



Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale in
Economia e Gestione delle Aziende
EM6-11

Tesi di Laurea

L'approccio dei Rough-Sets applicato alla valutazione dello stato di salute aziendale

Relatore

Ch. Prof. Margherita Gerolimetto

Correlatore

Ch. Prof. Marco Tolotti

Laureando

Alessandro Mazzocco
Matricola 833518

Anno Accademico

2016 / 2017

*Un sincero grazie alla professoressa Gerolimetto
che ha reso possibile questo lavoro.*

*Una dedica particolare ai miei genitori
Paola e Fabio, che mi sostengono sempre.*

Sommario

Introduzione	5
Il fallimento dell'azienda	8
1.1 I Modelli Statistici Classici	8
1.1.1 Multiple Discriminant Analysis.....	9
1.1.2 Conditional Probability Models.....	12
1.2 AIES.....	15
1.2.1 Neural Networks	15
1.2.2 Case Based Reasoning	16
1.2.3 Rough-Set Theory	17
1.3 Modelli Teorici.....	19
1.4 Considerazioni	20
2. La Teoria dei Rough-Set.....	22
2.1 Information Table e relazione di indiscernibilità	22
2.2 Approssimazioni e Rough-set.....	24
2.3 Dipendenza e Reduction degli attributi	27
2.4 Decision Table e regole decisionali	28
3. Analisi Empirica	30
3.1 I dati.....	30
3.2 Analisi Descrittiva	34
3.3 Applicazione della Rough-Set Theory.....	39
3.3.1 Discretizzazione degli Attributi.....	39
3.3.2 Reduct e Core	41
3.3.3 Induzione delle Regole decisionali	45
3.3.4 Validazione	48
Conclusione	54
Bibliografia	57
Elenco delle Formule.....	59
Elenco delle figure e tabelle.....	61
Appendice	63

Introduzione

Il fallimento di un'azienda consiste nella sua incapacità sopraggiunta di onorare i propri debiti, che siano nei confronti degli istituti di credito o dei fornitori, ma anche nella mancanza di liquidità per onorare i salari dei dipendenti e dei debiti nei confronti dello Stato. In generale il fallimento aziendale si verifica come discontinuità nelle operazioni dell'impresa, che va ad influenzare, quasi sempre in maniera negativa, le relazioni del tessuto economico-sociale di appartenenza, composto da lavoratori, fornitori, clienti, istituzioni e la collettività (Warner, 1977). Per questo motivo negli ultimi 50 anni molti studiosi si sono concentrati nello studio del fallimento aziendale, in particolar modo della previsione di questo, ricercandone le cause e cercando di identificare dei modelli che potessero permettere quantomeno l'identificazione di un segnale di allarme. A beneficio di tali studi sono innanzitutto gli attori interni all'azienda, i manager, interessati a prevenire il fallimento e a poter attuare politiche atte a scongiurarlo; successivamente gli investitori, e le istituzioni.

Il fine di questa Tesi di Laurea è quello di validare l'efficacia di utilizzo della teoria dei Rough-Sets nella discriminazione tra gruppi di aziende fallite e attive. Tale metodologia, che verrà approfondita in seguito, comporta grandi vantaggi dal punto di vista del trattamento dei dati, essendo in grado di analizzare e suddividere in gruppi, insiemi di dati aventi informazioni condivise ma ambigue. Nel ramo di studi in questione, questa caratteristica si rivela molto importante, poiché i dati utilizzati per predire il fallimento aziendale non possono che essere desunti dai bilanci aziendali; questi valori, come noto, rappresentano da un lato l'andamento dell'azienda secondo norme di prudenza e competenza, ma dall'altro lato possono risultare fuorvianti in un'analisi statistica. Basti pensare a strategie aziendali di lungo periodo che prevedono il presentarsi di perdite di esercizio a causa di investimenti effettuati o ad altre scelte di gestione. Al fine di contribuire alla tematica con un approccio parzialmente diverso, è stato deciso di analizzare aziende italiane di piccole e

medie dimensioni fallite e non fallite, in un arco temporale compreso tra il 2011 ed il 2015, ovvero in un periodo successivo alla recente crisi finanziaria. Non sono stati rintracciati in letteratura esempi di analisi Rough-Set eseguiti su questa tipologia di dati.

Il lavoro è stato strutturato in tre Capitoli: nel primo Capitolo verrà condotta una breve analisi della letteratura al fine di evidenziare i metodi utilizzati e i risultati ottenuti. Verrà prestata particolare attenzione alle modalità di campionamento delle aziende e di scelta delle variabili da utilizzare. Nel secondo capitolo verranno presentate le basi matematiche della teoria dei Rough-Sets, in maniera tale da chiarirne il funzionamento e le modalità di applicazione. Infine nel terzo Capitolo verrà applicata la teoria su un campione di aziende estratte dal *DataSet AIDA* (Analisi Informatizzata Database Aziendali). Il campione verrà scelto facendo ricorso alle informazioni apprese dalla letteratura (esclusi gli anni di interesse come visto sopra), così come per le variabili utilizzate. Le 12 variabili scelte verranno dapprima sottoposte ad un'analisi descrittiva al fine di ottenerne una maggiore conoscenza, dopodiché verranno discretizzate al fine di garantire una maggiore normalizzazione dei dati. In seguito il numero delle variabili o attributi sarà ridotto tramite un'operazione chiamata *Reduct*, attraverso la quale viene fornita una lista di tutte le possibili combinazioni di variabili ridotte senza avere perdita di informazioni. L'ultima parte consisterà nell'analisi vera e propria dei dati appena pre-processati: attraverso gli algoritmi di ROSE2, un software sviluppato dal Professore Roman Słowiński dell'University of Technology di Poznań in Polonia. Verranno indotte delle regole decisionali in grado di stabilire, dato uno o più valori di un attributo, quale sia il gruppo di appartenenza dell'oggetto avente queste caratteristiche. Infine sarà possibile validare i risultati tramite il test di classificazione *k-fold*, rilevando come il metodo porti dei risultati di classificazione piuttosto precisi.

