzati, quello che meglio spiega le difficoltà strutturali dell'impresa è il capitale circolante netto, che risulta persistentemente inadeguato in tutti gli anni osservati (-16.374 → -7.483 → -10.918 → -9.094).

Ciò indica l'incapacità dell'impresa di sostenere l'equilibrio finanziario nel breve periodo, nonostante l'apparente miglioramento reddituale.

In conclusione, anche se gli indicatori economici evidenziano una ripresa nell'ultimo esercizio, la presenza di squilibri finanziari profondi e radicati rende l'impresa strutturalmente instabile.

È proprio tale fragilità finanziaria a spiegare perché, pur in presenza di segnali positivi sul fronte reddituale, l'organizzazione non sia stata in grado di proseguire l'attività.

Per completezza di analisi, in questa sezione viene riportato anche il caso dell'impresa 11, la cui traiettoria mostra un andamento sensibilmente diverso rispetto alle altre imprese analizzate. Come evidenziato dal grafico, essa presenta infatti un trend crescente delle probabilità di salute, con valori nettamente superiori a quelli delle imprese caratterizzate da situazioni problematiche.

Nel dettaglio, pur essendo inizialmente classificata dal modello come "problematica", con una probabilità di salute pari al 64,50% nel quarto anno antecedente la cessazione, l'impresa viene successivamente identificata come "sana" nell'anno anteceente all'ultimo e nell'ultimo anno disponibile, con probabilità di essere sana rispettivamente pari al 72,53% e al 77,46%.

L'analisi degli indici di bilancio conferma la presenza di una struttura economico-finanziaria complessivamente equilibrata. Il rapporto di indebitamento (3,65 → 2,96 → 3,58 → 1,02) appare coerente con i valori medi rilevati per le imprese sane, mentre l'indipendenza finanziaria risulta particolarmente elevata (27,42% → 33,82% → 27,95% → 97,64%). Questo andamento è ulteriormente validato dal confronto tra l'ammontare complessivo dei debiti (3.259.739 → 2.849.550 → 4.526.675 → 43.061) e il patrimonio netto (1.477.566 → 1.690.146 → 1.938.302 → 1.936.965), che mostra un significativo riequilibrio nell'ultimo anno.

Nonostante il quadro strutturale tutto sommato positivo, emergono segnali di peggioramento sul fronte reddituale e finanziario. In particolare, l'impresa evidenzia un calo della redditività nell'ultimo esercizio: il ROE subisce una brusca inversione (–1,52% → 12,58% → 12,80% → –0,07%), accompagnata da una diminuzione dei margini operativi (EBITDA/Vendite 6,30% → 6,83% → 8,59% → 4,20%). Parallelamente, il capitale circolante netto mostra una riduzione progressiva (970.738 → 914.659 → 533.898 → –27.442), giungendo a valori negativi nell'ultimo anno e indicando l'incapacità dell'impresa di far fronte agli impegni di breve periodo (dato confermato dall'indice di liquidità).

Questo andamento suggerisce che, nonostante una struttura economico-finanziaria non critica e una performance reddituale complessivamente adeguata nella maggior parte del periodo analizzato, il deterioramento registrato nell'ultimo anno possa aver inciso in modo determinante sulla scelta di liquidare l'impresa. Tale caso risulta particolarmente significativo poiché evidenzia che la cessazione di un'attività non è necessariamente riconducibile esclusivamente

a squilibri economico-finanziari: possono intervenire fattori esterni, dinamiche di mercato, decisioni strategiche o scelte personali dei soci.

Per questo motivo, l'impresa 11 rappresenta un esempio utile a sottolineare i limiti informativi dei soli dati di bilancio e del modello predittivo: pur identificata come sana dal modello, l'impresa è comunque sfociata nella liquidazione, mostrando come alcuni eventi non siano interamente spiegabili attraverso le variabili considerate e come sia essenziale integrare valutazioni del contesto di analisi.

In conclusione, sulla base delle dimostrazioni condotte attraverso scenari applicativi concreti, è possibile affermare che il modello sviluppato risulta in grado di stimare in modo adeguato la probabilità di mantenimento della continuità aziendale di un'impresa, integrando l'analisi degli indicatori economico-finanziari e attribuendo a ciascuna relazione il peso informativo appropriato.

A differenza degli approcci tradizionali fondati su soglie statiche o su analisi univariate, la tipologia di modello proposta consente di valutare la salute di un'impresa come il risultato congiunto dell'interazione tra più variabili e delle relazioni che le legano, utilizzando la logica Bayesiana per aggiornare in modo coerente la probabilità di continuità al variare delle condizioni economico-finanziarie osservate.

Dal punto di vista applicativo, il modello può rappresentare un valido strumento di supporto per l'attività dei professionisti, permettendo di simulare scenari alternativi e di analizzare come varia la probabilità di continuità aziendale al mutare degli indicatori chiave, contribuendo così a rendere più efficiente la fase di valutazione preliminare.

Resta ferma la consapevolezza dei limiti del presente studio, legati al fatto che il modello, nello stato attuale, non può sostituire un'analisi puntuale e qualitativa del contesto in cui l'impresa opera. Come emerso nel corso dello studio, risulta infatti fondamentale considerare anche gli aspetti organizzativi, strategici e decisionali che caratterizzano la realtà aziendale, nonché le scelte operate dai soggetti apicali, influenzate da preferenze e valutazioni di natura personale.

Il modello proposto si configura pertanto come uno strumento di supporto all'analisi dell'andamento economico-finanziario delle imprese, ma non come un sostituto del processo

decisionale, che rimane necessariamente legato alla conoscenza del contesto specifico e alle valutazioni discrezionali degli attori coinvolti.

In conclusione, il lavoro dimostra come l'approccio basato sulle Reti Bayesiane rappresenti uno strumento efficace per l'analisi della salute d'impresa, offrendo una prospettiva probabilistica coerente con la complessità e l'incertezza che caratterizzano i fenomeni economico-aziendali.

Conclusione

Il presente lavoro ha analizzato il tema della continuità aziendale adottando un approccio statistico di tipo predittivo, con l'obiettivo di valutare l'applicabilità delle reti Bayesiane al monitoraggio dello stato di salute delle imprese. In un contesto economico caratterizzato da elevata complessità e incertezza, l'utilizzo di modelli probabilistici si configura come un valido supporto all'analisi dei fenomeni economico-aziendali.

Il modello sviluppato ha permesso di rappresentare la continuità aziendale come il risultato di un sistema di relazioni interdipendenti tra variabili economico-finanziarie osservate su più periodi temporali. La struttura a DAG multilivello ha consentito sia l'analisi statica di un singolo esercizio contabile, utile per comprendere la situazione economica dell'impresa in un determinato momento, sia l'analisi congiunta di più esercizi, integrando le informazioni disponibili in una prospettiva temporale più ampia. Tale impostazione ha reso possibile l'aggiornamento dinamico delle probabilità, evidenziando la capacità del modello di adattarsi al variare delle condizioni osservate e di offrire una lettura evolutiva del rischio di perdita della continuità aziendale.

Dal punto di vista metodologico, le reti Bayesiane si sono dimostrate uno strumento efficace per l'analisi di fenomeni complessi, in quanto consentono di gestire interdipendenze tra le variabili e di operare anche in presenza di informazione incompleta. Il contributo principale del lavoro risiede nell'applicazione di un approccio tipicamente statistico a un ambito di analisi economico-aziendale, evidenziando come le tecniche di Machine Learning probabilistico possano affiancare gli strumenti tradizionali di monitoraggio, fornendo indicazioni utili in ottica prospettica e decisionale.

L'analisi svolta evidenzia come la valutazione della continuità aziendale, pur potendo essere efficacemente supportata da modelli statistici predittivi, rimanga un fenomeno intrinsecamente complesso, influenzato da elementi che non sempre risultano osservabili attraverso i dati economico-finanziari. Il modello proposto si fonda infatti su informazioni quantitative disponibili, che consentono di stimare il rischio di perdita della continuità aziendale in funzione delle condizioni economico-finanziarie dell'impresa, ma non permettono di rappresentare integralmente la dimensione decisionale sottostante.

In particolare, le scelte degli imprenditori e degli organi di governance possono incidere in modo significativo sull'esito finale, indipendentemente dalla presenza di squilibri economico-finanziari. Decisioni strategiche, cambiamenti nelle preferenze individuali, processi di riorganizzazione o la volontà di cessare l'attività per motivazioni non strettamente economiche rappresentano fattori difficilmente quantificabili e non direttamente incorporabili nel modello.

Inoltre, la continuità aziendale può essere compromessa da condizioni di crisi economica esterna, quali shock macroeconomici, instabilità dei mercati di riferimento o eventi esogeni che incidono improvvisamente sulle prospettive operative dell'impresa. Tali fenomeni, spesso non anticipabili attraverso i soli dati storici di bilancio, contribuiscono ad aumentare il livello di incertezza e possono influenzare l'esito indipendentemente dalla situazione economico-finanziaria pregressa.

Le probabilità stimate devono pertanto essere interpretate come misure condizionate alle informazioni disponibili e non come previsioni deterministiche dell'evoluzione futura dell'impresa. In questo senso, il limite principale del modello non risiede nella struttura statistica adottata, ma nella difficoltà di rappresentare quantitativamente la complessità dei processi decisionali e del contesto economico in cui le imprese operano.

In conclusione, il presente lavoro evidenzia come l'approccio Bayesiano rappresenti uno strumento statisticamente solido e flessibile per l'analisi della continuità aziendale, confermando il ruolo centrale della statistica nella trasformazione dei dati economico-finanziari in informazioni utili al supporto dei processi decisionali.

Appendici

Appendice A – Indicatori di rischio secondo ISA 570

Di seguito si riportano gli elenchi completi degli eventi indicatori di rischio, presente nel principio ISA 570.

Indicatori finanziari

- situazione di deficit patrimoniale o di capitale circolante netto negativo;
- prestiti a scadenza fissa e prossimi alla scadenza senza che vi siano prospettive verosimili di rinnovo o di rimborso oppure eccessiva dipendenza da prestiti a breve termine per finanziare attività a lungo termine;
- indizi di cessazione del sostegno finanziario da parte dei creditori;
- bilanci storici o prospettici che mostrano flussi di cassa negativi;
- principali indici economico-finanziari negativi;
- consistenti perdite operative o significative perdite di valore delle attività utilizzate per generare i flussi di cassa;
- difficoltà nel pagamento di dividendi arretrati o discontinuità nella distribuzione di dividendi;
- incapacità di pagare i debiti alla scadenza;
- incapacità di rispettare le clausole contrattuali dei prestiti;
- cambiamento delle forme di pagamento concesse dai fornitori, dalla condizione “a credito” alla condizione “pagamento alla consegna”;
- incapacità di ottenere finanziamenti per lo sviluppo di nuovi prodotti ovvero per altri investimenti necessari.

Indicatori gestionali

- intenzione della direzione di liquidare l'impresa o di cessare le attività; perdita di membri della direzione con responsabilità strategiche senza una loro sostituzione;

- perdita di mercati fondamentali, di clienti chiave, di contratti di distribuzione, di concessioni o di fornitori importanti;

- difficoltà con il personale;

- scarsità nell'approvvigionamento di forniture importanti;

- comparsa di concorrenti di grande successo.

Altri indicatori

- capitale ridotto al di sotto dei limiti legali o non conformità del capitale ad altre norme di legge, come i requisiti di solvibilità o liquidità per gli istituti finanziari;

- procedimenti legali o regolamentari in corso che, in caso di soccombenza, possono comportare richieste di risarcimento cui l'impresa probabilmente non è in grado di far fronte;

- modifiche di leggi o regolamenti o delle politiche governative che si presume possano influenzare negativamente l'impresa;

- eventi catastrofici contro i quali non è stata stipulata una polizza assicurativa ovvero contro i quali è stata stipulata una polizza assicurativa con massimali insufficienti.

Fonte: International Auditing and Assurance Standards Board (IAASB), ISA 570 "Going Concern", 2014.

Appendice B – Indici di bilancio da monitorare secondo il CNDCEC

Di seguito riportati gli indici di bilancio degni di analisi evidenziati dal Consiglio Nazionale dei Dottori Commercialisti ed Esperti Contabili per l'individuazione preventiva di uno stato di crisi aziendale.

INDICI DI BILANCIO

Equazione patrimoniale

Crediti + Immobilizzazioni	=	Debiti vs banche+ Debiti vs fornitori + Altri debiti + Capitale

Indici di durata

Indice di durata delle scorte	$\frac{\text{GIACENZA MEDIA SCORTE}}{\text{CONSUMI}} x\,365$	Indica il tempo medio di realizzo delle scorte, ovvero il n. di giorni che intercorrono tra l'entrata e l'uscita dal magazzino degli stessi beni.
Indice di durata di crediti commerciali	$\frac{\text{CREDITI V/ CLIENTI}}{\text{RICAVI DI VENDITA}} x\,365$	Indica il tempo medio di incasso dei crediti, ovvero il tempo medio che intercorre tra la vendita e l'incasso.
Indice di durata dei debiti commerciali	$\frac{\text{DEBITI V/ FORNITORI}}{\text{ACQ BENI E SERVIZI}} x\,365$	Indica il tempo medio di pagamento dei debiti commerciali, ovvero il periodo che mediamente intercorre tra l'acquisto e l'esborso finanziario.

Indici di solidità patrimoniale

Indici di solidità del capitale sociale	$\frac{\text{CAPITALE PROPRIO}}{\text{CAPITALE SOCIALE}}$	Indica la presenza di componenti aggiuntive rispetto all'apporto nominale dei soci, ovvero segnala la capacità del patrimonio di accrescersi indipendentemente dagli incrementi del Capitale Sociale. La solidità del capitale sociale viene misurata con quozienti variabili da uno in poi.
Indice di auto copertura delle immobilizzazioni	$\frac{\text{CAPITALE PROPRIO}}{\text{IMMOBILIZZAZIONI}}$	Indica il grado di capitalizzazione dell'impresa. Indica quanta parte dei mezzi propri finanzia gli investimenti durevoli. Valore corretto ≈ 1
Indice di copertura globale delle immobilizzazioni	$\frac{\text{CAPITALE PROPRIO} + \text{PASSIVITA A ML TERMINE}}{\text{IMMOBILIZZAZIONI}}$	Valore corretto >1
Indice di indipendenza finanziaria	$\frac{\text{TOT PATRIMONIO NETTO}}{\text{TOT PASSIVO}}$	Indica l'incidenza del capitale proprio, ovvero in che misura il totale dei mezzi investiti nell'impresa sia stato finanziato con capitale di rischio. 33% <Valore corretto< 66% Un basso indice segnala un elevato indebitamento.

Indici di indebitamento

Indici di indebitamento a breve	$\frac{\text{PASSIVO CORRENTE}}{\text{PATRIMONIO NETTO}}$	Indica l'incidenza dei debiti a breve.
Indice di indebitamento M/L termine	$\frac{\text{PASSIVITÀ CONSOLIDATE}}{\text{PATRIMONIO NETTO}}$	Indica l'incidenza dei debiti a m/l termine.
Ricorso al capitale di terzi	$\frac{\text{PASSIVO CORRENTE} + \text{PASSIVITA ML TERMINE}}{\text{PATRIMONIO NETTO}}$	Valore corretto >1

Equazione monetaria

Entrate	=	Uscite

Indici di liquidità finanziaria

Liquidirà primaria	$\frac{\text{LIQUIDITÀ IMMEDIATA}}{\text{PASSIVITÀ CORRENTI}}$	Indica la capacità di coprire le uscite con le entrate correnti. Valore corretto ≈ 0
Liquidità secondaria	$\frac{\text{LIQUIDITA IMMEDIATE} + \text{LIQUIDITA DIFFERITE}}{\text{PASSIVITA CORRENTI}}$	Indica la capacità dell'azienda di far fronte agli impegni nel breve termine (entro un anno). Valore corretto ≈ 1 per il quale la cassa e i crediti riescono a far fronte ai debiti a breve.
Indice di disponibilità	$\frac{\text{ATTIVITÀ CORRENTI}}{\text{PASSIVITÀ CORRENTI}}$	Indica la capacità di far fronte agli impieghi con tutte le fonti disponibili comprese le rimanenze, che per natura sono le attività meno liquide. 1,5 < Valore corretto < 2

Margini finanziari

Margine di tesoreria	(LIQ DIF + LIQ IMM) - PASS CORRENTI	Tale valore dovrebbe essere > 0, se negativo segnala la presenza di tensioni finanziarie di breve periodo.
Margine di struttura primaria	PN - ATT. IMMOBILIZZATE	Tale differenza indica la capacità di copertura delle immobilizzazioni con mezzi propri.
Capitale circolante netto	A. CORRENT - P. CORRENTI Ovvero (PN + pass cons) - tot At. Imm	Il CCN segnala da un lato la capacità dell'impresa di far fronte ai propri impegni finanziari e dall'altro l'esistente situazione di equilibrio o squilibrio tra investimenti in immobilizzazioni e capitale permanente.

Equazione economica

Ricavi	=	Costi+ Uscite

Indici di redditività

ROS	$\frac{\text{REDDITO OPERATIVO (EBIT)}}{\text{RICAVI NETTI DI VENDITA}}$	Redditività delle vendite. Indica il margine residuale dei flussi dei ricavi dopo l'assorbimento dei costi operativi e prima dei componenti legali alla gestione finanziaria e straordinaria.
ROI	$\frac{\text{REDDITO OPERATIVO (EBIT)}}{\text{ATTIVITÀ NETTE (capitale investito)}}$	Redditività del capitale investito. In dica la capacità dell'impresa di sfruttare gli investimenti in essere producendo un flusso di reddito proporzionato alle risorse impiegate.
ROE	$\frac{\text{UTILE D'ESERCIZIO}}{\text{PATRIMONIO NETTO}}$	Redditività del capitale proprio. Indica la redditività che l'Azienda è in grado di generare con la propria attività rispetto al capitale proprio.
EBITDA margin	$\frac{\text{EBITDA}}{\text{FATTURATO}}$	Indica quanta parte del fatturato diventa margine operativo lordo.

Leverage (Indice di indebitamento)	$\frac{\text{TOT IMPIEGHI (tot passivo)}}{\text{PATRIMONIO NETTO}}$	Indice di indebitamento. Il valore corretto a nostro giudizio è compreso tra 1,5 e 3. Se il risultato fosse = 1 avrei perfetta indipendenza.

Indici finanziari

Incidenza degli oneri finanziari	$\frac{\text{ONERI FINANZIARI}}{\text{RICAVI DI VENDITA}}$	Un elevato livello di indebitamento può generare elevati oneri finanziari. Valore corretto < 5.
Costo medio dei mezzi di terzi	$\frac{\text{ONERI FINANZIARI}}{\text{MEZZI DI TERZI}}$	Indica il costo medio sostenuto per l'impiego di mezzi di terzi.
Incidenza oneri finanziari su EBIT	$\frac{\text{ONERI FINANZIARI}}{\text{EBIT}}$	Indica in che misura gli oneri finanziari hanno consumato il reddito della gestione caratteristica.