Il fallimento dell'azienda

La ricerca sulla previsione del fallimento aziendale si compone in letteratura di tre principali categorie di studio, suddivise in base alle metodologie utilizzate nell'analisi ed al fine perseguito (Aziz and Dar, 2006). Queste sono:

1. Modelli statistici classici;
2. Artificially Intelligent Expert System Models (AIES)
3. Modelli Teorici;

1.1 I Modelli Statistici Classici

Negli ultimi 50 anni il tema del fallimento dell'azienda ha portato molti ricercatori a ricercare le migliori metodologie statistiche che si applicassero allo studio di tale tematica. È facilmente riscontrabile dalla letteratura come la maggior parte, se non la quasi totalità degli studi fatti, siano stati applicati su di un orizzonte temporale superiore ad un anno, solitamente entro i 5 anni dal fallimento o crisi aziendali. Questi prevedono il campionamento svolto in maniera trasversale lungo un arco temporale, andando quindi ad analizzare l'azienda nelle sue fasi di declino prima dell'effettivo fallimento per coglierne gli aspetti critici. L'attenzione di queste ricerche è posta nei confronti dei sintomi del fallimento, al fine di acquistarne una maggiore consapevolezza.

Il pioniere di tali studi fu Beaver (1967a), il quale teorizzò un modello basato su un'analisi discriminante Univariata utilizzando indicatori finanziari di bilancio selezionati tramite un test di classificazione dicotomica in un gruppo di aziende fallite e non. Ad ogni indicatore o rapporto veniva assegnato un valore soglia, il quale veniva poi selezionato separatamente in base al valore o importanza datogli dall'azienda e al livello ottimale. Gli indicatori così selezionati erano poi utilizzati come variabili predittive. Tale metodo, nonostante la semplice applicabilità, aveva però lo svantaggio di assumere una relazione lineare tra lo

stato di fallimento e tutti gli indicatori utilizzati, risultando quindi di scarsa utilità e superato. Beaver comunque concluse che il miglior indice predittore nel caso di analisi Univariata era determinato dal rapporto: Flusso di cassa/Indice di Indebitamento.

Parallelamente a Beaver, Tamari, (1966) e Moses and Liao (1987) hanno sviluppato un modello denominato Indice di Rischio, il quale prevedeva l'assegnazione di un punteggio da 0 a 100 all'azienda. Questo veniva assegnato in base al valore degli indicatori di bilancio utilizzati nel modello ed evidenziava lo status più o meno sano dell'azienda, partendo da 100 come indice di ottima salute finanziaria.

1.1.1 Multiple Discriminant Analysis

Di grande interesse è la Multiple Discriminant Analysis, d'ora in avanti MDA, utilizzata per la prima volta da Altman (1968) e Beaver (1967, 1968) nell'ambito dello studio della salute aziendale. Il vantaggio della MDA è di andare oltre al limite della semplice analisi Univariata, ovvero quello di porre enfasi sulle singole variabili. Un rischio comune infatti può essere quello di giudicare come fallimentare un'azienda avente ad esempio un basso indice di profittabilità, nonostante al contempo possa avere una liquidità al di sopra della media. Risulta dunque chiaro come nello studio delle dinamiche aziendali non si possa fare affidamento ad analisi che considerano le variabili "*stand alone*", ma bensì serva passare ad una visione maggiormente allargata che prenda a riferimento più variabili allo stesso tempo.

La MDA consiste in una combinazione lineare di variabili, la quale riesce a fornire una migliore distinzione in termini di azienda fallita-non fallita. La funzione discriminante di un modello MDA lineare è la seguente:

$$(1) D_i = d_0 + d_1X_{i1} + d_2X_{i2} + \dots + d_nX_{in}$$

Dove $D_i \in \mathbb{R}$ rappresenta il valore discriminante della funzione (*discriminant score*) per l'azienda i ; X_{ij} il valore dell'attributo X_j (con $j=1, \dots, n$) per l'azienda i ; d_0 è l'intercetta e d_j è il coefficiente di discriminazione lineare per l'attributo j . D_i fornisce un'indicazione sullo stato di salute dell'azienda ed è proprio sulla base del *discriminant score* che viene effettuata la classificazione dell'azienda i nella categoria fallita o non fallita. Più precisamente, stabilito un limite ottimale di D_i , le aziende al di sotto di questo saranno considerate come fallite o prossime al fallimento, mentre quelle allo stesso livello o al di sopra in buona salute.

Altman (1968) utilizzò tale metodologia per andare oltre ai modelli dell'epoca ed essere in grado di analizzare il completo profilo di variabili dell'oggetto simultaneamente, invece che esaminarne in maniera sequenziale le caratteristiche. Formulò un modello basato su un campione di 66 aziende manifatturiere divise equamente in due gruppi: fallite (Gruppo 1) e non fallite (Gruppo 2). Il gruppo 1 racchiudeva aziende dichiarate fallite secondo le leggi dell'epoca tra il 1946 e il 1965 con un capitale medio di 6.4 Milioni di Dollari, ma con un intervallo compreso tra 0.7 Milioni e 25.9 Milioni; un gruppo non del tutto omogeneo. Per il secondo gruppo venne optata, nella scelta del campione, una stratificazione. Già usata in quello precedente per isolare le aziende manifatturiere, venne applicata anche alle dimensioni del capitale, ponendo come parametro 1-25 Milioni di Dollari con una media di 9.6. Come si può notare la media delle dimensioni delle aziende differisce nei due gruppi, ma secondo la tesi di Altman tale differenza non era particolarmente importante.

Dopo avere selezionato le aziende, vennero raccolti i dati dei bilanci dal conto economico e dallo stato patrimoniale. A causa dell'elevato numero di variabili considerabili, si crearono cinque categorie di indici: Liquidità, Profittabilità, Rapporto di Indebitamento, Solvibilità e Attività. Dalle 22 variabili risultanti di maggior utilità, ne vennero scelte 5, tenendo in considerazione la popolarità nella letteratura economica e la potenziale rilevanza con lo studio. Il risultato fu la seguente funzione discriminante:

$$(2) Z = .012X_1 + .014X_2 + .033X_3 + .006X_4 + .999X_5$$

Dove:

X_1 : Capitale Circolante/Totale Attivo. Questo indice viene trovato spesso negli studi su analisi aziendali ed è indicatore della liquidità netta relativa al capitale totale. Il capitale circolante viene definito come differenza tra attivo corrente e passività correnti; in questo modo vengono considerate sia la liquidità che le dimensioni dell'azienda. Solitamente, un'azienda che registra perdite consistenti avrà le attività correnti in flessione rispetto alle passività correnti.

X_2 : Utili trattenuti/Totale Attivo. Questa misura viene introdotta da Altman, e considera implicitamente l'età di una azienda; infatti un'azienda giovane avrà probabilmente un basso valore di questo indice dovuto al fatto che non ha ancora avuto tempo di realizzare utili su più basi annuali. Le aziende di recente formazione saranno quindi svantaggiate nel valore di questo indice; Altman affermò che nel 1965 il 50% delle aziende che dichiararono fallimento, lo fecero nei primi 5 anni di vita.

X_3 : EBIT/Totale Attività. È una misura della produttività delle attività dell'azienda al netto delle tasse e degli oneri finanziari.

X_4 : Valore di mercato del Patrimonio Netto/Valore totale dei Debiti. Questo indice mostra di che grado le attività dell'azienda possono perdere di valore prima che i debiti (considerati sia quelli di breve che medio lungo termine) le eccedano, creando problemi di insolvenza.

X_5 : Vendite/Totale Attività. Questo è l'indice di Turnover del Capitale, in grado di dimostrare la capacità di generare entrate delle attività aziendali; statisticamente non significativo applicato all'analisi Univariata, risulta di grande importanza nella MDA.

Le variabili con maggior contribuzione della discriminazione su base relativa furono in ordine X_3 , X_5 e X_4 . L'indice di Profittabilità, come previsto, è il più rilevante, poiché è difficile aspettarsi un fallimento da un'azienda che produce utili.

I risultati di questo studio furono piuttosto incoraggianti, fornendo un'accuratezza del 95% nel prevedere il fallimento di un'azienda ad un anno dall'avvenimento. Gli errori del I Tipo, ovvero aziende fallite classificate come non fallite, si sono attestati sul 6%; mentre per gli errori del II Tipo, ovvero aziende non fallite ma discriminate come tali, sul 3%.

A seguito di Altman, innumerevoli altri autori applicarono questo metodo ottenendo risultati comparabili in termini di capacità predittiva, come ad esempio Deakin (1972) e Blum (1974). Se da un lato questo metodo garantisce buoni risultati, bisogna ricordare che l'applicazione risulta piuttosto complessa, e la sua funzione è discriminatoria più che predittiva, non potendo dunque fornire alcuna stima della probabilità di fallimento.

1.1.2 Conditional Probability Models

I modelli di probabilità condizionali, a differenza della MDA, sono modelli non lineari che assumono una distribuzione normale cumulata nel caso dei modelli Probit, e di una distribuzione logistica nel caso di modelli Logit. (Maddala, 1991)

Nella letteratura il modello Logit è sicuramente il più utilizzato, e per somiglianza con il modello Probit, si decide di trattare solo il primo. Nell'analisi Logit, viene utilizzata una procedura di stima della massima probabilità non lineare per ottenere i valori dei parametri del modello così espresso:

$$(3) P_1(X_i) = \frac{1}{[1 + \exp - (b_0 + b_1X_{i1} + b_2X_{i2} + \dots + b_nX_{in})]} = 1 / [1 + \exp - (D_i)]$$

Dove $P_1(X_i)$ rappresenta la probabilità di fallimento dato il vettore di attributi X_i ; X_{ij} il valore dell'attributo j (con $j = 1, \dots, n$) per l'azienda i ; b_j il coefficiente dell'attributo j ; b_0 l'intercetta e D_i il logit dell'azienda i , ovvero la curva di regressione.

Il modello di regressione Logistica combina più caratteristiche o attributi di un'azienda in uno *score* indicante la probabilità di fallimento. La funzione Logistica implica che lo *score*

del logit P_1 abbia valori compresi tra 0 ed 1 e aumenti in D_i . Quando lo stato giuridico “fallita” è codificato come numero 0, un basso *score* del logit indicherà un’alta probabilità di fallimento, oltre che a dare un’indicazione di cattiva salute aziendale.