INDICATORI EXTRA-CONTABILI

Andamento del portafoglio di ordini

$$N° \; ORIDNI \; ANNO \; (N) - N° \; ORDINI \; ANNO \; (N-1)$$

Indice di rotazione e grado di affidabilità del personale

$$\frac{CESSAZIONI \; VOLONTARIE \; DEI \; RAPPORTI \; DI \; LAVORO}{N° \; DIPENDENTI}$$

$$\frac{CESSAZIONI \; DEI \; RAPPORTI \; DI \; LAVORO}{N° \; DIPENDENTI}$$

Grado di know-how in possesso dei dipendenti

$$N° \; CORSI \; DI \; FORMAZIONE \; SEGUITI \; IN \; UN \; ANNO$$

Oppure

$$\frac{N°\ ORE\ DEDICARE\ ALLA\ FROMAZIONE}{N°\ ORE\ LAVORATE\ TOTALI}$$

Customer satisfaction

$$\frac{N°\ RECLAMI}{TOT\ VENDITE}$$

Ricorrenza di guasti e riparazioni

$$\frac{N°\ GUASTI\ MACCHINARI\ IN\ UN\ ANNO}{N°\ RIPARAZIONI\ EFFETTTUATE\ IN\ UN\ ANNO}$$

Indice del costo del lavoro

$$\frac{COSTO\ DEL\ LAVORO}{RICAVI\ DI\ VENDITA}$$

Indici di produttività

Ricavi pro-capite	$\frac{RICAVI\ DI\ VENDITA}{N°DIPENDENTI}$
Valore della produzione pro-capite	$\frac{VALORE\ DELLA\ PRODUZIONE}{N°DIPENDENTI}$
Valore aggiunto pro-capite Costo medio del lavoro per dipendente	$\frac{VALORE\ AGGIUNTO}{N°DIPENDENTI}$ $\frac{COSTO\ DEL\ PERSONALE}{N°DIPENDENTI}$
Prezzo medio di vendita	$\frac{RICAVI\ DI\ VENDITA}{VENDITA\ IN\ UNITA\ DI\ MISURA}$

Fonte: Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC), La crisi d'impresa – Un approccio multidisciplinare alla diagnosi, gennaio 2005.

Appendice C – *Elenco variabili del database utilizzato nello studio*

Di seguito riportato l'elenco di tutte le variabili selezionate per i due database costruiti; nella colonna di sinistra vengono riportate tutte le variabili inizialmente selezionate; nella colonna di destra si trovano le variabili che sono state scelte e mantenuto nel database finale.

VARIABILI PRE PULIZIA - DATABASE INIZIALE	VARIABILI POST PULIZIA - DATABASE FINALE
...1	
Ragione sociale	ragione_sociale
Provincia	provincia
Chiusura bilancio\nUltimo anno disp.	chiusura_bilancio_ultimo_anno_disp
Ricavi delle vendite\nmigl EUR\nUltimo anno disp.	ricavi_delle_vendite_migl_eur_ultimo_anno_disp
Dipendenti\nUltimo anno disp.	dipendenti_ultimo_anno_disp
Indirizzo sede legale - Regione	indirizzo_sede_legale_regione
	area_geografica
Stato giuridico	stato_giuridico
Anno di costituzione	anno_di_costituzione
Procedura/Cessazione	Procedura/Cessazione
Data di inizio procedura/cessazione	Data di inizio procedura/cessazione
Data di chiusura della procedura	
ATECO 2007\ncodice	ateco_2007_codice
Oggetto sociale	oggetto_sociale
Indice di liquidità\nUltimo anno disp.	indice_di_liquidita_ultimo_anno_disp
Indice di liquidità\nAnno - 1	indice_di_liquidita_anno_1
Indice di liquidità\nAnno - 2	indice_di_liquidita_anno_2
Indice di liquidità\nAnno - 3	indice_di_liquidita_anno_3
Indice di liquidità\nAnno - 4	
Indice di liquidità\nAnno - 5	
Indice di indebitam. a breve\n%\nUltimo anno disp.	indice_di_indebitam_a_breve_percent_ultimo_anno_disp
Indice di indebitam. a breve\n%\nAnno - 1	indice_di_indebitam_a_breve_percent_anno_1
Indice di indebitam. a breve\n%\nAnno - 2	indice_di_indebitam_a_breve_percent_anno_2
Indice di indebitam. a breve\n%\nAnno - 3	indice_di_indebitam_a_breve_percent_anno_3
Indice di indebitam. a breve\n%\nAnno - 4	
Indice di indebitam. a breve\n%\nAnno - 5	
Indice di indebitam. a lungo\n%\nUltimo anno disp.	indice_di_indebitam_a_lungo_percent_ultimo_anno_disp
Indice di indebitam. a lungo\n%\nAnno - 1	indice_di_indebitam_a_lungo_percent_anno_1
Indice di indebitam. a lungo\n%\nAnno - 2	indice_di_indebitam_a_lungo_percent_anno_2

Indice di indebitam. a lungo\n%\nAnno - 3	indice_di_indebitam_a_lungo_percent_anno_3
Indice di indebitam. a lungo\n%\nAnno - 4	
Indice di indebitam. a lungo\n%\nAnno - 5	
Indice di copertura delle immob. (patrimoniale)\n%\nUltimo anno disp.	indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp
Indice di copertura delle immob. (patrimoniale)\n%\nAnno - 1	indice_di_copertura_delle_immob_patrimoniale_percent_anno_1
Indice di copertura delle immob. (patrimoniale)\n%\nAnno - 2	indice_di_copertura_delle_immob_patrimoniale_percent_anno_2
Indice di copertura delle immob. (patrimoniale)\n%\nAnno - 3	indice_di_copertura_delle_immob_patrimoniale_percent_anno_3
Indice di copertura delle immob. (patrimoniale)\n%\nAnno - 4	
Indice di copertura delle immob. (patrimoniale)\n%\nAnno - 5	
Rapporto di indebitamento\nUltimo anno disp.	rapporto_di_indebitamento_ultimo_anno_disp
Rapporto di indebitamento\nAnno - 1	rapporto_di_indebitamento_anno_1
Rapporto di indebitamento\nAnno - 2	rapporto_di_indebitamento_anno_2
Rapporto di indebitamento\nAnno - 3	rapporto_di_indebitamento_anno_3
Rapporto di indebitamento\nAnno - 4	
Rapporto di indebitamento\nAnno - 5	
Oneri finanz. su fatt. (%)\n%\nUltimo anno disp.	oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp
Oneri finanz. su fatt. (%)\n%\nAnno - 1	oneri_finanz_su_fatt_percent_percent_anno_1
Oneri finanz. su fatt. (%)\n%\nAnno - 2	oneri_finanz_su_fatt_percent_percent_anno_2
Oneri finanz. su fatt. (%)\n%\nAnno - 3	oneri_finanz_su_fatt_percent_percent_anno_3
Oneri finanz. su fatt. (%)\n%\nAnno - 4	
Oneri finanz. su fatt. (%)\n%\nAnno - 5	
Indice di indip. finanz. (%)\n%\nUltimo anno disp.	indice_di_indip_finanz_percent_percent_ultimo_anno_disp
Indice di indip. finanz. (%)\n%\nAnno - 1	indice_di_indip_finanz_percent_percent_anno_1
Indice di indip. finanz. (%)\n%\nAnno - 2	indice_di_indip_finanz_percent_percent_anno_2
Indice di indip. finanz. (%)\n%\nAnno - 3	indice_di_indip_finanz_percent_percent_anno_3
Indice di indip. finanz. (%)\n%\nAnno - 4	
Indice di indip. finanz. (%)\n%\nAnno - 5	
Posizione finanziaria netta\nmigl EUR\nUltimo anno disp.	posizione_finanziaria_netta_migl_eur_ultimo_anno_disp
Posizione finanziaria netta\nmigl EUR\nAnno - 1	
Posizione finanziaria netta\nmigl EUR\nAnno - 2	
Posizione finanziaria netta\nmigl EUR\nAnno - 3	
Posizione finanziaria netta\nmigl EUR\nAnno - 4	
Posizione finanziaria netta\nmigl EUR\nAnno - 5	
Rotaz. cap. investito (volte)\nUltimo anno disp.	rotaz_cap_investito_volte_ultimo_anno_disp
Rotaz. cap. investito (volte)\nAnno - 1	rotaz_cap_investito_volte_anno_1
Rotaz. cap. investito (volte)\nAnno - 2	rotaz_cap_investito_volte_anno_2

Rotaz. cap. investito (volte)\nAnno - 3	rotaz_cap_investito_volte_anno_3
Rotaz. cap. investito (volte)\nAnno - 4	
Rotaz. cap. investito (volte)\nAnno - 5	
Giac. media delle scorte (gg)\nUltimo anno disp.	
Giac. media delle scorte (gg)\nAnno - 1	
Giac. media delle scorte (gg)\nAnno - 2	
Giac. media delle scorte (gg)\nAnno - 3	
Giac. media delle scorte (gg)\nAnno - 4	
Giac. media delle scorte (gg)\nAnno - 5	
Durata media dei crediti al lordo IVA (gg)\nUltimo anno disp.	
Durata media dei crediti al lordo IVA (gg)\nAnno - 1	
Durata media dei crediti al lordo IVA (gg)\nAnno - 2	
Durata media dei crediti al lordo IVA (gg)\nAnno - 3	
Durata media dei crediti al lordo IVA (gg)\nAnno - 4	
Durata media dei crediti al lordo IVA (gg)\nAnno - 5	
Durata media dei debiti al lordo IVA (gg)\nUltimo anno disp.	
Durata media dei debiti al lordo IVA (gg)\nAnno - 1	
Durata media dei debiti al lordo IVA (gg)\nAnno - 2	
Durata media dei debiti al lordo IVA (gg)\nAnno - 3	
Durata media dei debiti al lordo IVA (gg)\nAnno - 4	
Durata media dei debiti al lordo IVA (gg)\nAnno - 5	
EBITDA\nmigl EUR\nUltimo anno disp.	ebitda_migl_eur_ultimo_anno_disp
EBITDA\nmigl EUR\nAnno - 1	ebitda_migl_eur_anno_1
EBITDA\nmigl EUR\nAnno - 2	ebitda_migl_eur_anno_2
EBITDA\nmigl EUR\nAnno - 3	ebitda_migl_eur_anno_3
EBITDA\nmigl EUR\nAnno - 4	
EBITDA\nmigl EUR\nAnno - 5	
EBITDA/Vendite (%)\n%\nUltimo anno disp.	ebitda_vendite_percent_percent_ultimo_anno_disp
EBITDA/Vendite (%)\n%\nAnno - 1	ebitda_vendite_percent_percent_anno_1
EBITDA/Vendite (%)\n%\nAnno - 2	ebitda_vendite_percent_percent_anno_2
EBITDA/Vendite (%)\n%\nAnno - 3	ebitda_vendite_percent_percent_anno_3
EBITDA/Vendite (%)\n%\nAnno - 4	
EBITDA/Vendite (%)\n%\nAnno - 5	
Redditività di tutto il capitale investito (ROI) (%)\n%\nUltimo anno disp.	
Redditività di tutto il capitale investito (ROI) (%)\n%\nAnno - 1	
Redditività di tutto il capitale investito (ROI) (%)\n%\nAnno - 2	

Redditività di tutto il capitale investito (ROI) (%)\n%\nAnno - 3	
Redditività di tutto il capitale investito (ROI) (%)\n%\nAnno - 4	
Redditività di tutto il capitale investito (ROI) (%)\n%\nAnno - 5	
Redditività del capitale proprio (ROE) (%)\n%\nUltimo anno disp.	redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp
Redditività del capitale proprio (ROE) (%)\n%\nAnno - 1	redditivita_del_capitale_proprio_roe_percent_percent_anno_1
Redditività del capitale proprio (ROE) (%)\n%\nAnno - 2	redditivita_del_capitale_proprio_roe_percent_percent_anno_2
Redditività del capitale proprio (ROE) (%)\n%\nAnno - 3	redditivita_del_capitale_proprio_roe_percent_percent_anno_3
Redditività del capitale proprio (ROE) (%)\n%\nAnno - 4	
Redditività del capitale proprio (ROE) (%)\n%\nAnno - 5	
Redditività delle vendite (ROS) (%)\n%\nUltimo anno disp.	redditivita_delle_vendite_ros_percent_percent_ultimo_anno_disp
Redditività delle vendite (ROS) (%)\n%\nAnno - 1	
Redditività delle vendite (ROS) (%)\n%\nAnno - 2	
Redditività delle vendite (ROS) (%)\n%\nAnno - 3	
Redditività delle vendite (ROS) (%)\n%\nAnno - 4	
Redditività delle vendite (ROS) (%)\n%\nAnno - 5	
Capitale circolante netto\nmigl EUR\nUltimo anno disp.	capitale_circolante_netto_migl_eur_ultimo_anno_disp
Capitale circolante netto\nmigl EUR\nAnno - 1	capitale_circolante_netto_migl_eur_anno_1
Capitale circolante netto\nmigl EUR\nAnno - 2	capitale_circolante_netto_migl_eur_anno_2
Capitale circolante netto\nmigl EUR\nAnno - 3	capitale_circolante_netto_migl_eur_anno_3
Capitale circolante netto\nmigl EUR\nAnno - 4	
Capitale circolante netto\nmigl EUR\nAnno - 5	
Margine di tesoreria\nmigl EUR\nUltimo anno disp.	margine_di_tesoreria_migl_eur_ultimo_anno_disp
Margine di tesoreria\nmigl EUR\nAnno - 1	
Margine di tesoreria\nmigl EUR\nAnno - 2	
Margine di tesoreria\nmigl EUR\nAnno - 3	
Margine di tesoreria\nmigl EUR\nAnno - 4	
Margine di tesoreria\nmigl EUR\nAnno - 5	
TOTALE PASSIVO\nmigl EUR\nUltimo anno disp.	totale_passivo_migl_eur_ultimo_anno_disp
TOTALE PASSIVO\nmigl EUR\nAnno - 1	totale_passivo_migl_eur_anno_1
TOTALE PASSIVO\nmigl EUR\nAnno - 2	totale_passivo_migl_eur_anno_2
TOTALE PASSIVO\nmigl EUR\nAnno - 3	totale_passivo_migl_eur_anno_3
TOTALE PASSIVO\nmigl EUR\nAnno - 4	
TOTALE PASSIVO\nmigl EUR\nAnno - 5	

TOTALE PATRIMONIO NETTO\nmigl EUR\nUltimo anno disp.	totale_patrimonio_netto_migl_eur_ultimo_anno_disp
TOTALE PATRIMONIO NETTO\nmigl EUR\nAnno - 1	totale_patrimonio_netto_migl_eur_anno_1
TOTALE PATRIMONIO NETTO\nmigl EUR\nAnno - 2	totale_patrimonio_netto_migl_eur_anno_2
TOTALE PATRIMONIO NETTO\nmigl EUR\nAnno - 3	totale_patrimonio_netto_migl_eur_anno_3
TOTALE PATRIMONIO NETTO\nmigl EUR\nAnno - 4	
TOTALE PATRIMONIO NETTO\nmigl EUR\nAnno - 5	
TOTALE ATTIVO\nmigl EUR\nUltimo anno disp.	totale_attivo_migl_eur_ultimo_anno_disp
TOTALE ATTIVO\nmigl EUR\nAnno - 1	totale_attivo_migl_eur_anno_1
TOTALE ATTIVO\nmigl EUR\nAnno - 2	totale_attivo_migl_eur_anno_2
TOTALE ATTIVO\nmigl EUR\nAnno - 3	totale_attivo_migl_eur_anno_3
TOTALE ATTIVO\nmigl EUR\nAnno - 4	
TOTALE ATTIVO\nmigl EUR\nAnno - 5	
TOTALE DEBITI\nmigl EUR\nUltimo anno disp.	totale_debiti_migl_eur_ultimo_anno_disp
TOTALE DEBITI\nmigl EUR\nAnno - 1	totale_debiti_migl_eur_anno_1
TOTALE DEBITI\nmigl EUR\nAnno - 2	totale_debiti_migl_eur_anno_2
TOTALE DEBITI\nmigl EUR\nAnno - 3	totale_debiti_migl_eur_anno_3
TOTALE DEBITI\nmigl EUR\nAnno - 4	
TOTALE DEBITI\nmigl EUR\nAnno - 5	
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nUltimo anno disp.	utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nAnno - 1	utile_perdita_di_esercizio_migl_eur_anno_1
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nAnno - 2	utile_perdita_di_esercizio_migl_eur_anno_2
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nAnno - 3	utile_perdita_di_esercizio_migl_eur_anno_ 3
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nAnno - 4	
UTILE/PERDITA DI ESERCIZIO\nmigl EUR\nAnno - 5	
Prezzo di mercato - anno\nUltimo anno disp.	
Prezzo di mercato - anno\nAnno - 1	
Prezzo di mercato - anno\nAnno - 2	
Prezzo di mercato - anno\nAnno - 3	
Prezzo di mercato - anno\nAnno - 4	
Prezzo di mercato - anno\nAnno - 5	

Appendice D – *Calcoli di approfondimento, Area Geografica*

Di seguito sono riportati i calcoli che hanno supportato le considerazioni svolte nell'analisi descrittiva della variabile relativa all'area geografica presenti nel paragrafo 3.3.1, *Analisi descrittiva delle variabili - Area Geografica.*

Per ciascuna zona individuata (Nord, Centro e Sud Italia) viene presentata una tabella che indica il numero di osservazioni per provincia; è inoltre inclusa una colonna denominata "gap", nella quale viene evidenziata la differenza tra il numero di imprese sane e quello delle imprese che hanno manifestato problematiche di continuità aziendale.

Successivamente vengono illustrati i calcoli delle probabilità, utili a valutare in quale misura, in ciascuna provincia, sia più probabile riscontrare una prevalenza di imprese sane rispetto a quelle con criticità, in relazione alla numerosità delle imprese presenti sul territorio.