Ohlson (1980) applicò questo metodo su un campione di aziende ottenuto secondo le seguenti caratteristiche:

- Periodo di studio compreso tra il 1970 ed il 1976
- Le aziende dovevano essere quotate
- L’azienda doveva appartenere al ramo dell’industria.

Il campione finale così ottenuto comprendeva 105 aziende dichiarate fallite e 2.058 aziende attive. Furono composti 3 modelli per ottenere una previsione del fallimento l’anno prima dell’avvenimento, due anni prima, e uno o due anni prima.

Le variabili utilizzate nel modello furono:

1. SIZE: un indice indicante la grandezza dell’azienda calcolato come: $\log(\text{total assets}/\text{GNP price-level index})$
2. Total Liabilities / Total Assets;
3. Working Capital / Total Assets;
4. Current Liabilities / Current Assets;
5. OENEG : una variabile dicotomica con valore 1 se i debiti totali eccedono le attività totali, 0 altrimenti;
6. Net Income / Total Assets
7. Funds provided by operations / Total Liabilities
8. INTWO: una variabile dicotomica avente valore 1 se negli ultimi due anni si sono registrati utili negativi, 0 altrimenti;
9. CHIN: un indicatore che mostra le variazioni degli utili tra gli ultimi due anni considerati

Ohlson ottenne interessanti risultati sia per quanto riguarda il modello stesso, con una percentuale di correttezza maggiore raggiunta nel modello 1 pari a 96,12%, sia rispetto ad altri risultati presenti in letteratura. Innanzitutto constatò come nonostante i dati e gli indici utilizzati non fossero comparabili con quelli di Beaver (1966), i risultati erano significativamente simili. Inoltre, al contrario di quanto sostenuto da Altman (1968), rilevò come la grandezza dell'azienda ricopra un ruolo importante nello studiarne il fallimento. Affermò infatti che i 4 fattori di maggior importanza derivati dai bilanci aziendali per stimare la probabilità di fallimento erano:

- SIZE (la grandezza);
- La struttura finanziaria, espressa da $\text{Total Liabilities} / \text{Total Assets}$;
- Una misura o una combinazione di misure di performance come $\text{Net Income} / \text{Total Assets}$
- Una misura della liquidità corrente come $\text{Working Capital} / \text{Total Assets}$

1.2 AIES

Gli *Artificially intelligent expert system models* rappresentano un ramo di studi altamente dipendente dalla tecnologia computazionale. Questi modelli cercano di porsi come alternativa all'utilizzo di modelli statistici classici, anche se nonostante tutto, per forza di cose, racchiudono al loro interno un'eredità statistica importante. Verranno ora descritti brevemente alcuni tra i più importanti modelli sopra citati.

1.2.1 Neural Networks

Il metodo dei Neural Networks (Reti Neurali) consiste nella classificazione di oggetti (in questo caso le aziende) tramite un processo emulativo del funzionamento del cervello umano. I "Neuroni" rappresentano dei nodi con interconnessioni di diverso peso organizzati su più strati; più precisamente ogni nodo della rete è un elemento in grado di processare una serie di segnali in entrata (in questo caso le informazioni riguardo alle aziende come gli indici di bilancio) e convertirli in un unico segnale in uscita. I risultati possibili sono due:

1. L'accettazione della decisione di classificazione ottenuta;
2. La ritrasmissione del segnale agli altri nodi, tra cui è compreso il nodo stesso

Questo processo iterativo continua fino a che non viene raggiunta una decisione di classificazione in base ai criteri richiesti, data come "con probabilità x , l'azienda fallirà".

Boritz and Kennedy (1995) applicarono questo metodo con lo scopo di verificarne l'effettiva possibilità di applicazione alla predizione del fallimento aziendale. Utilizzarono un campione di 171 società fallite e 6.153 società attive in un arco temporale compreso tra il 1971 ed il 1984.

La decisione su quali variabili utilizzare ricadde nella scelta delle medesime utilizzate da Altman (1968) e Ohlson (1980), al fine di poter presentare un paragone tra i risultati dei modelli.

Boritz e Kennedy ottennero risultati non del tutto soddisfacenti, in linea se non addirittura peggiori con gli altri metodi classici utilizzati, constatando dunque come l'utilizzo delle Reti Neurali non rappresentasse alcun miglioramento rispetto ad altri metodi più comuni, i quali inoltre detenevano il pregio di essere facilmente compresi e applicati tramite l'utilizzo di programmi statistici già sviluppati.

1.2.2 Case Based Reasoning

Il modello del Case Based Reasoning, ovvero "ragionamento basato sui casi", è basato su algoritmi che permettono la soluzione di un problema di classificazione tramite il ricorso alla memoria di altre classificazioni precedentemente operate. Questo è in breve il funzionamento:

1. Identificazione di un nuovo problema;
2. Recupero di informazioni riguardo a casi precedentemente risolti;
3. Adattamento delle informazioni recuperate al nuovo caso;
4. Valutazione della soluzione suggerita

Chiaramente, alla fine del processo qui sinteticamente descritto, le nuove informazioni ottenute vengono a loro volta memorizzate per un uso futuro.

Li and Sun (2009) applicarono questo metodo su un campione di 83 aziende cinesi quotate alla Borsa di Shanghai. Utilizzarono le informazioni di 30 variabili corrispondenti ad indici di bilancio relativi agli anni tra il 2000 e il 2005 e raffiguranti:

- La profittabilità
- Le attività
- I debiti
- La struttura finanziaria

Conclusero affermando che il modello ottenuto rappresenta un valido strumento atto a predire la scarsa salute aziendale delle società quotate cinesi, in quanto stabile e in grado di fornire una predizione piuttosto accurata; in particolar modo riuscirono ad ottenere una precisione media del 89,25% di classificazione sul campione analizzato.

1.2.3 Rough-Set Theory

La teoria dei Rough-Sets nasce con lo scopo di classificare oggetti utilizzando informazioni imprecise o imparziali. Da queste vengono indotte tramite algoritmi delle regole in grado di classificare un gruppo di oggetti in base a delle condizioni. Per le basi teoriche del metodo si rimanda al Cap. 2

Si riscontra un interessante utilizzo di questa metodologia da parte di Dimitras et al. (1998), i quali cercarono di predire il fallimento di un campione di 80 aziende greche. Questi raccolsero un cospicuo numero di aziende fallite in Grecia tra il 1986 ed il 1990 e ponendo criteri di selezione quali un minimo di 5 anni di attività e la disponibilità dei dati, ottennero un campione di 40 oggetti rispondenti alle caratteristiche ricercate; vennero raccolti i dati di bilancio risalenti a fino 5 anni prima del fallimento. Successivamente, ad ogni azienda fallita ne venne associata una attiva così da raggiungere un *set* totale di 80 oggetti da analizzare, definiti *learning sample*. Per le informazioni riguardo agli attributi utilizzati e alle norme di discretizzazione applicate agli attributi si rimanda al Cap.3

I risultati ottenuti furono soddisfacenti, riuscendo a classificare con una precisione del 100% le aziende fallite dalle non fallite nell' anno prima del fallimento, e con una precisione del 86,3% 2 anni prima. La Rough-Set Theory si dimostrò superiore all'analisi MDA ed all'analisi Logit; Dimitras et al. utilizzarono entrambe le metodologie statistiche con gli stessi dati ottenendo risultati significativamente inferiori.

Entrambi i modelli non ottennero una precisione superiore al 90% nell'anno -1; in particolare la MDA ottenne una precisione dell'85%, mentre la Logit ottenne una precisione

del 90%, dimostrando dunque la superiorità, in riferimento ai dati trattati, della Rough-Set Theory come strumento discriminante.

1.3 Modelli Teorici

In ultima analisi si citano per completezza i Modelli Teorici. A differenza degli altri 2 filoni di studi trattati, i modelli teorici sono basati su uno studio maggiormente qualitativo delle cause del fallimento, atto ad alimentare la discussione accademica in ambito economico-finanziario.

Tra questi si cita la *Gambler's ruin theory* (Scott 1980) in cui l'azienda viene vista come uno "scommettitore". Dato un certo ammontare di tempo, veniva calcolata una probabilità di registrare perdite in maniera continuativa, fino a portare all'azzeramento del Patrimonio e al successivo fallimento.

Un altro esempio è dato da (Laitinen and Laitinen 1998) con il *Cash management theory*: tale modello si focalizzava sulla liquidità nel breve periodo, andando a ricercare le cause di un possibile sbilanciamento tra entrate ed uscite. Nel caso di uno sbilanciamento a favore delle uscite prolungato nel tempo, chiaramente il rischio di fallimento aumentava.

1.4 Considerazioni

A seguito di questa breve panoramica sui principali metodi utilizzati in letteratura, si traggono le seguenti conclusioni:

- L'analisi MDA e Logit rappresentano la grande maggioranza dei metodi utilizzati poiché affiancati spesso ad altre tipologie di analisi per trarne poi un confronto; questo non pregiudica il fatto che rappresentino due buone metodologie per la ricerca sul fallimento aziendale, infatti la letteratura mostra come vi siano metodi in grado di ottenere una maggiore precisione, come ad esempio la teoria dei Rough-sets.
- Le variabili utilizzate sono costituite prevalentemente da indicatori desunti dai dati dei bilanci delle aziende riportanti informazioni su 4 macro-categorie quali: *a)* la capacità di generare profitto; *b)* la solvibilità e la disponibilità liquida; *c)* la struttura finanziaria dell'azienda; *d)* la presenza di debiti entro e oltre l'esercizio.
- La tipologia dei campioni utilizzati è principalmente di due tipi: *a)* campioni composti da un numero pari di aziende fallite e aziende attive; *b)* campioni composti da un ristretto numero di aziende fallite contrapposte a un vasto numero di aziende attive. Si riscontra inoltre come non vi sia accordo in letteratura riguardo alla scelta dei criteri da utilizzare nella selezione delle aziende da analizzare. Alcuni autori affermano l'importanza di campioni omogenei con società facenti parte della stessa industria ed aventi simili dimensioni, mentre altri ritengono non necessario questo tipo di selezione.
- L'approccio dei Rough-Sets si rivela molto interessante per i risultati ottenuti, in particolar modo grazie alla distintiva capacità di analizzare i dati ambigui, quali sono proprio i dati di bilancio.

2. La Teoria dei Rough-Set

La teoria dei Rough-set fu teorizzata da Z. Pawlak (1982) per la prima volta all'inizio degli anni ottanta, e successivamente sviluppata dallo stesso nel 1991 grazie alla crescente capacità di calcolo delle macchine. Questa teoria rappresenta uno strumento matematico col fine di ripartire o suddividere degli oggetti caratterizzati da informazioni incomplete o imprecise. Essa si basa sul fatto che ogni oggetto studiato è caratterizzato da una serie di informazioni che lo descrivono, espresse sotto forma di attributi. Tali attributi, o informazioni, qualora fossero uguali tra più oggetti, instaurano una relazione di indiscernibilità, ovvero rendono gli oggetti indistinguibili e raggruppati in base alle informazioni che li caratterizzano.