Area geografica - Nord Italia

	Imprese Sane "db.IMP.SANE.na.omit"	Imprese con problematiche di continuità aziendale "db.IMP.CATTIVE.na.omit"	Gap	Totale
Totale	23070	32175	-9105	55245
Milano	639	9383	-8744	10022
Mantova	4868	137	4731	5005
Bergamo	990	2945	-1955	3935
Verona	716	2450	-1734	3166
Brescia	1232	1802	-570	3034
Treviso	132	2247	-2115	2379
Torino	406	1759	-1353	2165
Como	399	1265	-866	1664
Monza e della Brianza	222	1258	-1036	1480
Sondrio	1344	57	1287	1401
Bologna	971	428	543	1399
Varese	659	728	-69	1387
Vicenza	258	1084	-826	1342
Padova	394	904	-510	1298
Modena	706	379	327	1085
Novara	881	146	735	1027
Trento	717	272	445	989
Vercelli	851	59	792	910

Venezia	85	812	-727	897
Parma	284	485	-201	769
Genova	470	281	189	751
Bolzano/Bozen	402	277	125	679
Valle d'Aosta	601	25	576	626
Cuneo	411	166	245	577
Forlì-Cesena	290	246	44	536
Trieste	377	135	242	512
Lecco	228	259	-31	487
Ravenna	353	122	231	475
Pavia	207	261	-54	468
Pordenone	260	189	71	449
Reggio nell'Emilia	318	119	199	437
Alessandria	258	85	173	343
Cremona	208	126	82	334
Ferrara	220	78	142	298
La Spezia	146	152	-6	298
Rovigo	153	127	26	280
Piacenza	196	74	122	270
Rimini	172	63	109	235
Belluno	94	135	-41	229
Lombardia	228	1	227	229
Savona	110	118	-8	228
Lodi	111	115	-4	226
Udine	73	129	-56	202
Gorizia	91	107	-16	198
Asti	99	81	18	180
Biella	76	56	20	132
Imperia	93	26	67	119
Verbano-Cusio-Ossola	69	21	48	90
Emilia-Romagna	2	1	1	3

Province con numerosità superiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	5
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	11
Totali	=	16

Province con numerosità inferiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	23
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	10
Totali	=	33

Calcolo Probabilità

	Numerosità > 1.000	*Numerosità < 1.000*
Probabilità che ci sia una prevalenza di imprese sane	31,25 %	69,70 %
Probabilità che ci sia una prevalenza di imprese con problemi di continuità aziendale	68,75 %	30,30 %

Area geografica - Centro Italia

	Imprese Sane "db.IMP.SANE.na.omit"	Imprese con problematiche di continuità aziendale "db.IMP.CATTIVE.na.omit"	Gap	Totale
Totale	13967	8409	5558	22376
Rieti	6346	60	6286	6406
Roma	183	3341	-3158	3524
Firenze	821	807	14	1628
Latina	576	206	370	782
Prato	110	637	-527	747
Ancona	371	366	5	737
Frosinone	450	158	292	608
Lucca	258	346	-88	604
Pesaro e Urbino	304	253	51	557
Chieti	300	253	47	553
Massa-Carrara	454	79	375	533
Pisa	206	267	-61	473
Viterbo	371	96	275	467
Perugia	266	193	73	459
Pistoia	291	146	145	437
Livorno	297	136	161	433
Pescara	329	100	229	429
Ascoli Piceno	183	230	-47	413
Arezzo	293	99	194	392
Siena	290	85	205	375
Teramo	163	164	-1	327
Macerata	165	152	13	317
L'Aquila	242	45	197	287
Terni	234	48	186	282
Grosseto	153	64	89	217
Fermo	127	74	53	201
Lazio	184	4	180	188

Province con numerosità superiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	2
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	1
Totali	=	3

Province con numerosità inferiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	19
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	5
Totali	=	24

Calcolo Probabilità

	Numerosità > 1.000	*Numerosità < 1.000*
Probabilità che ci sia una prevalenza di imprese sane	66,67 %	79,17 %
Probabilità che ci sia una prevalenza di imprese con problemi di continuità aziendale	33,33 %	20,83 %

Area geografica - Sud Italia

	Imprese Sane "db.IMP.SANE.na.omit"	Imprese con problematiche di continuità aziendale "db.IMP.CATTIVE.na.omit"	Gap	Totale
Totale	13267	5912	7355	19179
Napoli	2457	673	1784	3130
Salerno	1011	489	522	1500
Caserta	709	735	-26	1444
Bari	884	425	459	1309
Palermo	643	326	317	969
Catania	809	143	666	952
Foggia	465	228	237	693
Lecce	495	161	334	656
Cagliari	390	221	169	611
Sassari	458	148	310	606
Cosenza	443	151	292	594
Messina	401	171	230	572
Benevento	269	291	-22	560
Taranto	331	156	175	487
Reggio di Calabria	247	236	11	483
Avellino	298	101	197	399
Ragusa	298	75	223	373
Sud Sardegna	136	208	-72	344
Agrigento	219	117	102	336
Brindisi	240	69	171	309
Siracusa	243	66	177	309
Potenza	247	50	197	297
Trapani	270	10	260	280
Catanzaro	224	54	170	278
Barletta-Andria-Trani	203	72	131	275
Campobasso	152	54	98	206
Vibo Valentia	64	117	-53	181
Caltanissetta	135	36	99	171
Sicilia	3	160	-157	163
Matera	131	29	102	160
Crotone	93	37	56	130
Nuoro	90	39	51	129
Isernia	68	34	34	102

Oristano	72	22	50	94
Enna	69	8	61	77

Province con numerosità superiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	3
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	1
Totali	=	4

Province con numerosità inferiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	27
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	4
Totali	=	31

Calcolo Probabilità

	Numerosità > 1.000	*Numerosità < 1.000*
Probabilità che ci sia una prevalenza di imprese sane	75 %	87,10 %
Probabilità che ci sia una prevalenza di imprese con problemi di continuità aziendale	25 %	12,90 %

Totali

Province con numerosità superiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	10
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	13
Totali	=	23

Province con numerosità inferiore a 1.000 imprese

Totale province con prevalenza di imprese sane	=	69
Totale province con prevalenza di imprese soggette problematiche di continuità aziendale	=	19
Totali	=	88

Calcolo Probabilità

	Numerosità > 1.000	*Numerosità < 1.000*
Probabilità che ci sia una prevalenza di imprese sane	43,48 %	78,41 %
Probabilità che ci sia una prevalenza di imprese con problemi di continuità aziendale	56,52 %	21,59 %

Appendice E – *Calcoli di approfondimento, codice ATECO*

Di seguito è riportata la tabella che illustra la suddivisione settoriale secondo le prime due cifre del codice ATECO, con le relative frequenze distinte tra imprese sane e imprese che hanno manifestato problematiche di continuità aziendale.

Successivamente vengono presentati i calcoli delle probabilità, elaborati sulla base della prevalenza di imprese sane o problematiche in relazione alla numerosità delle imprese appartenenti a ciascun settore.

Sezione	**Divisori (prime 2 cifre)**	**Settore ufficiale**	**Frequenza Imprese Sane**	**Frequenza Imprese Problematiche**
A	01–03	*Agricoltura, silvicoltura e pesca*		
B	05–09	*Attività estrattiva*		
C	10–33	*Attività manifatturiere*	7386	8044
D	35	*Fornitura di energia elettrica, gas, vapore e aria condizionata*	457	105
E	36–39	*Fornitura di acqua; reti fognarie, attività di trattamento dei rifiuti e risanamento*	302	194
F	41–43	*Costruzioni*	6610	7011
G	45–47	*Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli*	9659	9815
H	49–53	*Trasporti e magazzinaggio*	1626	2430
I	55–56	*Servizi di alloggio e ristorazione*	3618	2333
J	58–63	*Informazione e comunicazione*	2498	2177
K	64–66	*Attività finanziarie e assicurative*	1089	450
L	68	*Attività immobiliari*	5970	4496
M	69–75	*Professioni tecniche e scientifiche*	4746	3668
N	77–82	*Attività amministrative e di servizi di supporto*	2673	2793
O	84	*Amministrazione pubblica e difesa; assicurazione sociale obbligatoria*		
P	85	*Istruzione*	574	351
Q	86–88	*Sanità*	791	988
R	89–93	*Attività artistiche, di intrattenimento e divertimento*	1001	880
S	94–96	*Altre attività di servizi*	672	635

T	97–98	*Attività di famiglie e convivenze come datori di lavoro per personale domestico; produzione di beni e servizi indifferenziati per uso proprio da parte di famiglie e convivenze*	1	1
U	99	*Attività di organizzazioni e organismi extraterritoriali*	2	8

Settori con numerosità superiore a 10.000 imprese

Totale settori con prevalenza di imprese sane	=	1
Totale settori con prevalenza di imprese soggette a problemi di continuità	=	3
Totali	=	4

Province con numerosità inferiore a 10.000 imprese

Totale settori con prevalenza di imprese sane	=	10
Totale settori con prevalenza di imprese soggette a problemi di continuità	=	4
Totali	=	14

Calcolo Probabilità

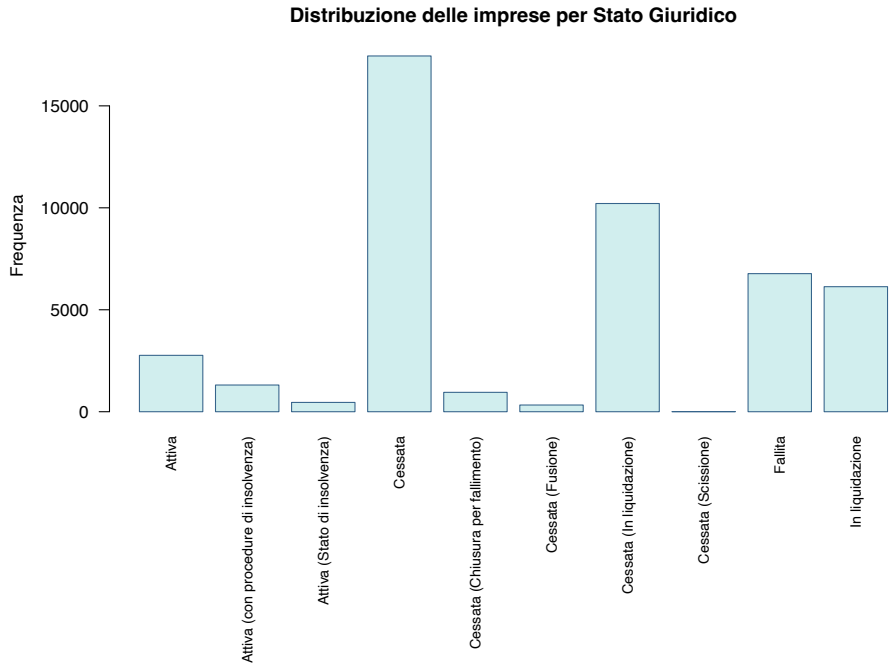
	Numerosità > 10.000	*Numerosità < 10.000*
Probabilità che il settore abbia una prevalenza di imprese sane	25 %	71,42 %
Probabilità che il settore abbia una prevalenza di imprese con problematiche di continuità aziendale	75 %	28,57 %

Appendice F – *Dati completi, Stato giuridico e Procedure Concorsuali*

Stato giuridico

Di seguito riportata la tabella relativa ai dati utilizzati per la costruzione del grafico in Figura 0.1: *Bar Plot della distribuzione delle osservazioni per Stato giuridico.*

	Frequenza	*Percentuale*
Attiva	2765	5.96 %
Attiva (con procedure di insolvenza)	1309	2.82 %
Attiva (Stato di insolvenza)	456	0.98 %
Cessata	17447	37.62 %
Cessata (Chiusura per fallimento)	950	2.05 %
Cessata (Fusione)	329	0.71 %
Cessata (In liquidazione)	10211	22.02 %
Cessata (Scissione)	8	0.02 %
Fallita	6772	14.6 %
In liquidazione	6132	13.22 %



Procedure concorsuali e di cessazione

Di seguito riportate tutte le tabelle complete relative alle frequenze delle procedure concorsuali e di cessazione in base allo stato giuridico precedentemente analizzato.

Impresa In liquidazione:

	Frequenza	Percentuale
Liquidazione volontaria	4537	*73,99*
Liquidazione coatta amministrativa	971	*15,83*
Liquidazione giudiziale	349	*5,69*
Scioglimento e liquidazione	103	*1,68*
Trasferimento in altra provincia	98	*1,6*
Scioglimento	49	*0,8*
Scioglimento per atto dell'autorità	9	*0,15*
Scioglimento senza liquidazione finalizzato alla cancellazione ai sensi dell'art. 40 C. 2 DL N. 76/2020	6	*0,1*
Chiusura dell'unità locale	3	*0,05*
Liquidazione controllata	2	*0,03*
Liquidazione giudiziaria	2	*0,03*
Cessazione delle attività nella provincia	1	*0,02*
Liquidazione	1	*0,02*
Trasformazione di natura giuridica	1	*0,02*

Impresa Attiva:

	Frequenza	Percentuale
Liquidazione volontaria	1337	*48,35*
Concordato preventivo	613	*22,17*
Liquidazione	264	*9,55*
Scioglimento	163	*5,9*
Trasferimento in altra provincia	118	*4,27*
Amministrazione giudiziaria	80	*2,89*
Concordato fallimentare	49	*1,77*
Fallimento	33	*1,19*
Liquidazione coatta amministrativa	25	*0,9*
Scioglimento e liquidazione	20	*0,72*
Accordo di ristrutturazione dei debiti	16	*0,58*

Concordato preventivo omologato	12	*0,43*
Stato di insolvenza	8	*0,29*
Cancellata d'ufficio ai sensi art, 2490 c,c,	5	*0,18*
Scioglimento per atto dell'autorità	5	*0,18*
Amministrazione straordinaria	3	*0,11*
Chiusura della liquidazione	3	*0,11*
Cessazione delle attività nella provincia	2	*0,07*
Chiusura del fallimento	2	*0,07*
Procedimento unitario	2	*0,07*
Accordi di ristrutturazione dei debiti	1	*0,04*
Amministrazione controllata	1	*0,04*
Fusione mediante incorporazione in altra società	1	*0,04*
Provvedimento di cancellazione dal registro delle imprese	1	*0,04*
Scioglimento senza liquidazione finalizzato alla cancellazione ai sensi dell'art, 40 C, 2 DL N, 76/2020	1	*0,04*

Impresa Cessata:

	Frequenza	Percentuale
Cancellazione dal registro delle imprese	10996	*63,03*
Cancellata d'ufficio ai sensi art. 2490 c.c.	3970	*22,75*
Provvedimento di cancellazione dal registro delle imprese	1805	*10,35*
Cessazione di ogni attivita'	220	*1,26*
Scioglimento	216	*1,24*
Altre cause	165	*0,95*
Cancellazione d'ufficio dell'iscrizione al r.i.	34	*0,19*
Cancellata d'ufficio ai sensi dell'art. 40 C.5 DL 76/2020 CONV. L. 120/2020	14	*0,08*
Trasferimento sede all'estero	10	*0,06*
Cessazione d'ufficio	9	*0,05*
Decreto cancellazione tribunale	4	*0,02*
Cancellata d'ufficio ai sensi dpr 23/7/2004 n.247	2	*0,01*
Cancellazione per comunicazione piano di riparto	2	*0,01*

Impresa Cessata (In liquidazione):

	Frequenza	Percentuale
Chiusura della liquidazione	10066	*98,58*
Chiusura per fallimento o liquidazione	114	*1,12*
Scioglimento e messa in liquidazione	27	*0,26*
Chiusura per liquidazione	4	*0,04*

Impresa Fallita:

	Frequenza	Percentuale
Chiusura del fallimento	2462	*36,36*
Fallimento	1996	*29,47*
Cancellazione dal registro delle imprese	821	*12,12*
Cancellata d'ufficio ai sensi art. 2490 c.c.	372	*5,49*
Provvedimento di cancellazione dal registro delle imprese	261	*3,85*
Liquidazione coatta amministrativa	202	*2,98*
Chiusura della liquidazione	188	*2,78*
Stato di insolvenza	96	*1,42*
Liquidazione giudiziale	94	*1,39*
Concordato preventivo	63	*0,93*
Chiusura per fallimento o liquidazione	58	*0,86*
Cessazione d'ufficio	31	*0,46*
Altre cause	25	*0,37*
Amministrazione straordinaria	20	*0,3*
Liquidazione volontaria	15	*0,22*
Cancellazione d'ufficio dell'iscrizione al r.i.	14	*0,21*
Fusione mediante incorporazione in altra società	8	*0,12*
Trasferimento in altra provincia	8	*0,12*
Amministrazione giudiziaria	7	*0,1*
Cancellata d'ufficio ai sensi dell'art. 40 C.5 DL 76/2020 CONV. L. 120/2020	4	*0,06*
Concordato fallimentare	4	*0,06*
Scioglimento e liquidazione	4	*0,06*
Cessazione di ogni attivita'	3	*0,04*
Scioglimento	3	*0,04*
Chiusura per fallimento	2	*0,03*
Procedimento unitario	2	*0,03*

Scioglimento senza liquidazione finalizzato alla cancellazione ai sensi dell'art. 40 C. 2 DL N. 76/2020	2	*0,03*
Cessione azienda	1	*0,01*
Chiusura per liquidazione	1	*0,01*
Decreto cancellazione tribunale	1	*0,01*
Mancata ricostituzione della pluralità dei soci	1	*0,01*
Piano di ristrutturazione omologato (CCI)	1	*0,01*
Scioglimento per atto dell'autorità	1	*0,01*
Scissione	1	*0,01*

Appendice G – *Dati completi, Indici e voci di bilancio*

Nella presente appendice sono riportati tutti i dati utilizzati per la costruzione dei grafici del paragrafo 4.3.2, *indici e voci di bilancio, andamento medio e mediano*.

Le tabelle illustrano, per ciascun indice, i valori medi o mediani rilevati annualmente, dall'ultimo esercizio disponibile fino al terzo anno precedente, al fine di analizzare l'evoluzione temporale delle diverse grandezze considerate.

Indice di liquidità - media

	Imprese sane	Imprese problematiche
Anno -3	1,960	1,329
Anno -2	1,965	1,345
Anno -1	1,991	1,343
Ultimo anno	2,034	1,500

Indice di indebitamento a breve termine - media

	Imprese sane	Imprese problematiche
Anno -3	0,757	0,825
Anno -2	0,759	0,824
Anno -1	0,767	0,829
Ultimo anno	0,775	0,848

Indice di indebitamento a lungo termine - media

	Imprese sane	Imprese problematiche
Anno -3	0,243	0,174
Anno -2	0,241	0,176
Anno -1	0,233	0,171
Ultimo anno	0,226	0,152

Indice di copertura delle immobilizzazioni patrimoniali - media

	Imprese sane	Imprese problematiche
Anno -3	0,987	1,058
Anno -2	0,967	1,000
Anno -1	0,940	0,792
Ultimo anno	0,895	0,359

Rapporto di indebitamento - media

	Imprese sane	Imprese problematiche
Anno -3	10,787	24,024
Anno -2	10,075	23,706
Anno -1	9,111	20,604
Ultimo anno	9,776	10,612

Oneri finanziari su fatturato - media

	Imprese sane	Imprese problematiche
Anno -3	2,381	3,651
Anno -2	2,520	3,539
Anno -1	2,847	3,535
Ultimo anno	3,192	3,329

Indice di indipendenza finanziaria - media

	Imprese sane	Imprese problematiche
Anno -3	31,901	18,334
Anno -2	31,415	17,274
Anno -1	32,178	13,465
Ultimo anno	33,418	8,630

Rotazione del capitale investito - media

	Imprese sane	Imprese problematiche
Anno -3	0,939	1,139
Anno -2	1,008	1,142
Anno -1	1,002	1,182
Ultimo anno	0,976	1,413

EBITDA

EBITDA, valore assoluto in migliaia di euro – media

	Imprese sane	Imprese problematiche
Anno -3	291,231	44,935
Anno -2	335,756	50,335
Anno -1	371,293	-10,623
Ultimo anno	381,559	-133,093

EBITDA margin, in valore percentuale – mediana

	Imprese sane	Imprese problematiche
Anno -3	7,870 %	4,230 %
Anno -2	8,030 %	3,890 %
Anno -1	7,920 %	2,760 %
Ultimo anno	8,020 %	-1,470 %

Redditività del capitale proprio, ROE - media

	Imprese sane	Imprese problematiche
Anno -3	5,887	-11,103
Anno -2	6,346	-14,407
Anno -1	3,516	-21,349
Ultimo anno	2,084	-39,089

Capitale circolante netto - media

	Imprese sane	Imprese problematiche
Anno -3	628,366	183,994
Anno -2	741,616	103,387
Anno -1	1.530,571	-40,514
Ultimo anno	856,910	-362,931

Totale attivo e Totale passivo - media

	Imprese sane	Imprese problematiche
Anno -3	4.445,119	3.594,638
Anno -2	4.939,641	3.651,996
Anno -1	5.924,907	3.643,285
Ultimo anno	5.385,761	3.170,331

Totale patrimonio netto - media

	Imprese sane	Imprese problematiche
Anno -3	1.712,609	531,482
Anno -2	1.856,083	474,712
Anno -1	2.049,771	323,927
Ultimo anno	2.257,544	-83,913

Totale debiti - mediana

	Imprese sane	Imprese problematiche
Anno -3	195,360	268,627
Anno -2	226,543	293,003
Anno -1	241,053	316,836
Ultimo anno	243,879	257,056

Utile perdita – media

	Imprese sane	Imprese problematiche
Anno -3	149,051	-123,937
Anno -2	173,867	-111,284
Anno -1	245,319	-203,284
Ultimo anno	229,433	-407,740

Appendice H – *Dati completi, correlazione e test di Mann-Whitney*

Coefficienti di correlazione tra indici e voci di bilancio

Di seguito sono riportati tutti i coefficienti di correlazione calcolati tra le variabili relative agli indici o alle voci di bilancio. I risultati sono raccolti nella tabella sottostante, distinta per imprese sane e imprese problematiche: le correlazioni riferite alle imprese sane sono ordinate in base al valore assoluto, mentre quelle relative alle imprese problematiche sono riportate per corrispondenza. Sono evidenziati in giallo i coefficienti con valore superiore a 0,8, soglia oltre la quale si considera la presenza di ripetizione delle informazioni.

Variabile_1	**Variabile_2**	**Correlazione imprese sane**	**Correlazione imprese problematiche**
i_indebitamento_breve	i_indebitamento_lungo	-0,999751655	-0,999039424
EBITDA_margin	ROS	0,882574755	0,946061157
tot_debiti	tot_passivo_attivo	0,863069977	0,852439435
EBITDA	utile_perdita	0,80688636	0,837163862
ROE	utile_perdita	0,789739138	0,764403212
tot_passivo_attivo	tot_patrimonio_netto	0,762472152	0,107403484
n_dipendenti	ricavi_di_vendita	0,747311496	0,534561187
i_liquidità	margine_tesoreria	0,716701117	0,774923189
ROS	utile_perdita	0,694329538	0,812054786
ROE	ROS	0,689536668	0,74306038
EBITDA	ricavi_di_vendita	0,680971073	-0,004178828
EBITDA	ROS	0,648763072	0,85314578
EBITDA	EBITDA_margin	0,637644263	0,881789563
ricavi_di_vendita	rotazione_cap_investito	0,625239363	0,260355588
EBITDA	tot_patrimonio_netto	0,624237989	0,60457834
i_indipendenza_fin	tot_patrimonio_netto	0,619633191	0,833318756
EBITDA	ROE	0,607959368	0,687940245
ricavi_di_vendita	tot_passivo_attivo	0,606909661	0,648006711
CCN	tot_patrimonio_netto	0,591757527	0,659529463
i_indipendenza_fin	i_liquidità	0,585218471	0,654416253
ricavi_di_vendita	tot_debiti	0,584470781	0,619849606
CCN	i_liquidità	0,57214171	0,701407748
tot_patrimonio_netto	utile_perdita	0,570095816	0,603269934
EBITDA	tot_passivo_attivo	0,567366318	0,012385145

EBITDA_margin	utile_perdita	0,559977659	0,770257816
EBITDA_margin	ROE	0,541427506	0,704009549
CCN	margine_tesoreria	0,533277698	0,68304508
i_indipendenza_fin	rapporto_indebitamento	-0,522147322	0,525327358
ricavi_di_vendita	utile_perdita	0,511215479	-0,146592294
n_dipendenti	rotazione_cap_investito	0,510437246	0,119811425
ricavi_di_vendita	tot_patrimonio_netto	0,494586612	0,022052192
CCN	tot_passivo_attivo	0,470826068	-0,001545677
i_copertura_immobilizzazizoni	rapporto_indebitamento	0,470576286	0,80953843
CCN	utile_perdita	0,455130259	0,494331647
tot_debiti	tot_patrimonio_netto	0,450925886	-0,256313655
EBITDA	tot_debiti	0,446553741	-0,186423007
n_dipendenti	tot_debiti	0,440782251	0,43118226
EBITDA	n_dipendenti	0,440221942	-0,144856922
oneri_fatturato	PFN	0,434786387	0,526559201
CCN	ricavi_di_vendita	0,432091899	-0,032659093
CCN	EBITDA	0,429664245	0,442157063
i_indipendenza_fin	margine_tesoreria	0,429289998	0,574212188
n_dipendenti	tot_passivo_attivo	0,411217515	0,378175639
tot_passivo_attivo	utile_perdita	0,410380573	-0,152004076
margine_tesoreria	PFN	-0,403966163	-0,529745983
CCN	i_indipendenza_fin	0,377311272	0,611775405
i_liquidità	PFN	-0,364401917	-0,44286377
i_indipendenza_fin	utile_perdita	0,361068553	0,576440082
i_indebitamento_lungo	oneri_fatturato	0,356874241	0,285259969
i_indebitamento_breve	oneri_fatturato	-0,356821759	-0,285307704
ricavi_di_vendita	ROE	0,343216451	0,030856604
i_indipendenza_fin	ROS	0,33834661	0,548539499
EBITDA_margin	tot_patrimonio_netto	0,333877062	0,537854253
rapporto_indebitamento	tot_debiti	0,331853854	-0,062909409
ROE	rotazione_cap_investito	0,329969669	-0,095592501
EBITDA_margin	i_indipendenza_fin	0,326973581	0,520628847
i_indebitamento_lungo	PFN	0,325996941	0,319496061
i_indebitamento_breve	PFN	-0,325971821	-0,31948498
ROS	tot_patrimonio_netto	0,324843229	0,554857983
i_liquidità	rapporto_indebitamento	-0,321438365	0,23794609

i_indipendenza_fin	PFN	-0,315566039	-0,393242869
i_indipendenza_fin	tot_debiti	-0,313829191	-0,475435005
i_liquidità	tot_debiti	-0,308037519	-0,461390988
n_dipendenti	utile_perdita	0,297803233	-0,235526925
ROE	tot_patrimonio_netto	0,296052844	0,564034265
i_liquidità	tot_patrimonio_netto	0,292187323	0,54511913
i_indebitamento_breve	tot_debiti	-0,288243168	-0,319163842
i_indebitamento_lungo	tot_debiti	0,288162817	0,318375981
EBITDA	i_copertura_immobilizzazizoni	0,284768468	0,462168704
margine_tesoreria	utile_perdita	0,282067538	0,477188274
n_dipendenti	tot_patrimonio_netto	0,281782794	-0,136733585
CCN	n_dipendenti	0,280112524	-0,187273906
EBITDA	rotazione_cap_investito	0,273811455	-0,017482161
oneri_fatturato	rotazione_cap_investito	-0,273504328	-0,25291699
CCN	ROE	0,273359193	0,416637882
i_liquidità	utile_perdita	0,271389654	0,443924527
CCN	PFN	-0,262483082	-0,343766077
i_copertura_immobilizzazizoni	tot_passivo_attivo	0,260077389	0,051926061
i_liquidità	ROS	0,258423342	0,417034839
CCN	tot_debiti	0,257421277	-0,297054265
tot_debiti	utile_perdita	0,253145497	-0,360413604
i_indipendenza_fin	ROE	0,252591301	0,571849493
EBITDA	i_indipendenza_fin	0,250542369	0,517320354
i_copertura_immobilizzazizoni	i_indebitamento_breve	-0,249906486	-0,007949608
i_copertura_immobilizzazizoni	i_indebitamento_lungo	0,24987704	0,007975864
EBITDA_margin	i_copertura_immobilizzazizoni	0,248029765	0,419999548
margine_tesoreria	tot_patrimonio_netto	0,245324687	0,488789661
i_liquidità	oneri_fatturato	-0,244242203	-0,353150723
i_copertura_immobilizzazizoni	tot_debiti	0,243073713	-0,12474677
margine_tesoreria	rapporto_indebitamento	-0,242420091	0,208325509
rotazione_cap_investito	utile_perdita	0,241617508	0,029941874
margine_tesoreria	oneri_fatturato	-0,238001854	-0,379753826
i_copertura_immobilizzazizoni	tot_patrimonio_netto	0,234602353	0,671059415
PFN	utile_perdita	-0,232857042	-0,352739841
i_liquidità	ROE	0,232287517	0,401563604
margine_tesoreria	tot_debiti	-0,230477198	-0,634653291

margine_tesoreria	ROS	0,230091112	0,366991402
margine_tesoreria	ROE	0,227125166	0,329536727
CCN	ROS	0,2252683	0,43701623
oneri_fatturato	tot_debiti	0,222734287	0,375240066
PFN	rapporto_indebitamento	0,215783556	-0,105621472
EBITDA_margin	i_liquidità	0,209014696	0,377868074
i_indebitamento_breve	tot_passivo_attivo	-0,204922732	-0,268464499
i_indebitamento_lungo	tot_passivo_attivo	0,204872145	0,268144119
PFN	ROE	-0,199470487	-0,231575935
PFN	tot_debiti	0,188912386	0,506567453
i_indipendenza_fin	oneri_fatturato	-0,187765213	-0,317546461
ricavi_di_vendita	ROS	0,187446336	0,03193068
i_copertura_immobilizzazizoni	ricavi_di_vendita	0,186332404	0,008416737
i_copertura_immobilizzazizoni	PFN	0,185808741	-0,138763461
oneri_fatturato	ROE	-0,183741774	-0,216245877
EBITDA	margine_tesoreria	0,183632154	0,356848253
n_dipendenti	ROE	0,177189755	-0,128735623
EBITDA_margin	tot_passivo_attivo	0,168722222	0,030170749
EBITDA_margin	margine_tesoreria	0,165272765	0,332716962
i_copertura_immobilizzazizoni	oneri_fatturato	0,164513894	-0,128848962
rapporto_indebitamento	ROE	0,164336626	0,421689147
i_indebitamento_breve	i_indipendenza_fin	0,163223957	0,150731839
i_indebitamento_lungo	i_indipendenza_fin	-0,163188611	-0,14994092
i_copertura_immobilizzazizoni	n_dipendenti	0,16199632	-0,062789162
PFN	tot_patrimonio_netto	-0,161537492	-0,314883776
EBITDA	i_liquidità	0,157999337	0,363204328
i_indebitamento_breve	rapporto_indebitamento	-0,153250397	0,00484243
i_indebitamento_lungo	rapporto_indebitamento	0,153235524	-0,004835198
i_copertura_immobilizzazizoni	i_liquidità	-0,152175104	0,293088899
i_copertura_immobilizzazizoni	ROE	0,151811701	0,41483838
oneri_fatturato	utile_perdita	-0,149244404	-0,328434141
i_indebitamento_lungo	rotazione_cap_investito	-0,147872509	-0,239480086
i_indebitamento_breve	rotazione_cap_investito	0,147718003	0,239616843
rapporto_indebitamento	ricavi_di_vendita	0,144670425	0,057465825
CCN	EBITDA_margin	0,141336925	0,385456524
n_dipendenti	rapporto_indebitamento	0,141093779	-0,047878408

rapporto_indebitamento	tot_patrimonio_netto	-0,141090305	0,620011507
PFN	rotazione_cap_investito	-0,140448647	-0,19880756
PFN	ROS	-0,134429872	-0,239853676
i_copertura_immobilizzazizoni	ROS	0,134213323	0,421031811
i_copertura_immobilizzazizoni	margine_tesoreria	-0,133929802	0,247541644
ROS	tot_passivo_attivo	0,129569744	0,013559363
oneri_fatturato	tot_passivo_attivo	0,127305434	0,252461351
ROE	tot_passivo_attivo	0,127284366	0,100449536
CCN	oneri_fatturato	-0,122653053	-0,26068426
oneri_fatturato	rapporto_indebitamento	0,121737216	-0,113653589
margine_tesoreria	rotazione_cap_investito	0,121658934	0,1419207
rapporto_indebitamento	tot_passivo_attivo	0,118659486	0,059195684
CCN	i_indebitamento_breve	-0,118636572	-0,122029663
CCN	i_indebitamento_lungo	0,118601524	0,122356989
CCN	i_copertura_immobilizzazizoni	-0,117645927	0,351247764
i_indipendenza_fin	n_dipendenti	-0,112807288	-0,273206477
CCN	rotazione_cap_investito	0,110540988	-0,045752532
rotazione_cap_investito	tot_passivo_attivo	-0,108598127	-0,487437966
EBITDA_margin	i_indebitamento_breve	-0,108135578	0,029502616
EBITDA_margin	i_indebitamento_lungo	0,107962336	-0,029341887
i_copertura_immobilizzazizoni	utile_perdita	0,098858883	0,428160383
EBITDA	PFN	-0,095388912	-0,222583587
oneri_fatturato	ricavi_di_vendita	-0,094935423	0,086564991
margine_tesoreria	ricavi_di_vendita	0,094301094	-0,29823953
i_indipendenza_fin	rotazione_cap_investito	-0,091814566	-0,03337873
rapporto_indebitamento	rotazione_cap_investito	0,089735075	-0,021722638
PFN	ricavi_di_vendita	-0,089198264	0,207165873
EBITDA	i_indebitamento_breve	-0,087406647	0,032781495
EBITDA	i_indebitamento_lungo	0,087289556	-0,0326159
EBITDA_margin	rotazione_cap_investito	-0,085878752	-0,035818394
EBITDA_margin	ricavi_di_vendita	0,085678705	0,011828994
CCN	rapporto_indebitamento	-0,083945229	0,356826473
i_indipendenza_fin	tot_passivo_attivo	0,08102952	-0,121308266
EBITDA_margin	rapporto_indebitamento	-0,079171948	0,37667426
ROS	rotazione_cap_investito	0,076324562	0,011305254
i_indebitamento_breve	tot_patrimonio_netto	-0,070720533	0,051996223

i_indebitamento_lungo	tot_patrimonio_netto	0,070691738	-0,051687864
rapporto_indebitamento	ROS	-0,068993716	0,399352927
rotazione_cap_investito	tot_patrimonio_netto	-0,067864542	-0,121963209
EBITDA_margin	PFN	-0,065057677	-0,205646646
EBITDA_margin	n_dipendenti	-0,064990416	-0,157469089
i_indebitamento_lungo	ROE	-0,063868145	-0,061241388
i_indebitamento_breve	ROE	0,063806035	0,061456674
rapporto_indebitamento	utile_perdita	-0,057856771	0,417520288
i_indebitamento_lungo	i_liquidità	0,055980996	-0,005141773
i_indebitamento_breve	i_liquidità	-0,055932917	0,005835815
EBITDA	rapporto_indebitamento	0,051102794	0,437231185
i_liquidità	rotazione_cap_investito	0,047899463	0,029194665
margine_tesoreria	n_dipendenti	0,046858054	-0,279197377
ROE	tot_debiti	0,046201655	-0,131546858
oneri_fatturato	ROS	-0,044639841	-0,228329127
n_dipendenti	oneri_fatturato	-0,040921984	0,132204532
n_dipendenti	PFN	-0,034813847	0,20464946
EBITDA_margin	tot_debiti	0,033082211	-0,190084968
rotazione_cap_investito	tot_debiti	-0,032073186	-0,346080729
i_liquidità	ricavi_di_vendita	0,031526146	-0,106565122
i_indebitamento_breve	ROS	-0,030293226	0,062562177
i_indebitamento_lungo	ROS	0,030124882	-0,062431852
i_indebitamento_lungo	utile_perdita	-0,028175927	-0,11293236
i_indebitamento_breve	utile_perdita	0,028116719	0,1131816
i_liquidità	tot_passivo_attivo	-0,025911606	-0,130332557
i_liquidità	n_dipendenti	-0,024708815	-0,170511835
i_indebitamento_breve	ricavi_di_vendita	-0,023190553	-0,096418527
i_indebitamento_lungo	ricavi_di_vendita	0,023080291	0,096076121
i_indebitamento_breve	n_dipendenti	-0,021941005	-0,084624083
i_indebitamento_lungo	n_dipendenti	0,021891682	0,08409053
PFN	tot_passivo_attivo	0,021817273	0,34480621
EBITDA_margin	oneri_fatturato	0,020700133	-0,193542087
margine_tesoreria	tot_passivo_attivo	-0,019020852	-0,38744122
EBITDA	oneri_fatturato	-0,017803817	-0,201908725
n_dipendenti	ROS	0,015376362	-0,156945674
i_copertura_immobilizzazizoni	rotazione_cap_investito	-0,010691374	-0,06496794

oneri_fatturato	tot_patrimonio_netto	-0,009797576	-0,236980194
i_copertura_immobilizzazizoni	i_indipendenza_fin	-0,007764812	0,596753945
ROS	tot_debiti	0,006602209	-0,208960436
i_indebitamento_breve	margine_tesoreria	0,004650694	0,053690157
i_indebitamento_lungo	margine_tesoreria	-0,004587638	-0,053297719
i_indipendenza_fin	ricavi_di_vendita	0,004350163	-0,150541501

Risultati completi della verifica statistica delle differenze tra i due gruppi

Nella seguente tabella vengono riportati tutti i risultati completi ottenuti nell'analisi della differenza tra il gruppo di imprese sane e il gruppo di imprese problematiche. Per questa analisi si è utilizzato il t.test (risultati riportati nella colonna corrispondente), per il confronto tra le medie dei gruppi, e il test di Mann-Whitney (risultati riportati nella colonna corrispondente), che confronta la differenza tra i gruppi attraverso il metodo dei ranghi, per maggiore robustezza dei risultati data la distribuzione non normale delle variabili.

Sono evidenziati in giallo i valori che presentano un p-value superiore a 0,05 limite considerato per l'analisi di non significatività.