Di seguito verranno fornite le basi teoriche di tale metodo e i passaggi per la sua corretta applicazione.

2.1 Information Table e relazione di indiscernibilità

Approcciandosi ad un'analisi dei dati con il metodo dei Rough-set, il punto di partenza è rappresentato dalla cosiddetta *information table*. Si tratta di una tabella di dati, in cui ad ogni riga corrisponderà un oggetto, mentre gli attributi relativi saranno rappresentati nelle colonne. Dunque, all'interno di ogni cella vi sarà un valore, esprimibile sia in maniera quantitativa che qualitativa, riferito all'oggetto della riga e all'attributo della colonna in questione. È possibile pre-trattare i dati, ovvero prevederne una discretizzazione; ciò è opportuno nelle analisi di tipo quantitativo, dove il dominio degli attributi può essere ripartito in classi, espresse da numeri naturali, con un risultato finale dipendente dalla modalità di discretizzazione utilizzata.

Secondo la Rough-Set Theory, ogni oggetto dell'universo U considerato è caratterizzato da una serie di informazioni o attributi $Q = (q_1, q_2, \dots, q_m)$ che lo definiscono. L'information

table dunque, può essere rappresentata come $S = (U, Q)$, ovvero l'universo U di tutti gli oggetti sotto studio e l'insieme finito Q degli attributi.

Definiamo il dominio di un attributo come V_q ed una funzione di informazione $f: U \times Q \rightarrow V$, tale che definito un $q \in Q$ e $x \in U$ sia $f(x, q) \in V_q$, in modo tale da restituire il valore dell'attributo relativo a quell'oggetto; in tal modo per qualsiasi sottoinsieme $P \subseteq Q$ può essere definita una relazione binaria I_p quale la seguente:

$$(4) I_p = \{(x, y) \in U \times U : f(x, q) = f(y, q), \forall q \in P\}$$

Questa rappresenta la relazione di indiscernibilità, ovvero una legge di equivalenza che assume gli oggetti indiscernibili come equivalenti, ovvero gli oggetti aventi le stesse informazioni rispetto agli attributi di P . Un esempio per rendere più chiara la relazione di indiscernibilità viene fornito dallo stesso Pawlawk (1991): in un'ottica clinica, i pazienti di un'ospedale rappresentano gli oggetti, dunque le righe della decision table, mentre i sintomi che questi presentano sono i valori all'interno delle celle, dipendenti dall'attributo presente in colonna. Suddividendo dunque i pazienti in base agli stessi sintomi si creeranno delle divisioni di U in classi di equivalenza denominate insiemi elementari o, nel caso in cui $P=Q$, atomi. Queste suddivisioni rappresentano le fondamenta su cui si basa la conoscenza relativa agli oggetti di U .

2.2 Approssimazioni e Rough-set

Da ciò che si è visto sopra, è possibile identificare dei sottoinsiemi $X \subseteq U$ rappresentati dagli insiemi elementari. In base all'incertezza ed all'ambiguità relativi alle informazioni a disposizione, tali sottoinsiemi possono essere più o meno precisi. Si evidenziano quindi due casi specifici ed opposti:

1. Un sottoinsieme X di tipo ordinario, preciso, ottenuto come semplice unione di insiemi elementari.
2. Un sottoinsieme X indefinito a causa dell'incongruenza dei dati.

Nel secondo caso, data l'impossibilità di ottenere un sottoinsieme X definito, questo sarà rappresentato tramite una coppia di sottoinsiemi definiti, chiamati approssimazione inferiore ed approssimazione superiore.

Per quanto riguarda l'approssimazione inferiore P_{\downarrow} , questa è costituita dall'unione di tutti gli insiemi elementari che sicuramente appartengono a X . Dunque, gli elementi costituenti $P_{\downarrow}(X)$ non sono altro che gli oggetti $x \in U$ appartenenti alle classi di equivalenza racchiuse in X derivate dalla relazione di indiscernibilità I_p :

$$(5) P_{\downarrow}(X) = \{x \in U : I_p(x) \subseteq X\}$$

L'approssimazione superiore, al contrario, è costituita da tutti gli insiemi elementari aventi un'intersezione non vuota con X , ovvero che potrebbero appartenere ad X . Così, facenti parte di $P^{\uparrow}(X)$ saranno tutti gli oggetti $x \in U$ che appartengono alle classi di equivalenza derivate dalla relazione di indiscernibilità I_p contenenti almeno un oggetto x appartenente ad X .

$$(6) P^\uparrow(X) = \bigcup_{x \in X} \{I_p(x)\}$$

Dunque, $P^\uparrow(X)$ risulta essere l'unione più grande degli insiemi elementari all'interno di X , e $P_\downarrow(X)$ l'unione più piccola dei medesimi all'interno di X .