Variabile	t. test	t. wilcoxn
indice_di_liquidita_ultimo_anno_disp	0	0
indice_di_liquidita_anno_1	0	0
indice_di_liquidita_anno_2	0	0
indice_di_liquidita_anno_3	0	0
indice_di_indebitam_a_breve_percent_ultimo_anno_disp	0	0
indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	0	0
indice_di_indip_finanz_percent_percent_ultimo_anno_disp	0	0
indice_di_indip_finanz_percent_percent_anno_1	0	0
indice_di_indip_finanz_percent_percent_anno_2	0	0
indice_di_indip_finanz_percent_percent_anno_3	0	0
rotaz_cap_investito_volte_ultimo_anno_disp	0	0
redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp	0	0
redditivita_del_capitale_proprio_roe_percent_percent_anno_1	0	0
redditivita_del_capitale_proprio_roe_percent_percent_anno_2	0	0
redditivita_del_capitale_proprio_roe_percent_percent_anno_3	0	0
redditivita_delle_vendite_ros_percent_percent_ultimo_anno_disp	0	0
indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp	1,02E-315	0
indice_di_indebitam_a_lungo_percent_anno_3	2,50E-289	3,038E-271
indice_di_indebitam_a_breve_percent_anno_3	1,01E-287	2,2641E-266
indice_di_indebitam_a_breve_percent_anno_2	1,93E-275	4,0828E-302
indice_di_indebitam_a_lungo_percent_anno_2	6,49E-275	1,7893E-300
indice_di_indebitam_a_lungo_percent_anno_1	1,69E-254	2,5852E-305
indice_di_indebitam_a_breve_percent_anno_1	2,03E-254	6,7576E-306
rotaz_cap_investito_volte_anno_3	6,93E-231	5,5422E-299
rotaz_cap_investito_volte_anno_1	8,80E-175	1,5545E-215

ebitda_vendite_percent_percent_ultimo_anno_disp	2,22E-125	0
oneri_finanz_su_fatt_percent_percent_anno_3	2,08E-107	0
rotaz_cap_investito_volte_anno_2	3,88E-101	4,8368E-129
oneri_finanz_su_fatt_percent_percent_anno_2	4,31E-74	0
ebitda_migl_eur_ultimo_anno_disp	1,82E-49	0
ebitda_vendite_percent_percent_anno_1	3,01E-40	0
ebitda_migl_eur_anno_1	2,34E-35	0
utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	2,69E-35	0
oneri_finanz_su_fatt_percent_percent_anno_1	6,46E-34	7,7284E-207
rapporto_di_indebitamento_anno_3	1,37E-33	0
ebitda_vendite_percent_percent_anno_2	7,50E-32	0
totale_patrimonio_netto_migl_eur_ultimo_anno_disp	3,07E-31	0
utile_perdita_di_esercizio_migl_eur_anno_2	5,32E-25	0
totale_patrimonio_netto_migl_eur_anno_1	6,38E-25	0
rapporto_di_indebitamento_anno_2	9,92E-25	4,2607E-241
rapporto_di_indebitamento_anno_1	4,16E-23	1,06149E-07
ebitda_migl_eur_anno_2	1,00E-22	0
indice_di_copertura_delle_immob_patrimoniale_percent_anno_1	6,27E-21	0
totale_patrimonio_netto_migl_eur_anno_2	6,65E-21	0
posizione_finanziaria_netta_migl_eur_ultimo_anno_disp	8,96E-18	0
ateco_2007_codice	6,22E-17	2,55412E-40
totale_patrimonio_netto_migl_eur_anno_3	7,69E-17	0
capitale_circolante_netto_migl_eur_ultimo_anno_disp	1,66E-15	0
utile_perdita_di_esercizio_migl_eur_anno_3	3,47E-14	0
capitale_circolante_netto_migl_eur_anno_2	3,08E-13	0
margine_di_tesoreria_migl_eur_ultimo_anno_disp	3,86E-12	0
utile_perdita_di_esercizio_migl_eur_anno_1	5,63E-12	0
ebitda_migl_eur_anno_3	3,19E-11	2,0411E-188
ebitda_vendite_percent_percent_anno_3	1,09E-10	0
capitale_circolante_netto_migl_eur_anno_3	6,00E-10	1,0223E-298
ricavi_delle_vendite_migl_eur_ultimo_anno_disp	5,32E-06	1,10643E-07
indice_di_copertura_delle_immob_patrimoniale_percent_anno_3	1,40E-05	0,067200885
totale_attivo_migl_eur_ultimo_anno_disp	4,63E-05	0
totale_passivo_migl_eur_ultimo_anno_disp	4,63E-05	0
totale_passivo_migl_eur_anno_2	0,01392153	0,187796299
totale_attivo_migl_eur_anno_2	0,014682217	0,189512622
oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	0,016355619	4,23456E-98

totale_attivo_migl_eur_anno_1	0,018195633	2,94232E-14
totale_passivo_migl_eur_anno_1	0,018196076	2,94944E-14
indice_di_copertura_delle_immob_patrimoniale_percent_anno_2	0,043742819	1,02414E-59
capitale_circolante_netto_migl_eur_anno_1	0,056838207	0
totale_attivo_migl_eur_anno_3	0,069166497	1,84923E-11
totale_passivo_migl_eur_anno_3	0,069272868	1,5896E-11
dipendenti_ultimo_anno_disp	0,146242618	0,000426148
totale_debiti_migl_eur_anno_3	0,269359232	1,5702E-165
rapporto_di_indebitamento_ultimo_anno_disp	0,339491335	0
totale_debiti_migl_eur_anno_1	0,479597314	1,8812E-126
totale_debiti_migl_eur_anno_2	0,800340766	2,7702E-112
totale_debiti_migl_eur_ultimo_anno_disp	0,849870206	1,48836E-07

Appendice I – *Legenda grafici DAG*

In questa appendice si riporta la legenda degli acronimi utilizzato dal grafico 4.1 al 4.10 per la rappresentazione dei grafici DAG.

- **CCN** – Capitale circolante netto *(in migliaia di euro)*
- **EBITDA** – Margine operativo lordo *(in migliaia di euro)*
- **EBITDA%** – EBITDA su vendite (*in percentuale*)
- **I.Cop.Immo.** – Indice di copertura delle immobilizzazioni patrimoniali
- **I.Indeb.** – Indice di indebitamento a lungo termine
- **I.Indip.Fin** – Indice di indipendenza finanziaria
- **I.Liq.** – Indice di liquidità
- **O/F** – Oneri finanziari su fatturato
- **Rapp.Ind.** – Rapporto di indebitamento
- **ROE** – Redditività del capitale proprio (ROE)
- **Rot.CI** – Rotazione del capitale investito
- **T.D.** – Totale debiti *(in migliaia di euro)*
- **T.PN** – Totale patrimonio netto *(in migliaia di euro)*
- **UT.P.** – Utile (o perdita) di esercizio *(in migliaia di euro)*

Appendice L – *Variabili utilizzate per l'apprendimento della rete*

Nella tabella in seguito riportata viene presentato l'elenco delle 19 variabili inizialmente utilizzate per l'apprendimento della rete Bayesiana.

N.	Variabile	Descrizione
1	ricavi_delle_vendite_migl_eur_ultimo_anno_disp	Ricavi di vendita espressi in migliaia di euro
2	dipendenti_ultimo_anno_disp	Numero di dipendenti
3	indice_di_liquidita_ultimo_anno_disp	Indice di liquidità
4	indice_di_indebitam_a_lungo_percent_ultimo_anno_disp	Indice di indebitamento a lungo termine %
5	indice_di_copertura_delle_immob_patrimoniale_ percent_ultimo_anno_disp	Indice di copertura delle immobilizzazioni patrimoniali %
6	rapporto_di_indebitamento_ultimo_anno_disp	Rapporto di indebitamento
7	oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp	Oneri finanziari / fatturato %
8	indice_di_indip_finanz_percent_percent_ultimo_anno_disp	Indice di indipendenza finanziaria %
9	posizione_finanziaria_netta_migl_eur_ultimo_anno_disp	Posizione finanziaria netta espressa in migliaia di euro
10	rotaz_cap_investito_volte_ultimo_anno_disp	Rotazione del capitale investito
11	ebitda_migl_eur_ultimo_anno_disp	EBITDA espresso in migliaia di euro
12	ebitda_vendite_percent_percent_ultimo_anno_disp	EBITDA/ vendite %
13	redditivita_del_capitale_proprio_roe_ percent_percent_ultimo_anno_disp	ROE
14	redditivita_delle_vendite_ros_ percent_percent_ultimo_anno_disp	ROS
15	capitale_circolante_netto_migl_eur_ultimo_anno_disp	Capitale circolante netto espresso in migliaia di euro
16	margine_di_tesoreria_migl_eur_ultimo_anno_disp	Margine di tesoreria espresso in migliaia di euro
17	totale_patrimonio_netto_migl_eur_ultimo_anno_disp	Totale del patrimonio netto espresso in migliaia di euro
18	totale_debiti_migl_eur_ultimo_anno_disp	Totale dei debiti espresso in migliaia di euro
19	utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp	Totale utile o perdita di esercizio espresso in migliaia di euro

Appendice M – *Dati per analisi inferenziale predittiva e classificazione*

In questa appendice vengono riportati tutti i dati relativi agli scenari analizzati nel capitolo, organizzati in due sezioni distinte. La prima sezione riguarda il campione delle imprese in stato giuridico attivo, quindi considerate sane; la seconda sezione raccoglie invece le imprese caratterizzate da procedure concorsuali o che sono state oggetto di fallimento.

Imprese in stato giuridico Attivo – sane

Per ciascuna delle imprese considerate negli scenari vengono presentati esclusivamente i risultati dell'analisi, ovvero la probabilità di condizione sana, la probabilità di condizione problematica, il valore restituito dall'algoritmo *predict* e la classe attribuita dal modello. Gli indici di bilancio associati a ciascun caso sono invece riportati direttamente nel capitolo dedicato all'analisi degli scenari.

impresa	Prob. Sana	Prob. Problematica	indicatore predetto	classe attribuita
A	86,07%	13,92%	0,98	sana
B	79,83%	19,87%	0,87	sana
C	59,81%	40,28%	0,61	prob

Nella pagina seguente sono riportate tutte le ulteriori imprese appartenenti al campione utilizzato per l'addestramento e la valutazione del modello, complete dei relativi valori previsti.

Il campione presenta inoltre le seguenti statistiche descrittive relative ai valori generati dall'algoritmo predict:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max
0,5305	0,6514	0,7169	0,7246	0,7817	0,9810

Questi valori descrivono la distribuzione del punteggio assegnato dal modello, mostrando come la maggior parte delle osservazioni si collochi in un intervallo che riflette una condizione tendenzialmente stabile, pur in presenza di una certa variabilità interna al campione.

n. impresa	PROB SANA	PROB PROB	predict	classe	i. liquidità	i.indebitam. lungo	Rapporto indebitam	i.copertura immo	Oneri/fatt	i.indip fin	Rotaz cap investito	Ebitda /vendite	ebitda	ROE	CCN	Tot PN	Tot debiti	Utile/ perdita
1	85,48%	14,71%	0,97	sana	0,89	0,91	3,59	2,02	30,04	27,84	0,01	70,99	77.496,86	1,66	-7.115,69	309.481,22	684.982,97	5.126,82
2	82,45%	17,22%	0,92	sana	0,19	0,55	1,64	0,50	4,51	61,16	0,46	34,49	117,88	22,88	357,19	447,78	274,92	102,46
3	79,18%	20,58%	0,86	sana	2,56	0,51	4,05	0,16	0,00	24,68	0,50	68,26	231,62	98,08	395,27	168,50	513,64	165,26
4	78,25%	21,58%	0,85	sana	1,73	0,13	1,60	0,51	0,21	62,56	0,63	57,68	397,78	57,14	233,23	608,20	361,40	347,53
5	78,16%	22,08%	0,84	sana	1,97	0,15	1,30	0,39	1,36	77,16	0,04	39,38	1.163,48	20,38	13.091,32	54.818,11	15.832,54	11.170,49
6	76,44%	23,80%	0,82	sana	3,49	0,00	1,39	0,02	0,00	71,97	1,47	65,26	204,09	90,11	148,90	153,26	59,69	138,09
7	76,21%	23,87%	0,82	sana	2,94	0,00	1,79	0,02	0,16	55,99	1,13	68,32	287,26	98,56	238,99	208,32	123,16	205,32
8	74,98%	24,96%	0,80	sana	1,61	0,39	3,61	0,88	0,94	27,68	1,09	16,57	942,80	70,91	1.365,40	1.438,03	2.949,08	1.019,67
9	74,28%	25,49%	0,79	sana	2,24	0,00	1,80	0,04	0,00	55,48	2,01	38,04	143,25	98,07	101,38	103,83	82,01	101,83
10	73,82%	25,97%	0,78	sana	1,82	0,00	2,18	0,06	0,20	45,95	0,89	61,02	202,13	78,27	163,28	170,79	198,40	133,68
11	73,56%	26,20%	0,78	sana	1,98	0,00	1,35	0,58	0,13	73,97	2,24	31,78	157,39	62,16	66,52	163,33	51,41	101,53
12	73,12%	26,95%	0,78	sana	1,67	0,79	4,49	0,00	13,32	22,27	0,17	77,96	3.164,03	11,84	1.904,23	4.821,08	13.277,35	570,61
13	72,67%	27,63%	0,77	sana	1,14	0,24	2,36	1,10	1,12	42,45	0,54	37,40	482,03	29,05	124,36	952,40	1.186,73	276,63
14	72,67%	27,33%	0,77	sana	1,20	0,00	1,93	0,05	0,00	51,92	1,33	34,51	1.025,09	62,44	593,69	1.160,74	1.023,88	724,83
15	71,38%	28,26%	0,75	sana	2,48	0,28	1,61	0,70	0,64	62,24	1,13	21,02	192,13	20,94	245,69	493,14	219,94	103,25
16	70,69%	29,56%	0,74	sana	1,15	0,00	1,48	0,04	0,04	67,57	0,80	18,87	890,73	11,25	3.244,46	3.962,35	1.756,91	445,96
17	70,54%	29,51%	0,74	sana	1,34	0,00	1,59	0,59	4,03	62,95	1,92	7,65	58.604,83	3,46	104.240,35	242.973,43	116.654,63	8.417,52
18	70,31%	29,67%	0,74	sana	5,66	0,00	1,10	0,02	6,81	91,25	1,14	42,60	163,57	44,49	133,13	298,03	28,57	132,59
19	69,83%	30,39%	0,73	sana	3,49	0,00	1,31	0,22	0,21	76,09	1,21	21,44	145,90	35,25	308,54	394,31	123,89	139,01
20	69,59%	30,57%	0,73	sana	0,74	0,06	4,89	0,00	0,61	20,45	0,65	23,58	144,28	69,06	236,22	193,44	749,15	133,59
21	68,86%	31,20%	0,72	sana	4,50	0,54	2,11	0,10	0,97	47,42	1,19	18,36	292,33	36,99	936,48	630,14	586,30	233,10
22	68,63%	31,55%	0,71	prob	3,06	0,34	2,70	0,39	1,08	37,07	1,03	24,48	324,17	45,13	455,30	471,25	332,19	212,66
23	68,48%	31,53%	0,71	prob	0,94	0,12	1,91	0,10	1,30	52,34	0,95	13,38	633,03	13,16	2.623,91	2.592,96	2.311,13	341,19

n. impresa	PROB SANA	PROB PROB	predict	classe	i. liquidità	i.indebitam. lungo	Rapporto indebitam	i.copertura immo	Oneri/fatt	i.indip fin	Rotaz cap investito	Ebitda /vendite	ebitda	ROE	CCN	Tot PN	Tot debiti	Utile/ perdita
24	68,46%	31,14%	0,72	prob	1,49	0,03	2,21	0,02	0,00	45,23	0,74	31,04	193,20	38,87	211,48	382,09	442,22	148,50
25	68,14%	31,84%	0,71	prob	1,40	0,26	2,92	0,98	0,71	34,28	1,60	13,15	325,21	42,81	368,56	527,84	847,87	225,98
26	67,82%	32,49%	0,70	prob	1,32	0,05	1,87	0,36	1,01	53,37	1,23	5,04	5.515,31	20,53	25.492,06	47.069,17	37.694,41	9.662,95
27	67,54%	32,50%	0,70	prob	2,10	0,03	1,74	0,27	0,05	57,55	1,15	19,77	8.373,25	23,81	16.868,29	21.054,35	12.163,65	5.012,98
28	67,02%	32,84%	0,70	prob	2,80	0,01	1,38	0,37	0,01	72,61	1,03	10,50	1.607,80	9,54	6.653,95	10.499,86	3.644,33	1.001,77
29	66,70%	33,64%	0,69	prob	1,30	0,05	4,37	0,34	0,63	22,89	1,35	14,37	160,87	70,76	157,45	179,95	536,74	127,34
30	66,50%	33,34%	0,69	prob	3,17	0,00	1,45	0,24	0,03	69,01	0,99	19,90	181,79	20,07	406,96	635,36	187,01	127,53
31	64,67%	35,23%	0,67	prob	2,41	0,15	2,01	0,08	0,31	49,78	1,32	13,11	163,07	22,93	530,01	470,54	407,26	107,88
32	64,05%	36,09%	0,66	prob	1,44	0,11	2,73	0,31	0,45	36,61	1,65	11,43	699,99	36,79	1.232,10	1.343,04	2.077,01	494,05
33	63,64%	36,49%	0,66	prob	3,63	0,00	1,35	0,11	0,06	73,95	1,62	13,25	256,20	17,24	837,94	865,15	212,78	149,14
34	62,97%	37,18%	0,65	prob	5,96	0,01	1,35	0,07	0,04	73,85	1,47	33,74	326,96	42,68	511,38	485,48	103,86	207,19
35	61,84%	38,04%	0,63	prob	3,06	0,00	1,48	0,11	0,00	67,51	2,68	9,92	190,33	34,03	432,91	475,65	209,93	161,85
36	60,87%	39,09%	0,62	prob	0,33	0,10	4,40	0,95	3,87	22,70	0,62	7,93	1.172,09	1,91	2.945,71	5.375,13	17.175,65	102,45
37	59,91%	40,33%	0,61	prob	0,87	0,05	3,17	0,10	0,29	31,52	1,49	7,46	1.543,24	18,63	3.885,44	4.269,61	8.603,31	795,43
38	59,90%	40,10%	0,61	prob	2,49	0,00	2,45	0,01	0,00	40,84	1,63	10,84	188,30	36,51	597,96	432,02	400,73	157,73
39	59,67%	40,25%	0,61	prob	2,21	0,00	1,83	0,56	0,11	54,61	1,80	8,83	472,38	9,02	1.057,97	1.589,73	854,43	143,43
40	59,37%	40,62%	0,61	prob	1,64	0,21	1,95	0,56	0,47	51,26	2,73	2,10	3.291,77	2,14	18.873,53	29.132,19	26.299,52	623,54
41	57,25%	42,40%	0,58	prob	1,51	0,00	2,87	0,26	0,05	34,90	1,49	6,72	425,90	14,99	1.479,75	1.472,58	2.319,41	220,69
42	57,18%	42,67%	0,58	prob	0,35	0,06	21,82	5,13	0,74	4,58	1,69	5,17	4.896,01	18,05	-9.760,68	2.523,74	44.736,82	455,62
43	54,09%	45,68%	0,55	prob	0,58	0,00	3,40	1,23	0,09	29,37	2,62	3,97	316,20	14,33	-410,53	891,09	1.478,56	127,73
44	52,75%	47,51%	0,53	prob	1,08	0,00	14,09	0,00	56,30	7,10	0,33	2,20	305,21	5,06	2.867,12	2.740,57	35.737,31	138,68

Imprese in stato giuridico non regolarmente attivo – problematiche

Anche in questa sezione, per quanto riguarda le imprese oggetto dell'analisi per scenari del capitolo, vengono presentati esclusivamente i risultati dell'analisi, ovvero la probabilità di condizione sana, la probabilità di condizione problematica, il valore restituito dall'algoritmo *predict* e la classe attribuita dal modello. Gli indici di bilancio associati a ciascun caso sono invece riportati direttamente nel capitolo dedicato all'analisi degli scenari.

impresa	Prob. Sana	Prob. Problematica	indicatore predetto	classe attribuita
D	41,94%	57,98%	0,4112	prob
E	56,46%	43,21%	0,5736	prob
F	69,30%	30,67%	0,7237	sana

Nella pagina seguente sono riportate tutte le ulteriori imprese appartenenti al campione utilizzato per l'addestramento e la valutazione del modello, complete dei relativi valori previsti.

Il campione presenta inoltre le seguenti statistiche descrittive relative ai valori generati dall'algoritmo *predict*:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max
0,1653	0,6285	0,6952	0,6771	0,7655	0,9563

Questi valori descrivono la distribuzione del punteggio assegnato dal modello, mostrando come la maggior parte delle osservazioni si collochi in un intervallo che riflette la situazione potenzialmente squilibrata.

È importante sottolineare che, all'interno di questo campione, molti valori risultano potenzialmente fuorvianti se analizzati superficialmente. In numerosi casi, infatti, le imprese oggetto di analisi hanno già avviato il processo di liquidazione e hanno proceduto alla dismissione di crediti e immobilizzazioni. Ciò genera indicatori controintuitivi e anomali, che non riflettono la normale operatività aziendale ma gli effetti delle operazioni straordinarie tipiche delle fasi di cessazione.

n. impresa	PROB SANA	PROB PROB	predict	classe	i. liquidità	i.indebitam. lungo	Rapporto indebitam	i.copertura immo	Oneri/fatt	i.indip fin	Rotaz cap investito	Ebitda /vendite	ebitda	ROE	CCN	Tot PN	Tot debiti	Utile/ perdita	Stato giuridico
1	22,48%	77,47%	0,1652	prob	0,79	0,00	-2,12	0,00	0,09	-47,17	0,00	44,84	194,36	-82,62	-147,96	-268,47	717,15	221,81	Cessata
2	41,94%	57,98%	0,4112	prob	0,16	0,04	-6,21	-4,80	0,00	-16,10	0,02	71,60	130,06	-11,85	-5.847,80	-1.047,66	7.276,69	124,18	Fallita
3	42,82%	57,26%	0,4197	prob	0,07	0,55	-6,29	-5,94	1,71	-15,90	0,00	80,88	245,57	-57,62	-1.245,04	-417,20	3.016,80	240,39	Fallita
4	51,47%	48,53%	0,5193	prob	0,96	0,00	4,28	0,17	1,25	23,35	2,62	4,85	1.682,69	21,76	2.741,46	3.094,83	9.947,94	673,45	Cessata
5	56,14%	43,76%	0,5702	prob	1,61	0,00	2,63	0,01	0,00	37,97	3,73	8,79	439,78	56,91	503,51	508,01	819,30	289,13	Cessata
6	56,46%	43,21%	0,5736	prob	1,30	0,00	2,27	0,41	0,06	44,08	3,09	7,49	395,44	27,56	465,29	750,13	911,19	206,77	Cessata
7	57,79%	42,29%	0,5881	prob	4,66	0,00	1,27	0,00	0,69	78,56	3,71	15,03	410,20	41,67	578,42	578,42	157,88	241,01	Cessata (In liqu.)
8	60,49%	39,59%	0,6197	prob	2,94	0,00	1,86	0,06	0,00	53,66	1,10	17,37	471,50	12,73	1.530,55	1.315,38	788,39	167,42	Cessata
9	61,16%	38,89%	0,6257	prob	1,74	0,13	1,92	0,08	0,72	52,20	2,08	5,49	243,39	9,43	1.099,80	1.107,13	968,40	104,41	Fallita
10	61,87%	37,89%	0,6366	prob	1,94	0,00	2,15	0,18	0,03	46,56	1,73	14,65	236,39	31,56	456,01	427,98	385,60	135,09	Cessata
11	62,27%	37,85%	0,6401	prob	1,59	0,00	2,78	0,05	0,11	35,97	1,14	14,36	546,55	32,54	1.299,28	1.185,08	1.937,31	385,59	In liqu.
12	62,36%	37,69%	0,6374	prob	1,50	0,02	2,60	0,08	0,10	38,53	0,99	11,85	430,07	20,29	1.358,19	1.414,51	2.256,78	287,01	In liqu.
13	65,07%	35,19%	0,6721	prob	2,75	0,15	1,48	0,04	0,08	67,70	2,88	11,67	493,60	35,48	964,61	993,16	469,58	352,41	Cessata (per fall.)
14	65,10%	34,80%	0,6739	prob	0,73	0,10	6,90	1,55	1,41	14,49	1,18	13,40	306,54	54,18	-99,56	259,49	1.454,98	140,58	Fallita
15	65,81%	34,21%	0,6816	prob	1,44	0,13	2,13	0,58	0,39	46,91	3,17	15,81	427,62	61,23	203,57	378,65	398,57	231,84	Cessata (per fall.)
16	66,76%	33,38%	0,6904	prob	0,68	0,00	10,41	0,01	0,05	9,61	1,32	7,47	237,40	89,98	-646,31	217,48	2.045,68	195,69	Cessata
17	66,92%	32,84%	0,6969	prob	1,04	0,00	23,82	0,08	0,00	4,20	0,74	7,92	983,48	92,92	643,89	704,86	16.084,15	654,99	Cessata (In liqu.)
18	67,12%	33,33%	0,6935	prob	1,10	0,04	2,91	0,37	34,45	34,38	0,01	-176,91	-80,06	21,42	223,85	1.337,13	2.423,81	286,46	Cessata
19	67,97%	32,40%	0,7044	prob	2,39	0,33	2,06	0,30	0,00	48,43	4,43	10,19	979,01	91,38	1.053,62	1.051,24	1.103,36	960,61	Fallita
20	68,43%	31,60%	0,7126	prob	2,20	0,00	1,83	0,00	0,00	54,52	0,25	73,40	316,11	23,85	943,06	942,93	786,74	224,87	In liqu.
21	68,46%	31,80%	0,7130	prob	0,59	0,39	3,48	0,09	0,31	28,76	0,87	8,90	468,94	16,24	-530,01	1.735,62	4.131,75	281,91	Fallita

n. impresa	PROB SANA	PROB PROB	predict	classe	i. liquidità	i.indebitam. lungo	Rapporto indebitam	i.copertura immo	Oneri/fatt	i.indip fin	Rotaz cap investito	Ebitda /vendite	ebitda	ROE	CCN	Tot PN	Tot debiti	Utile/ perdita	Stato giuridico
22	68,80%	31,10%	0,7203	sana	0,79	0,00	2,52	0,48	0,00	39,74	2,67	21,48	583,34	83,72	211,01	404,01	612,49	338,25	Cessata
23	69,30%	30,67%	0,7237	sana	0,06	0,00	4,24	0,02	1,27	23,58	0,78	31,78	334,36	60,89	223,92	309,99	988,03	188,75	Fallita
24	70,12%	30,24%	0,7335	sana	1,04	0,12	1,93	0,71	0,32	51,78	0,82	23,96	295,76	18,24	229,14	782,71	720,52	142,74	Fallita
25	70,27%	29,57%	0,7391	sana	2,89	0,03	1,76	0,00	1,24	56,66	0,65	34,59	378,97	48,09	745,49	627,52	368,26	301,79	Cessata (In liqu.)
26	73,23%	26,90%	0,7743	sana	4,73	0,95	5,64	4,44	6,13	17,73	0,18	38,37	350,75	26,49	781,57	892,51	4.125,73	236,43	Cessata
27	73,29%	26,64%	0,7752	sana	0,59	0,16	5,29	2,75	1,14	18,89	1,33	20,28	231,47	82,17	-212,83	147,14	622,60	120,91	In liqu.
28	73,55%	26,16%	0,7807	sana	2,61	0,00	1,46	0,10	0,05	68,30	0,72	81,67	310,74	48,07	310,22	359,96	167,06	173,02	Cessata
29	75,19%	24,88%	0,8018	sana	2,98	0,00	1,60	0,05	0,00	62,50	1,08	44,02	692,15	80,56	627,36	606,21	309,20	488,40	Cessata (In liqu.)
30	75,36%	24,49%	0,8052	sana	0,61	0,00	14,43	6,23	11,12	6,93	0,00	75,77	224,57	85,29	-993,34	198,27	2.552,25	169,11	In liqu.
31	75,73%	24,31%	0,8099	sana	1,02	0,60	2,32	0,01	7,59	43,07	0,10	31,26	241,16	3,98	25,78	2.786,21	3.345,18	110,89	Cessata
32	75,80%	24,38%	0,8082	sana	6,59	0,00	1,14	0,09	0,03	88,09	0,11	54,24	79,55	71,41	120,43	159,44	21,55	113,85	Cessata (In liqu.)
33	84,31%	15,56%	0,9489	sana	0,09	0,08	11,11	7,13	0,00	9,00	0,01	-24,51	-34,83	127,17	-8.160,76	1.742,22	16.428,82	2.215,62	Fallita
34	84,59%	15,36%	0,9563	sana	0,86	0,00	2,30	0,48	0,20	43,48	0,06	88,00	134,65	126,47	114,93	220,54	286,64	278,91	In liqu.

Analisi inferenziale su scala temporale

La tabella seguente riporta le probabilità che ciascuna impresa del campione presenti una condizione di continuità aziendale in ognuno degli anni analizzati, così da permettere l'osservazione dell'evoluzione temporale delle dinamiche di salute aziendale.

La colonna "PROB. Salute" indica la percentuale calcolata dal modello, ovvero la probabilità che l'impresa risulti sana sulla base degli indici di bilancio riferiti all'esercizio considerato.

La colonna "Previsione" riporta invece i valori stimati tramite l'algoritmo *predict*, accompagnati dalla classe assegnata dal modello Bayesiano.

L'ultima colonna, "Stato giuridico", specifica la situazione attuale dell'impresa nel caso in cui risulti cessata.

Le imprese oggetto di analisi approfondita nel capitolo sono evidenziate con sfondo grigio.

La tabella riassuntiva è riportata nella pagina seguente.

ID impresa	Anno -3			Anno -2			Anno -1			Ultimo anno			stato giuridico
	PROB. Salute	PREVISIONE	CLASSE	PROB. Salute	PREVISIONE	CLASSE	PROB. Salute	PREVISIONE	CLASSE	PROB. Salute	PREVISIONE	CLASSE	
1	58,52%	0,60	PROB	54,82%	0,55	PROB	41,40%	0,40	PROB	56,89%	0,58	PROB	Cessata
2	55,32%	0,56	PROB	42,76%	0,42	PROB	43,83%	0,43	PROB	44,42%	0,44	PROB	Cessata
3	47,26%	0,47	PROB	63,33%	0,65	PROB	64,00%	0,66	PROB	31,30%	0,28	PROB	Cessata
4	52,27%	0,53	PROB	44,51%	0,44	PROB	45,16%	0,45	PROB	25,00%	0,20	PROB	Cessata (In liquidazione)
5	62,03%	0,64	PROB	54,14%	0,55	PROB	50,21%	0,51	PROB	63,09%	0,65	PROB	Cessata (In liquidazione)
6	51,83%	0,52	PROB	43,67%	0,43	PROB	48,70%	0,48	PROB	47,91%	0,47	PROB	Cessata (In liquidazione)
7	60,55%	0,62	PROB	66,59%	0,69	PROB	57,67%	0,59	PROB	60,35%	0,62	PROB	Cessata (In liquidazione)
8	75,55%	0,81	SANA	74,68%	0,80	SANA	73,23%	0,78	SANA	71,59%	0,75	SANA	Fallita
9	42,06%	0,41	PROB	71,39%	0,75	SANA	73,71%	0,78	SANA	66,91%	0,70	PROB	Cessata
10	64,73%	0,67	PROB	65,78%	0,68	PROB	60,29%	0,62	PROB	72,80%	0,77	SANA	Cessata
11	64,50%	0,67	PROB	67,13%	0,70	PROB	72,53%	0,77	SANA	77,46%	0,84	SANA	Cessata (In liquidazione)
12	64,60%	0,67	PROB	65,97%	0,68	PROB	74,06%	0,79	SANA	68,25%	0,71	PROB	Fallita
13	75,38%	0,81	SANA	62,53%	0,64	PROB	57,48%	0,58	PROB	58,66%	0,60	PROB	In liquidazione
14	54,48%	0,55	PROB	52,94%	0,53	PROB	64,80%	0,67	PROB	64,58%	0,66	PROB	Cessata (Chiusura per fallimento)
15	61,12%	0,63	PROB	57,25%	0,58	PROB	59,56%	0,61	PROB	53,99%	0,55	PROB	In liquidazione
16	70,12%	0,73	SANA	70,21%	0,74	SANA	70,31%	0,74	SANA	70,46%	0,74	SANA	In liquidazione
17	54,88%	0,56	PROB	55,44%	0,56	PROB	41,86%	0,41	PROB	61,82%	0,64	PROB	Cessata
18	76,86%	0,83	SANA	69,87%	0,73	SANA	57,57%	0,58	PROB	44,41%	0,44	PROB	In liquidazione
19	52,80%	0,53	PROB	57,33%	0,58	PROB	55,14%	0,56	PROB	37,88%	0,36	PROB	Cessata
20	57,17%	0,58	PROB	57,08%	0,58	PROB	60,06%	0,61	PROB	64,80%	0,67	PROB	Cessata

21	54,13%	0,55	PROB	50,67%	0,51	PROB	56,23%	0,57	PROB	37,57%	0,36	PROB	Cessata
22	66,38%	0,69	PROB	57,72%	0,59	PROB	59,67%	0,61	PROB	59,82%	0,61	PROB	Cessata
23	55,74%	0,56	PROB	56,38%	0,57	PROB	60,37%	0,62	PROB	52,05%	0,52	PROB	Cessata
24	60,14%	0,62	PROB	63,60%	0,66	PROB	63,83%	0,66	PROB	68,50%	0,72	PROB	Cessata (In liquidazione)
25	52,37%	0,52	PROB	58,08%	0,59	PROB	56,90%	0,57	PROB	61,15%	0,62	PROB	In liquidazione

Indicatori di bilancio delle imprese analizzate nello scenario evolutivo

Indicatore di bilancio	ID IMPRESA				
	4	11	17	19	21
"indice_di_liquidita_ultimo_anno_disp"	1,23	0,36	0,89	0,91	1,50
"indice_di_liquidita_anno_1"	0,64	0,83	0,84	0,73	0,56
"indice_di_liquidita_anno_2"	0,57	1,04	0,88	0,41	0,85
"indice_di_liquidita_anno_3"	0,06	0,91	0,74	0,33	0,83
"indice_di_indebitam_a_lungo_percent_ultimo_anno_disp"	0,00	0,00	0,00	0,12	0,00
"indice_di_indebitam_a_lungo_percent_anno_1"	0,00	0,58	0,00	0,08	0,00
"indice_di_indebitam_a_lungo_percent_anno_2"	0,00	0,34	0,00	0,14	0,00
"indice_di_indebitam_a_lungo_percent_anno_3"	0,05	0,42	0,00	0,00	0,00
"rapporto_di_indebitamento_ultimo_anno_disp"	7,01	1,02	4,77	26,38	2,75
"rapporto_di_indebitamento_anno_1"	9,40	3,58	7,15	11,81	3,90
"rapporto_di_indebitamento_anno_2"	7,40	2,96	5,38	11,16	19,88
"rapporto_di_indebitamento_anno_3"	11,23	3,65	6,89	9,91	14,31
"indice_di_copertura_delle_immob_patrimoniale_ percent_ultimo_anno_disp"	0,83	0,00	0,18	2,25	0,00
"indice_di_copertura_delle_immob_patrimoniale_ percent_anno_1"	1,06	2,03	1,07	1,35	0,49
"indice_di_copertura_delle_immob_patrimoniale_ percent_anno_2"	1,34	0,97	0,81	2,70	1,37
"indice_di_copertura_delle_immob_patrimoniale_ percent_anno_3"	1,72	1,01	1,37	5,41	1,52
"oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp"	5,18	3,12	0,53	3,79	0,53
"oneri_finanz_su_fatt_percent_percent_anno_1"	1,66	2,16	0,55	3,21	0,31
"oneri_finanz_su_fatt_percent_percent_anno_2"	1,69	2,53	0,64	3,59	0,24
"oneri_finanz_su_fatt_percent_percent_anno_3"	1,57	1,87	0,06	3,39	0,21
"indice_di_indip_finanz_percent_percent_ultimo_anno_disp"	14,27	97,64	20,98	3,79	36,33
"indice_di_indip_finanz_percent_percent_anno_1"	10,64	27,95	13,98	8,47	25,61
"indice_di_indip_finanz_percent_percent_anno_2"	13,52	33,82	18,58	8,96	5,03
"indice_di_indip_finanz_percent_percent_anno_3"	8,90	27,42	14,51	10,09	6,99

"rotaz_cap_investito_volte_ultimo_anno_disp"	0,70	1,01	0,58	0,89	1,26
"rotaz_cap_investito_volte_anno_1"	1,98	0,73	1,77	1,17	1,41
"rotaz_cap_investito_volte_anno_2"	3,06	1,03	1,93	1,31	0,41
"rotaz_cap_investito_volte_anno_3"	1,87	1,03	1,32	2,31	0,80
"ebitda_migl_eur_ultimo_anno_disp"	-7.499,00	96.088,00	39.693,00	6.592,00	-17.742,00
"ebitda_migl_eur_anno_1"	14.656,00	479,86	0,04	16.148,00	11.013,00
"ebitda_migl_eur_anno_2"	23.121,00	375.691,00	15.927,00	-1.001,00	12.036,00
"ebitda_migl_eur_anno_3"	39.493,00	350.615,00	7.505,00	15.645,00	29.763,00
"ebitda_vendite_percent_percent_ultimo_anno_disp"	-12,13	4,20	58,40	3,33	-21,55
"ebitda_vendite_percent_percent_anno_1"	5,20	8,59	0,02	7,53	4,98
"ebitda_vendite_percent_percent_anno_2"	5,99	6,83	8,98	-0,54	3,95
"ebitda_vendite_percent_percent_anno_3"	11,15	6,30	6,97	5,96	6,31
"redditivita_del_capitale_proprio_roe_percent_ percent_ultimo_anno_disp"	-149,45	-0,07	31,90	-78,61	-93,62
"redditivita_del_capitale_proprio_roe_percent_percent_anno_ 1"	-12,62	12,80	-37,36	14,38	6,98
"redditivita_del_capitale_proprio_roe_percent_percent_anno_ 2"	1,23	12,58	30,14	10,23	-10,26
"redditivita_del_capitale_proprio_roe_percent_percent_anno_ 3"	7,74	-1,52	15,62	8,94	11,90
"capitale_circolante_netto_migl_eur_ultimo_anno_disp"	13.505,00	-27.442,00	-9.094,00	23.495,00	18.268,00
"capitale_circolante_netto_migl_eur_anno_1"	16.004,00	533.898,00	-10.918,00	14,35	17.846,00
"capitale_circolante_netto_migl_eur_anno_2"	25.345,00	914.659,00	-7.483,00	0,32	-7.627,00
"capitale_circolante_netto_migl_eur_anno_3"	17.172,00	973.938,00	-16.374,00	-43,28	-20.711,00
"totale_patrimonio_netto_migl_eur_ultimo_anno_disp"	12.617,00	1.936.965,00	22.785,00	8.364,00	20.768,00
"totale_patrimonio_netto_migl_eur_anno_1"	15.149,00	1.938.302,00	12.401,00	14,94	40.211,00
"totale_patrimonio_netto_migl_eur_anno_2"	17.046,00	1.690.146,00	17.033,00	12,79	37.406,00
"totale_patrimonio_netto_migl_eur_anno_3"	16.837,00	1.477.566,00	11.851,00	11.482,00	41.244,00
"totale_debiti_migl_eur_ultimo_anno_disp"	59.253,00	43.061,00	85.815,00	199.328,00	36,40

"totale_debiti_migl_eur_anno_1"	107.665,00	4.526.675,00	76,28	151.715,00	116,79
"totale_debiti_migl_eur_anno_2"	72.848,00	2.849.550,00	74.656,00	122.961,00	697.536,00
"totale_debiti_migl_eur_anno_3"	124.885,00	3.259.739,00	69.825,00	93.919,00	544.087,00
"utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"	-18.856,00	-1.337,00	7.269,00	-6.575,00	-19.443,00
"utile_perdita_di_esercizio_migl_eur_anno_1"	-1.912,00	248.158,00	-4.633,00	2.149,00	2.805,00
"utile_perdita_di_esercizio_migl_eur_anno_2"	0,21	212.578,00	5.134,00	1.308,00	-3.838,00
"utile_perdita_di_esercizio_migl_eur_anno_3"	1.304,00	-22.434,00	1.851,00	1.027,00	4.908,00

Appendice N – Codice R

Di seguito si riporta l'estratto della parte di modellazione dello script utilizzato per l'implementazione computazionale del modello descritto nel corso di questo elaborato.