Operando una differenza tra l'approssimazione superiore e quella inferiore si ottiene la frontiera $B_{nP}(X)$, ovvero la regione dell'incertezza, in cui gli elementi appartenenti ad essa, impiegando le informazioni a disposizione, non possono essere dichiarati con certezza se appartenenti o meno ad X . Tale frontiera è rappresentata da:

$$(7) B_{nP} = P^\uparrow(X) - P_\downarrow(X)$$

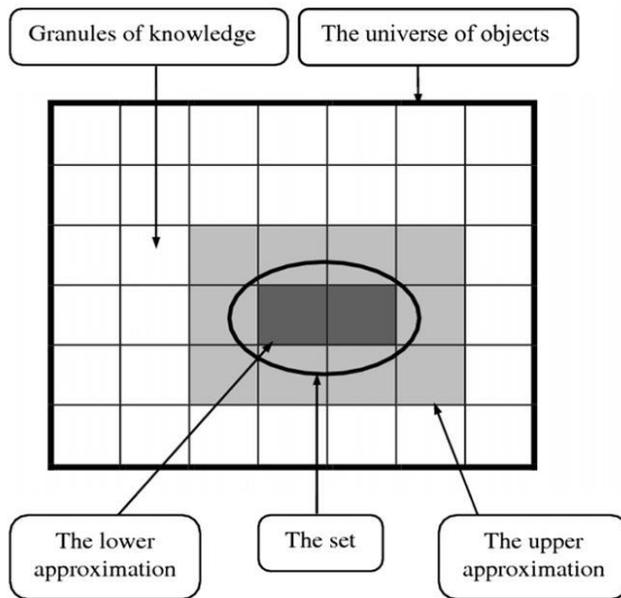
Nella frontiera dunque si trovano tutti quegli oggetti aventi informazioni ambigue e non chiare; sarebbe vuota se si trattasse di un insieme esatto o *crisp*, ovvero espressione dell'unione di insiemi elementari. In caso contrario l'insieme diventa *rough*, approssimato, e delimitato dagli insiemi superiore e inferiore. Tutti gli insiemi $X \subseteq U$ caratterizzati dalle stesse approssimazioni superiori e inferiori formano il *Rough-set*; complessivamente vale la seguente relazione:

$$(8) P_\downarrow(X) \subseteq X \subseteq P^\uparrow(X)$$

Dunque, $B_{nP}(X)$ rappresenta la regione dell'incertezza di X , sicchè non è possibile affermare con certezza l'appartenenza o meno di un suo elemento ad X . Infatti x appartiene sicuramente ad X qualora questo rientri in $P_{\downarrow}(X)$, mentre non si può dire lo stesso qualora questo appartenga a $P^{\uparrow}(X)$: è possibile che vi appartenga ma non si può averne la certezza.

Figura 1 – L'approssimazione dei Rough-Sets

Z. Pawlak, A. Skowron / Information Sciences 177 (2007) 3–27



Considerando $X \neq \emptyset$ si può definire l'accuratezza dell'approssimazione di X tramite questa relazione:

$$(9) \quad \alpha_P(X) = \frac{|P_{\downarrow}(X)|}{|P^{\uparrow}(X)|}$$

$\alpha_P(X)$ può assumere valori tra 0 ed 1; per un valore risultante =1, X rappresenta un insieme esatto o *crisp*, mentre per un valore $\neq 1$ rappresenta un insieme approssimato o *rough*.

Oltre all'accuratezza, è possibile anche determinare la qualità dell'approssimazione di X in questo modo:

$$(10) \quad \gamma_P(X) = \frac{|P_{\downarrow}(X)|}{|X|}$$

Quest'indice esprime la frequenza relativa degli oggetti di X correttamente attribuiti all'approssimazione inferiore per mezzo degli attributi di P .

Ancora, è possibile stabilire la probabilità o grado di certezza con cui x appartiene ad X :

$$(11) \mu_X^P(x) = \frac{|X \cap I_P(x)|}{|I_P(x)|}$$

In tal modo è possibile determinare la probabilità che un oggetto x appartenga ad un sottoinsieme dell'universo $X \subseteq U$.

2.3 Dipendenza e Reduction degli attributi

Molto importanti da un punto di vista pratico sono gli aspetti relativi alla dipendenza tra gli attributi e l'eliminazione di quelli superflui. Prendendo due insiemi di attributi $T \subseteq Q$, questi sono dipendenti tra di loro a patto che i valori degli attributi in T siano dati dai valori degli attributi in P , ovvero $I_P \subseteq I_T$. Semplificando, la ripartizione effettuata tramite gli attributi di P è raffinata tanto quanto quella ottenuta dagli attributi di T , così che è sufficiente utilizzare gli attributi di P per ottenere la partizione $U|I_T$. Per quanto riguarda gli attributi superflui, rappresentati *nell'information table*, questi possono essere esclusi, senza andare ad inficiare l'informazione originale.

Ad esempio, per determinare se un attributo p è superfluo in P , dati $P \subseteq Q$ e $p \in P$, si elimina p da I_P , e se $I_P = I_{P-\{p\}}$ allora l'attributo è removibile senza andare a minare le informazioni originarie, mentre in caso contrario l'attributo risulta essere indispensabile in P . P si definisce indipendente quando tutti i suoi attributi sono indispensabili, al netto di quelli

eliminati precedentemente. Infatti il sottoinsieme P' , definito anche come $Red(P)$ (*Reduction*) ed effetto della riduzione degli attributi superflui di P , può essere definito tale solo qualora esso sia indipendente e valga la suddetta relazione $I_{P'} = I_P$. Una volta eliminati tutti gli attributi non necessari, e quindi avendo solo gli attributi indispensabili, si delinea il nucleo o *core* di P :

$$(12) Core_x(P) = \cap Red_x(P)$$

ovvero quell'insieme di attributi che non possono essere eliminati pena la perdita di informazioni. Grazie a questa proprietà della teoria dei Rough-set è dunque possibile ottenere gli stessi risultati, senza perdita di informazioni, con un minor numero di attributi rispetto a quelli che potrebbero essere utilizzati.