Il codice completo è disponibile su richiesta.

```
##################-----STIMARE IL MODELLO----------############################

#------TENTATIVO CON NODO CONTINUTIA COME fIGLIO E TUTE V COME CONTINUE
fit.mod.continue <- bn.fit(
  dag.BL,
  data = db.c.nodo.figlio,
  method = "mle-g"
)

prev.db.continuo <- predict(
  object = fit.mod.continue,
  node = "continuita",
  data = db.c.nodo.figlio
)
fit.mod.continue
prev.db.continuo

prev.db.continuo <- ifelse(prev.db.continuo >= 0.5, "sana", "problematica")

etich.nuove.continuo   <-   ifelse(db.c.nodo.figlio$continuita   ==   1,   "sana",
"problematica")

mat.conf.db.continuo <- confusionMatrix(
  factor(prev.db.continuo, levels = c("problematica","sana")),
  factor(etich.nuove.continuo, levels = c("problematica","sana"))
)
#e qui matrice confusione ok

#-----TENATIVO CON NODO GENITORE CONTINUITA COME FATTORE - conditional gaussian
db.cond.gauss <- as.data.frame(db.unito.ua)
db.cond.gauss$continuita <- factor(db.cond.gauss$continuita,
                                   levels = c("sana","problematica"))
str(db.cond.gauss)
str(db.unito.ua)
v_ua <- c(
```

```
  "indice_di_liquidita_ultimo_anno_disp",
  "indice_di_indebitam_a_lungo_percent_ultimo_anno_disp",
  "indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp",
  "rapporto_di_indebitamento_ultimo_anno_disp",
  "oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp",
  "indice_di_indip_finanz_percent_percent_ultimo_anno_disp",
  "rotaz_cap_investito_volte_ultimo_anno_disp",
  "ebitda_migl_eur_ultimo_anno_disp",
  "ebitda_vendite_percent_percent_ultimo_anno_disp",
  "redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp",
  "capitale_circolante_netto_migl_eur_ultimo_anno_disp",
  "totale_patrimonio_netto_migl_eur_ultimo_anno_disp",
  "totale_debiti_migl_eur_ultimo_anno_disp",
  "utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"
)

db.cond.gauss[v_ua] <- lapply(db.cond.gauss[v_ua],as.numeric)

sapply(db.cond.gauss,class)

fit.mod.cond.gauss <- bn.fit(dag.cont.genitore, data = db.cond.gauss, method = "mle-
cg")

#db.prev <- db.cond.gauss
#db.prev$continuita <- NULL

set.seed(123)
prev.db.cond.gauss <- predict(
  object = fit.mod.cond.gauss,
  node = "continuita",
  data = db.cond.gauss,
  method = "bayes-lw",
  n=1000
)

conf.CD      <-      confusionMatrix(factor(prev.db.cond.gauss,      levels      =
c("problematica","sana")),
                factor(db.cond.gauss$continuita, levels = c("problematica","sana")))

parents(dag.cont.genitore, "continuita")
children(dag.cont.genitore, "continuita")

#--------TAN CON DISCRETIZZAZIONE
library(bnclassify)
library(caret)
```

```
db.tan <- as.data.frame(db.unito.ua)

db.tan$continuita <- factor(db.tan$continuita,
                            levels = c("sana", "problematica"))

# vettore delle variabili predittive (tutti i nomi tranne continuita)
features <- setdiff(names(db.tan), "continuita")

#DISCRETIZZAZIONE
# trasformo ogni variabile continua in 3 classi: basso / medio / alto
for (v in features) {

  x <- db.tan[[v]]

  #quantili 0%, 33%, 66%, 100%
  qs <- quantile(x, probs = c(0, 0.33, 0.66, 1), na.rm = TRUE)
  qs <- unique(qs)

  if (length(qs) >= 3) {

    db.tan[[v]] <- cut(
      x,
      breaks = qs,
      include.lowest = TRUE,
      labels = c("basso", "medio", "alto")[1:(length(qs)-1)]
    )
  } else {
    db.tan[[v]] <- factor("medio")
  }
}

# controllo devo avere fattori
str(db.tan)

#COSTRUZIONE TAN
set.seed(123)

tan.mod <- tan_cl(
  class   = "continuita",
  dataset = db.tan)

tan.mod   # struttura della rete TAN

#FIT MODELLO
```

```
tan.fit <- lp(
  tan.mod,
  dataset=db.tan,
  smooth = 1)   # lp = local probabilities

#MATRICE DI CONFUSIONE
# previsioni della classe (sana / problematica)
pred.tan <- predict(tan.fit, db.tan)

# matrice di confusione
conf.tan <- confusionMatrix(
  factor(pred.tan,         levels = c("problematica", "sana")),
  factor(db.tan$continuita, levels = c("problematica", "sana"))
)

conf.tan

#QUINDI CHE FUNZIONANO HO

mat.conf.db.continuo  #modello  generato  con  v  continuità  come  solo  figlia,  db
totalmente con v continute
    #modello  -> fit.mod.continue
    #db       -> db.c.nodo.figlio
    #dag      -> dag.BL

conf.CD #modello generato con v continuità come GENITORE, db MISTO
#modello  -> fit.mod.cond.gauss
#db       -> db.cond.gauss
#dag      -> dag.cont.genitore

conf.tan #modello generato con TAN e variabili discretizzate secondo livelli basso
medio alto
    #modello  -> tan.fit
    #db       -> db.tan
    #dag      -> tan.mod

#TAN SEMBREREBBE MEGLIO -> però è discretizzato

lapply(db.tan, levels)

#per vedere i limiti dei livelli
for (v in features) {
  x <- db.unito.ua[[v]]        # valori originali
  cat("\n----", v, "----\n")
  print(quantile(x, probs = c(0, .33, .66, 1), na.rm = TRUE))
```

```
}

#per vedere n casi per classe
for (v in features) {
  cat("\n----", v, "----\n")
  print(table(db.tan[[v]]))
}


################################################################################
#################################-INFERENZA-#####################################

#-----------------------------
#modello  -> fit.mod.continue
#db       -> db.c.nodo.figlio
#dag      -> dag.BL
#-----------------------------

#strumetni:
#     cpquery -> inferenza probabilistica -calcola probabilità
#     predict -> classificazione          - restituisce classe piu probabile
#     grain   -> inferenza probabilistica con tanti dati

summary(fit.mod.continue$continuita$coefficients)
summary(fit.mod.continue$continuita$fitted.values)
range(fit.mod.continue$continuita$fitted.values)  #-0,25 (prob) / 0,52 / 1,93 (sana)

#caso si prova
# CASO prova: AZIENDA sana del cambione di training

caso_prova <- list(
  indice_di_liquidita_ultimo_anno_disp                                    = 1.61,
  indice_di_indebitam_a_lungo_percent_ultimo_anno_disp                    =0.45,
  indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp=0.78,
  rapporto_di_indebitamento_ultimo_anno_disp                              =5.10,
  oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp                   =0.19,
  indice_di_indip_finanz_percent_percent_ultimo_anno_disp                 =19.61,
  rotaz_cap_investito_volte_ultimo_anno_disp                              =1.87,
  ebitda_migl_eur_ultimo_anno_disp                                        =16.99,
  ebitda_vendite_percent_percent_ultimo_anno_disp                         =11.91,
  redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp=73,
  capitale_circolante_netto_migl_eur_ultimo_anno_disp                     =26.09,
  totale_patrimonio_netto_migl_eur_ultimo_anno_disp                       =15.09,
  totale_debiti_migl_eur_ultimo_anno_disp                                 =60.75,
  utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp                    =10.96
```

```
)

# Probabilità che sia "sana"
set.seed(123)
p_prob_prova <- cpquery(
  fit.mod.continue,
  event    = (continuita >0.5),
  evidence = caso_prova,
  method   = "lw",
  n        = 1e5
)

p_prob_prova  #0,70955 probabilita che l'impresa sia sana
predict(fit.mod.continue, node = "continuita", data = as.data.frame(caso_prova))
#0,7434583 valore stimato del nodo continutia -> molto vicino a 1 = sana

###--- INFERENZA CON NUOVI CASI AIDA

#prova con aida
caso_1 <- list(
  indice_di_liquidita_ultimo_anno_disp                                    = 0.58,
  indice_di_indebitam_a_lungo_percent_ultimo_anno_disp                    = 0.26,
  indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp=1.29,
  rapporto_di_indebitamento_ultimo_anno_disp                              =1.77,
  oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp                   =7.95,
  indice_di_indip_finanz_percent_percent_ultimo_anno_disp                 =56.44,
  rotaz_cap_investito_volte_ultimo_anno_disp                              =0.13,
  ebitda_migl_eur_ultimo_anno_disp                                        =18,
  ebitda_vendite_percent_percent_ultimo_anno_disp                         =28.45,
  redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp=0.60,
  capitale_circolante_netto_migl_eur_ultimo_anno_disp                     =-24,
  totale_patrimonio_netto_migl_eur_ultimo_anno_disp                       =276,
  totale_debiti_migl_eur_ultimo_anno_disp                                 =213,
  utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp                    =0.5
)

# Probabilità che sia "sana"
set.seed(123)
p_prob_1 <- cpquery(
  fit.mod.continue,
  event    = (continuita >0.5),
  evidence = caso_1,
  method   = "lw",
  n        = 1e5
)
```

```
p_prob_1  # 0,75037 probabilita che l'impresa sia sana
predict(fit.mod.continue,  node  =  "continuita",  data  =  as.data.frame(caso_1))
#0,7984982 valore stimato del nodo continutia -> molto vicino a 1 = sana

#-----------
db.inferenza.sane <- read_excel("inferenza sane.xls")
names(db.inferenza.sane)
names(db.c.nodo.figlio)

names(db.inferenza.sane) <- c(
  "...1",
  "ragione_sociale",
  "indice_di_liquidita_ultimo_anno_disp",
  "indice_di_indebitam_a_lungo_percent_ultimo_anno_disp",
  "rapporto_di_indebitamento_ultimo_anno_disp",
  "indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp",
  "oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp",
  "indice_di_indip_finanz_percent_percent_ultimo_anno_disp",
  "rotaz_cap_investito_volte_ultimo_anno_disp",
  "ebitda_vendite_percent_percent_ultimo_anno_disp",
  "ebitda_migl_eur_ultimo_anno_disp",
  "redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp",
  "capitale_circolante_netto_migl_eur_ultimo_anno_disp",
  "totale_patrimonio_netto_migl_eur_ultimo_anno_disp",
  "totale_debiti_migl_eur_ultimo_anno_disp",
  "utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"
)
db.inferenza.sane <- db.inferenza.sane[ , -1]

db.inferenza.prob <- read_excel("inferenza prob.xls")
names(db.inferenza.prob)
names(db.c.nodo.figlio)

names(db.inferenza.prob) <- c(
  "...1",
  "ragione_sociale",
  "indice_di_liquidita_ultimo_anno_disp",
  "indice_di_indebitam_a_lungo_percent_ultimo_anno_disp",
  "rapporto_di_indebitamento_ultimo_anno_disp",
  "indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp",
  "oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp",
  "indice_di_indip_finanz_percent_percent_ultimo_anno_disp",
  "rotaz_cap_investito_volte_ultimo_anno_disp",
  "ebitda_vendite_percent_percent_ultimo_anno_disp",
```

```
  "ebitda_migl_eur_ultimo_anno_disp",
  "redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp",
  "capitale_circolante_netto_migl_eur_ultimo_anno_disp",
  "totale_patrimonio_netto_migl_eur_ultimo_anno_disp",
  "totale_debiti_migl_eur_ultimo_anno_disp",
  "utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp",
  "stato"
)
db.inferenza.prob <- db.inferenza.prob[ , -1]

str(db.inferenza.sane)
str(db.inferenza.prob)
db.inferenza.sane[db.inferenza.sane == "NS"] <- NA
db.inferenza.sane[db.inferenza.sane == "ND"] <- NA

db.inferenza.prob[db.inferenza.prob == "NS"] <- NA
db.inferenza.prob[db.inferenza.prob == "ND"] <- NA

db.inferenza.sane$indice_di_liquidita_ultimo_anno_disp                              <-
as.numeric(db.inferenza.sane$indice_di_liquidita_ultimo_anno_disp)

db.inferenza.prob$indice_di_liquidita_ultimo_anno_disp                              <-
as.numeric(db.inferenza.prob$indice_di_liquidita_ultimo_anno_disp)
db.inferenza.prob$oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp             <-
as.numeric(db.inferenza.prob$oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp)
db.inferenza.prob$indice_di_indip_finanz_percent_percent_ultimo_anno_disp           <-
as.numeric(db.inferenza.prob$indice_di_indip_finanz_percent_percent_ultimo_anno_dis
p)
db.inferenza.prob$rotaz_cap_investito_volte_ultimo_anno_disp                        <-
as.numeric(db.inferenza.prob$rotaz_cap_investito_volte_ultimo_anno_disp)
db.inferenza.prob$ebitda_vendite_percent_percent_ultimo_anno_disp                   <-
as.numeric(db.inferenza.prob$ebitda_vendite_percent_percent_ultimo_anno_disp)
db.inferenza.prob$redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_
disp                                                                                <-
as.numeric(db.inferenza.prob$redditivita_del_capitale_proprio_roe_percent_percent_u
ltimo_anno_disp)
db.inferenza.prob$indice_di_indebitam_a_lungo_percent_ultimo_anno_disp              <-
as.numeric(db.inferenza.prob$indice_di_indebitam_a_lungo_percent_ultimo_anno_disp)
db.inferenza.prob$indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_
disp                                                                                <-
as.numeric(db.inferenza.prob$indice_di_copertura_delle_immob_patrimoniale_percent_u
ltimo_anno_disp)

#tolgo NA
db.inferenza.sane <- na.omit(db.inferenza.sane)
```

```
db.inferenza.prob <- na.omit(db.inferenza.prob)

#
db.inf.prob.stato <- db.inferenza.prob
db.inferenza.prob <- db.inferenza.prob[ , -15]
str(db.inf.prob.stato)
str(db.inferenza.prob)

#ottimo ora abbiamo i nostri scenari nuovi di aida
    #db.inferenza.sane
    #db.inferenza.prob

previsioni.test.sane <- predict(fit.mod.continue, node = "continuita", data =
db.inferenza.sane) #valori stimati del nodo continutia -> molto vicino a 1 = sana
summary(previsioni.test.sane)

#probabilita che le imprese (sane) analizzate siano sane
set.seed(123)
prob_sana <- sapply(1:nrow(db.inferenza.sane), function(i) {
  evid_i <- as.list(db.inferenza.sane[i, , drop = FALSE])

  cpquery(
    fit.mod.continue,
    event    = (continuita > 0.5),
    evidence = evid_i,
    method   = "lw",
    n        = 1e5
  )
})

prob_sana.PROB <- sapply(1:nrow(db.inferenza.sane), function(i) {
  evid_i <- as.list(db.inferenza.sane[i, , drop = FALSE])

  cpquery(
    fit.mod.continue,
    event    = (continuita < 0.5),
    evidence = evid_i,
    method   = "lw",
    n        = 1e5
  )
})

prob_sana
write.table(prob_sana,sep="\t")
write.table(prob_sana.PROB,sep = "\t")
```

```
summary(prob_sana)

names(db.inferenza.sane)

summary(previsioni.test.sane)
previsioni.test.sane

names(db.inferenza.prob)

previsioni.test.prob <- predict(fit.mod.continue, node = "continuita", data =
db.inferenza.prob) #valori stimati del nodo continutia -> molto vicino a 1 = sana
summary(previsioni.test.prob)

#probabilita che le imprese (prob) analizzate siano sane
set.seed(123)
prob_problem <- sapply(1:nrow(db.inferenza.prob), function(i) {
  evid_i <- as.list(db.inferenza.prob[i, , drop = FALSE])

  cpquery(
    fit.mod.continue,
    event    = (continuita > 0.5),
    evidence = evid_i,
    method   = "lw",
    n        = 1e5
  )
})

prob_problem.prob <- sapply(1:nrow(db.inferenza.prob), function(i) {
  evid_i <- as.list(db.inferenza.prob[i, , drop = FALSE])

  cpquery(
    fit.mod.continue,
    event    = (continuita < 0.5),
    evidence = evid_i,
    method   = "lw",
    n        = 1e5
  )
})

prob_problem
summary(prob_problem)

write.table(prob_problem,sep = "\t")
write.table(prob_problem.prob,sep = "\t")
write.table(previsioni.test.prob,sep = "\t")
```

```
summary(previsioni.test.prob)

write.table(db.inf.prob.stato$stato,sep = "\t")

##--- INFERENZA CON CASO EVOLUZIONE NEL TEMPO

db.inferenza.temporale <- read_excel("inferenza temporale .xls")
names(db.inferenza.temporale)
names(db_IMP.SANE_def)

names(db.inferenza.temporale) <- c(
  "...1"                                                                ,
  "Ragione sociale"                                                     ,
  "Stato giuridico"                                                     ,
  "indice_di_liquidita_ultimo_anno_disp"                                ,
  "indice_di_liquidita_anno_1"                                          ,
  "indice_di_liquidita_anno_2"                                          ,
  "indice_di_liquidita_anno_3"                                        ,
  "indice_di_indebitam_a_lungo_percent_ultimo_anno_disp"               ,
  "indice_di_indebitam_a_lungo_percent_anno_1"                          ,
  "indice_di_indebitam_a_lungo_percent_anno_2"                          ,
  "indice_di_indebitam_a_lungo_percent_anno_3"                         ,
  "rapporto_di_indebitamento_ultimo_anno_disp"                          ,
  "rapporto_di_indebitamento_anno_1"                                    ,
  "rapporto_di_indebitamento_anno_2"                                    ,
  "rapporto_di_indebitamento_anno_3"                                  ,
  "indice_di_copertura_delle_immob_patrimoniale_percent_ultimo_anno_disp",
  "indice_di_copertura_delle_immob_patrimoniale_percent_anno_1"         ,
  "indice_di_copertura_delle_immob_patrimoniale_percent_anno_2"         ,
  "indice_di_copertura_delle_immob_patrimoniale_percent_anno_3"       ,
  "oneri_finanz_su_fatt_percent_percent_ultimo_anno_disp"              ,
  "oneri_finanz_su_fatt_percent_percent_anno_1"                         ,
  "oneri_finanz_su_fatt_percent_percent_anno_2"                         ,
  "oneri_finanz_su_fatt_percent_percent_anno_3"                          ,
  "indice_di_indip_finanz_percent_percent_ultimo_anno_disp"             ,
  "indice_di_indip_finanz_percent_percent_anno_1"                       ,
  "indice_di_indip_finanz_percent_percent_anno_2"                       ,
  "indice_di_indip_finanz_percent_percent_anno_3"                        ,
  "rotaz_cap_investito_volte_ultimo_anno_disp"                          ,
  "rotaz_cap_investito_volte_anno_1"                                    ,
  "rotaz_cap_investito_volte_anno_2"                                    ,
  "rotaz_cap_investito_volte_anno_3"                                  ,
  "ebitda_migl_eur_ultimo_anno_disp"                                   ,
  "ebitda_migl_eur_anno_1"                                              ,
```

```
  "ebitda_migl_eur_anno_2"                                          ,
  "ebitda_migl_eur_anno_3"                                           ,
  "ebitda_vendite_percent_percent_ultimo_anno_disp"                   ,
  "ebitda_vendite_percent_percent_anno_1"                             ,
  "ebitda_vendite_percent_percent_anno_2"                             ,
  "ebitda_vendite_percent_percent_anno_3"                              ,
  "redditivita_del_capitale_proprio_roe_percent_percent_ultimo_anno_disp",
  "redditivita_del_capitale_proprio_roe_percent_percent_anno_1"       ,
  "redditivita_del_capitale_proprio_roe_percent_percent_anno_2"       ,
  "redditivita_del_capitale_proprio_roe_percent_percent_anno_3"        ,
  "capitale_circolante_netto_migl_eur_ultimo_anno_disp"               ,
  "capitale_circolante_netto_migl_eur_anno_1"                         ,
  "capitale_circolante_netto_migl_eur_anno_2"                         ,
  "capitale_circolante_netto_migl_eur_anno_3"                        ,
  "totale_patrimonio_netto_migl_eur_ultimo_anno_disp"                 ,
  "totale_patrimonio_netto_migl_eur_anno_1"                           ,
  "totale_patrimonio_netto_migl_eur_anno_2"                           ,
  "totale_patrimonio_netto_migl_eur_anno_3"                           ,
  "totale_debiti_migl_eur_ultimo_anno_disp"                           ,
  "totale_debiti_migl_eur_anno_1"                                     ,
 "totale_debiti_migl_eur_anno_2"                                      ,
 "totale_debiti_migl_eur_anno_3"                                      ,
 "utile_perdita_di_esercizio_migl_eur_ultimo_anno_disp"               ,
 "utile_perdita_di_esercizio_migl_eur_anno_1"                         ,
 "utile_perdita_di_esercizio_migl_eur_anno_2"                         ,
 "utile_perdita_di_esercizio_migl_eur_anno_3"  )

db.inferenza.temporale<- db.inferenza.temporale[ , -1]
db.stato.inf.tempo <- db.inferenza.temporale

str(db.inferenza.temporale)
db.inferenza.temporale[db.inferenza.temporale == "NS"] <- NA
    db.stato.inf.tempo[db.stato.inf.tempo == "n.s."] <- NA
db.inferenza.temporale[db.inferenza.temporale == "ND"] <- NA
    db.stato.inf.tempo[db.stato.inf.tempo == "n.d."] <- NA

db.inferenza.temporale   <-   data.frame(lapply(db.inferenza.temporale,   function(x)
as.numeric(as.character(x))))

db.inferenza.temporale <- na.omit(db.inferenza.temporale)
    db.stato.inf.tempo <- na.omit (db.stato.inf.tempo)
#-#

#creare liste v. per suddivisione di anno
```

```
vars_inferenza.ua <- grep("ultimo_anno_disp$", names(db.inferenza.temporale), value
= TRUE)
vars_inferenza.ua
vars_inf.anno1 <- gsub("ultimo_anno_disp", "anno_1", vars_inferenza.ua)
vars_inf.anno2 <- gsub("ultimo_anno_disp", "anno_2", vars_inferenza.ua)
vars_inf.anno3 <- gsub("ultimo_anno_disp", "anno_3", vars_inferenza.ua)

#creare db suddivisi per anno
df_anno1  <- db.inferenza.temporale[, vars_inf.anno1]
df_anno2  <- db.inferenza.temporale[, vars_inf.anno2]
df_anno3  <- db.inferenza.temporale[, vars_inf.anno3]
df_ultimo <- db.inferenza.temporale[, vars_inferenza.ua]

      # solo perche il dag ha altri nomi nei nodi:
      #rinomino le colonne degli anni storici con i nomi dei nodi della BN
      names(df_anno1)  <- vars_inferenza.ua
      names(df_anno2)  <- vars_inferenza.ua
      names(df_anno3)  <- vars_inferenza.ua

#funzione per il calcolo della porb che ogni imp sia sana
calcola_prob <- function(df_evidence) {
  set.seed(123)
  sapply(1:nrow(df_evidence), function(i) {
    evid_i <- as.list(df_evidence[i, , drop = FALSE])
    cpquery(
      fit.mod.continue,
      event    = (continuita > 0.5),
      evidence = evid_i,
      method   = "lw",
      n        = 1e5
    )
  })
}

#applicazione della funzione per anno
prob_anno1  <- calcola_prob(df_anno1)
prob_anno2  <- calcola_prob(df_anno2)
prob_anno3  <- calcola_prob(df_anno3)
prob_ultimo <- calcola_prob(df_ultimo)

risultati_continuita.temporale <- data.frame(
  impresa   = 1:nrow(db.inferenza.temporale),
  p_anno3   = prob_anno3,
  p_anno2   = prob_anno2,
  p_anno1   = prob_anno1,
```

```
  p_ultimo  = prob_ultimo
)
risultati_continuita.temporale
write.table(risultati_continuita.temporale,sep = "\t")

#calcolo previsioni
calcola_previsioni <- function(df_evidence) {
  set.seed(123)
  sapply(1:nrow(df_evidence), function(i) {
    predict(
      fit.mod.continue,
      node = "continuita",
      data = as.data.frame(df_evidence[i, ])   # QUI LA MAGIA: uso la riga corretta!
    )

  })
}

#applicazione della funzione per anno
previs_anno1  <- calcola_previsioni(df_anno1)
previs_anno2  <- calcola_previsioni(df_anno2)
previs_anno3  <- calcola_previsioni(df_anno3)
previs_ultimo <- calcola_previsioni(df_ultimo)

risultati_previsione.temporale <- data.frame(
  impresa   = 1:nrow(db.inferenza.temporale),
  prev_anno3   = previs_anno3,
  prev_anno2   = previs_anno2,
  prev_anno1   = previs_anno1,
  prev_ultimo  = previs_ultimo
)
risultati_previsione.temporale
write.table(risultati_previsione.temporale,sep = "\t")

risultati_continuita.temporale
#-#
imp.sp.scelte <- c(4, 11, 17, 19, 21)

df_sp.inf <- risultati_continuita.temporale %>%
  filter(impresa %in% imp.sp.scelte) %>%
  pivot_longer(
    cols = c(p_anno3, p_anno2, p_anno1, p_ultimo),
    names_to = "anno",
    values_to = "prob"
  ) %>%
```

```
  mutate(
    anno = factor(
      anno,
      levels = c("p_anno3", "p_anno2", "p_anno1", "p_ultimo"),
      labels = c("Anno -3", "Anno -2", "Anno -1", "Ultimo Anno")
    )
  )

ggplot(df_sp.inf, aes(x = anno, y = prob, group = impresa, color = factor(impresa)))
+
  geom_line(linewidth = 1.1) +
  geom_point(size = 2.5) +
  labs(
    x = "Anno",
    y = "Probabilità che l'impresa sia SANA",
    color = "ID Impresa",
    title = "Trend temporale per 5 imprese selezionate"
  ) +
  theme_minimal()

write.table(db.stato.inf.tempo,sep = "\t")
names(db.stato.inf.tempo)


imp.4 <- db.inferenza.temporale[4,]
write.table(imp.4,sep = "\t")

imp.11 <- db.inferenza.temporale[11,]
write.table(imp.11,sep = "\t")

imp.17 <- db.inferenza.temporale[17,]
write.table(imp.17,sep = "\t")

imp.19 <- db.inferenza.temporale[19,]
write.table(imp.19,sep = "\t")

imp.21 <- db.inferenza.temporale[21,]
write.table(imp.21,sep = "\t")
```

Elenco delle figure

Figura 2.1: Rappresentazione grafica dell'intersezione tra insiemi. Fonte: elaborazione grafica personale 23

Figura 2.2: Rappresentazione di un grafo aciclico diretto semplificato. Fonte: elaborazione grafica personale 27

Figura 2.3: Rappresentazione grafica di una catena causale. Fonte: elaborazione grafica personale 31

Figura 2.4: Rappresentazione grafica di connessioni divergenti. Fonte: elaborazione grafica personale 31

Figura 2.5: Rappresentazione grafica di connessioni divergenti. Fonte: elaborazione grafica personale 32

Figura 2.6: Inferenza predittiva..... 37

Figura 2.7: Inferenza diagnostica. Fonte: elaborazione grafica personale..... 38

Figura 2.8: Inferenza inter-causale. Fonte: elaborazione grafica personale 38

Figura 2.9: Relazione a catena con evidenza su nodo genitore. Fonte: elaborazione grafica personale 41

Figura 2.10: Relazione a catena con evidenza su nodo figlio. Fonte: elaborazione grafica personale 41

Figura 2.11: Relazione a catena con evidenza su nodo radice. Fonte: elaborazione grafica personale 42

Figura 2.12: Relazione a catena con evidenza su nodo foglio. Fonte: elaborazione grafica personale 42

Figura 2.13: Rappresentazione grafica di una struttura a polialbero. Fonte: elaborazione grafica personale 43

Figura 2.14: Rappresentazione rete a connessioni multiple. Fonte: elaborazione grafica personale 48

Figura 3.1: Bar Plot della distribuzione delle osservazioni per area geografica. Fonte: elaborazione personale su dati AIDA 76

Figura 3.2: Bar Plot della distribuzione delle osservazioni per codice ATECO. Fonte: elaborazione personale su dati AIDA 80

Figura 3.3: Bar Plot della distribuzione delle osservazioni per Stato giuridico. Fonte: elaborazione personale su dati AIDA 81

Figura 3.4: Box Plot della variabile “ricavi di vendita” (trasformazione log); confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA 87

Figura 3.5: Box Plot della variabile “margine di tesoreria”; confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA 88

Figura 3.6: Box Plot della variabile “ROS”; confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA 90

Figura 3.7: Box Plot della variabile "posizione finanziaria netta" confronto imprese sane e problematiche. Fonte: elaborazione personale su dati AIDA 91

Figura 3.8: Scatter Plot, andamento medio "indice di liquidità". Fonte: elaborazione personale su dati AIDA 92

Figura 3.9: Scatter Plot, andamento medio "indice di indebitamento a breve termine ". Fonte: elaborazione personale su dati AIDA 93

Figura 3.10: Scatter Plot, andamento medio "indice di indebitamento a lungo termine ". Fonte: elaborazione personale su dati AIDA 93

Figura 3.11: Scatter Plot, andamento medio "indice di copertura delle immobilizzazioni". Fonte: elaborazione personale su dati AIDA 94

Figura 3.12: Scatter Plot, andamento medio "rapporto di indebitamento". Fonte: elaborazione personale su dati AIDA..... 95

Figura 3.13: Scatter Plot, andamento medio "oneri finanziari su fatturato". Fonte: elaborazione personale su dati AIDA 95

Figura 3.14: Scatter Plot, andamento medio "indice di indipendenza finanziaria". Fonte: elaborazione personale su dati AIDA 96

Figura 3.15: Scatter Plot, andamento medio "rotazione del capitale investito". Fonte: elaborazione personale su dati AIDA 97

Figura 3.16 : Scatter Plot, andamento medio "EBITDA". Fonte: elaborazione personale su dati AIDA 97

Figura 3.17: Scatter Plot, andamento mediano "EBITDA margin". Fonte: elaborazione personale su dati AIDA 98

Figura 3.18: Scatter Plot, andamento medio "ROE". Fonte: elaborazione personale su dati AIDA 98

Figura 3.19: Scatter Plot, andamento medio "capitale circolante netto". Fonte: elaborazione personale su dati AIDA 99

Figura 3.20: Scatter Plot, andamento medio "totale attvio/passivo". Fonte: elaborazione personale su dati AIDA 99

Figura 3.21: Scatter Plot, andamento medio "patrimonio netto". Fonte: elaborazione personale su dati AIDA 100

Figura 3.22: Scatter Plot, andamento mediano "totale debito". Fonte: elaborazione personale su dati AIDA 100

Figura 3.23: Scatter Plot, andamento medio "utile/perdita d'esercizio". Fonte: elaborazione personale su dati AIDA 101

Figura 3.24: esempi grafici per livello di correlazione. Fonte: elaborazione personale attraverso software R 104

Figura 3.25: Matrice di correlazione tra indici e voci di bilancio di imprese sane. Fonte: elaborazione personale su dati AIDA 105

Figura 3.26:Matrice di correlazione tra indici e voci di bilancio di imprese che hanno riscontrato problematiche di continuità aziendale. Fonte: elaborazione personale su dati AIDA 107

Figura 4.1: Dag a 19 nodi. Fonte: Elaborazione personale su dati AIDA 117

Figura 4.2: DAG imprese sane – ultimo anno disponibile. Fonte: elaborazione personale su dati AIDA..... 123

Figura 4.3:imprese sane – anno t-1. Fonte: elaborazione personale su dati AIDA..... 123

Figura 4.4: imprese sane – anno t-2. Fonte: elaborazione personale su dati AIDA..... 124

Figura 4.5: imprese sane – anno t-3. Fonte: elaborazione personale su dati AIDA..... 124

Figura 4.6: DAG imprese problematiche – ultimo anno disp. Fonte: elaborazione personale su dati AIDA..... 128

Figura 4.7: DAG imprese problematiche – anno t-1. Fonte: elaborazione personale su dati AIDA 128

Figura 4.8:DAG imprese problematiche – anno t-3. Fonte: elaborazione personale su dati AIDA 129

Figura 4.9: DAG imprese problematiche – anno t-3. Fonte: elaborazione personale su dati AIDA..... 129

Figura 4.10: Rete Bayesiana parametrizzata delle imprese sane - ultimo anno disponibile.. 131

Figura 4.11: Rete Bayesiana parametrizzata delle imprese problematiche - ultimo anno disponibile..... 131

Figura 4.12: Rete Bayesiana previsionale. Fonte: elaborazione personale su dati AIDA 137

Figura 4.13: Scatterplot andamento delle probabilità di salute di un campione di imprese nei 3 anni antecedenti la cessazione. Fonte: elaborazione personale su dati AIDA 153

Di seguito riportata la tabella relativa ai dati utilizzati per la costruzione del grafico in Figura 0.1: *Bar Plot della distribuzione delle osservazioni per Stato giuridico.* 187

Elenco delle tabelle

Tabella 3.1: Frequenze della distribuzione delle osservazioni per area geografica. Fonte: elaborazione personale su dati Aida 75

Tabella 3.2: Suddivisione settoriale codici ATECO. Fonte: ISTAT (2009)..... 79

Tabella 3.3: Statistiche descrittive della variabile "numero di dipendenti". Fonte: elaborazione personale su dati AIDA..... 84

Tabella 3.4: Statistiche descrittive della variabile "ricavi di vendita". Fonte: elaborazione personale su dati AIDA..... 85

Tabella 3.5: Statistiche descrittive della variabile "numero di dipendenti" con trasformazione logaritmica. Fonte: elaborazione personale su dati AIDA..... 86

Tabella 3.6: Statistiche descrittive della variabile "margine di tesoreria". Fonte: elaborazione personale su dati AIDA..... 88

Tabella 3.7: Statistiche descrittive della variabile "ROS". Fonte: elaborazione personale su dati AIDA..... 89

Tabella 3.8: Statistiche descrittive della variabile "posizione finanziaria netta". Fonte: elaborazione personale su dati AIDA 91

Tabella 4.1: Struttura della Rete Bayesiana multilivello, imprese sane. Fonte: elaborazione personale su dati AIDA..... 121

Tabella 4.2: Struttura della Rete Bayesiana multilivello, imprese problematiche. Fonte: elaborazione personale su dati AIDA. 125

Bibliografia e Sitografia

- Alberici, Adalberto (1975), *Analisi dei bilanci e previsione delle insolvenze: affidabilità bancaria e informativa del mercato mobiliare*, Milano, ISEDI – Istituto Editoriale Internazionale.

- Altman, Edward I. (1968), *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, vol. 23, n. 4, pp. 589–609.

- Anderson, David R.; Sweeney, Dennis J.; Williams, Thomas A. (2014), *Statistica per le analisi economico-aziendali*, Milano, Apogeo Education – Maggioli Editore.

- Appetiti, Sandro (1984), *L'utilizzo dell'analisi discriminatoria per la previsione delle insolvenze: ipotesi e test per un'analisi dinamica*, Temi di Discussione n. 27, Roma, Banca d'Italia – Servizio Studi.

- Beaver, William H. (1966), *Financial Ratios as Predictors of Failure*, Journal of Accounting Research, vol. 4, pp. 71–111.

- Benzi, Margherita (1999), *Reti bayesiane e ragionamento causale*, Sistemi intelligenti, anno XI, n. 2, pp. 239–260.

- Breiman, Leo (2001), *Random Forests*, Machine Learning, vol. 45, n. 1, pp. 5–32.

- Cenatiempo, Serena; D'Agostini, Giulio; Vannelli, Aldo (2010), *Reti bayesiane: da modelli di conoscenza a strumenti inferenziali decisionali*, Notiziario Tecnico Telecom Italia, n. 3, pp. 16–25.

- Codice Civile, artt. 2086, 2423 e 2423-bis.

- Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC) (2005), *Crisi d'impresa: strumenti per l'individuazione di una procedura dall'allerta*, Roma.

- Cover, Thomas; Hart, Peter (1967), *Nearest Neighbor Pattern Classification*, IEEE Transactions on Information Theory, vol. 13, n. 1, pp. 21–27.

- Desai, Vikram; Bucaro, Anthony C.; Kim, Joung W.; Srivastava, Rajendra; Desai, Renu (2023), *Toward a Better Expert System for Auditor Going Concern Opinions Using Bayesian Network Inflation Factors*, International Journal of Accounting Information Systems, vol. 49, art. 100617.

- Ellsberg, Daniel (1961), *Risk, Ambiguity and the Savage Axioms*, The Quarterly Journal of Economics, vol. 75, n. 4, pp. 643–669.

- Gazzetta Ufficiale della Repubblica Italiana.

- ISA Italia n. 570, Continuità aziendale, Principi di revisione internazionale.

- ISTAT – Istituto Nazionale di Statistica (2009), *Classificazione delle attività economiche ATECO 2007: derivata dalla NACE Rev. 2* (Metodi e norme n. 40), Roma.

- ISTAT – Istituto Nazionale di Statistica (2024), Rapporto Annuale 2024 – La situazione del Paese, Capitolo 14 *Struttura e dinamica delle imprese industriali e dei servizi*, Roma.

- Kim, Jin H.; Pearl, Judea (1983), *A Computational Model for Causal and Diagnostic Reasoning in Inference Systems*, Proceedings of the 8th International Joint Conference on Artificial Intelligence.

- Koller, Daphne; Friedman, Nir (2009), *Probabilistic Graphical Models: Principles and Techniques*, Cambridge (MA), MIT Press.

- Korb, Kevin B.; Nicholson, Ann E. (2011), *Bayesian Artificial Intelligence*, 2ª ed., Boca Raton, CRC Press – Taylor & Francis Group.

- Maffei, Marco (2017), *Il principio della continuità aziendale e il controllo della sua corretta applicazione*, Torino, Giappichelli.

- Mullainathan, Sendhil; Spiess, Jann (2017), *Machine Learning: An Applied Econometric Approach*, Journal of Economic Perspectives, vol. 31, n. 2, pp. 87–106.

- Nagarajan, Radhakrishnan; Scutari, Marco; Lèbre, Sophie (2013), *Bayesian Networks in R: With Applications in Systems Biology*, New York, Springer.

- OIC 11 – Organismo Italiano di Contabilità, Finalità e postulati del bilancio d'esercizio.

- Pearl, Judea (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo (CA), Morgan Kaufmann.

- Pourret, Olivier; Naim, Patrick; Marcot, Bruce G. (2008), *Bayesian Networks: A Practical Guide to Applications*, Chichester, John Wiley & Sons.

- Quattrocchio, M. Luciano (2019), *Diritto ed economia dell'impresa*, Torino, Giappichelli.

- Savage, Leonard J. (1954), *The Foundations of Statistics*, New York, John Wiley & Sons.

- Scutari, Marco; Denis, Jean-Baptiste (2015), *Bayesian Networks: With Examples in R, Boca Raton*, CRC Press – Taylor & Francis Group.

- Shmueli, Galit (2010), *To Explain or to Predict?*, Statistical Science, vol. 25, n. 3, pp. 289–310.

- Sun, Lili; Shenoy, Prakash P. (2007), *Using Bayesian Networks for Bankruptcy Prediction: Some Methodological Issues*, European Journal of Operational Research, vol. 180, n. 2, pp. 738–753.

- Zappia, Carlo (2014), *Leonard Savage, the Ellsberg Paradox, and the Debate on Subjective Probabilities: Evidence from the Archives*, History of Political Economy, vol. 46, n. 1, pp. 121–154.