2.4 Decision Table e regole decisionali

Una volta eliminati gli attributi non necessari e operate le discretizzazioni del caso sul valore di questi, ci si ritrova ad avere una tabella chiamata Tabella di decisione o *Decision table* con tutti i dati necessari per operare le analisi Rough-set. La questione più rilevante è quella di suddividere gli elementi in sottoinsiemi tramite un procedimento logico con le informazioni a disposizione. Al fine di fare ciò, è necessario che le informazioni siano rappresentate tramite due tipologie di attributi:

1. Attributi decisionali (D)
2. Attributi condizionali (C)

Lo scopo è quello di rilevare una connessione tra condizioni e decisione, in un'ottica "if...then". Per ogni oggetto x appartenente ad U viene posta la condizione "if" davanti agli attributi condizionali, mentre "then", ovvero la conseguenza delle condizioni in essere, si

riferisce agli attributi decisionali. È vero però che non tutte le regole così formatesi possono essere certe, in quanto alcune informazioni potrebbero essere caratterizzate da inconsistenza o incertezza. Tramite il metodo del Rough-set è possibile ottenere come risultato finale una lista delle regole decisionali, sia certe che incerte, più o meno approssimate.

Nel campo pratico, gli attributi decisionali vengono utilizzati per determinare la relazione di indiscernibilità I_D , ovvero andare a suddividere l'universo in classi di decisione elementari in base agli attributi decisionali. Gli attributi condizionali a loro volta fungono a creare la relazione di indiscernibilità I_C , la quale ripartisce l'universo in insiemi elementari in base agli attributi condizionali stessi, che serviranno ad approssimare le classi di decisione. Per ogni classe di decisione cioè, si possono determinare l'approssimazione superiore ed inferiore degli attributi condizionali, ovvero una distinzione tra gli oggetti che sicuramente appartengono alla classe e quelli che potrebbero appartenervi. Si tratta quindi di trovare le condizioni per le quali un oggetto appartiene con certezza ad una classe e, al contempo, quelle per le quali vi è incertezza. Ciò che viene a crearsi è una corrispondenza "rough" tra le classi condizionali e decisionali, in cui ogni classe di decisione viene approssimata da un insieme superiore ed uno inferiore della classe decisionale. Da questo deriva l'estrapolazione di regole certe ed incerte, le prime date dagli oggetti appartenenti all'insieme di approssimazione inferiore, le seconde da quelli rientranti nella regione di frontiera. Da queste regole chiaramente, il campo di applicazione può essere allargato ad un più vasto universo di oggetti dei quali siano conosciuti solo gli attributi condizionali.

3. Analisi Empirica

3.1 I dati

Nel momento dell'applicazione della Teoria dei Rough-Set, si presentano due principali scelte da compiere per procedere con l'analisi: la scelta degli oggetti, e degli attributi o variabili che li definiscono. Iniziando con gli oggetti, che in questo caso sono rappresentati da aziende, queste sono state scelte secondo dei criteri che garantissero uno standard di omogeneità sufficiente come visto in parte della letteratura. Per l'estrazione delle aziende è stato utilizzato il Database AIDA (Analisi Informatizzata Delle Aziende), uno strumento tramite il quale è possibile visionare ed esportare i bilanci di circa 1.200.000 società italiane. Questa banca dati offre la possibilità di operare dei filtri nella ricerca, andando così a ridurre le differenze tra le aziende selezionate.

I criteri di ricerca sono stati innanzitutto posti con riguardo alla dimensione aziendale, al fine di non rischiare di andare a comparare realtà troppo diverse tra di loro. È così che è stato impostato un filtro per i ricavi delle vendite, compresi tra 1.000.000 € e 10.000.000 €, mentre il numero di dipendenti è stato selezionato tra 50 e 200; entrambi presenti in almeno uno degli ultimi due bilanci disponibili. La forma giuridica scelta è stata quella delle società di capitali più diffuse in Italia, ovvero S.p.a. ed S.r.l., anche qua per andare ad escludere possibili casi di contaminazione da società di persone. La scelta di questo tipo di dimensionamento dipende dal fatto che l'analisi vuole essere concentrata su aziende di medie e piccole dimensioni, tipiche del territorio italiano. Per quanto riguarda la base temporale, ricorrendo ancora alla letteratura si nota come nella ricerca sulla predizione del fallimento, venga data molta importanza agli anni di vita delle società, ritrovando spesso negli articoli una scelta di includere aziende che siano in attività da almeno 5 anni. In questo caso la scelta è stata quella di includere società con una vita di almeno 10 anni, ovvero costituite entro l'anno 2000 ed il cui ultimo bilancio disponibile risalga tra il 2015 e il 2011. Il motivo di questa decisione è che, volendo analizzare aziende fallite dopo la recente crisi

finanziaria, si desidera che queste siano state costituite prima dell'avvenimento di questa, e non in concomitanza; in tal modo si eliminano altre eventuali e prevedibili alterazioni dei dati.

In conclusione, come ultima discriminante, sono state selezionate solamente aziende facenti parte della categoria merceologica C del codice ATECO 2007, ovvero le imprese manifatturiere. Anche questa scelta ha delle motivazioni, dovute a due fattori: la prima è che in buona parte della bibliografia, indipendentemente dal metodo utilizzato, vengono analizzate aziende manifatturiere; la seconda invece, dipende dal fatto che l'industria manifatturiera, soprattutto quella di piccole e medie dimensioni, rappresenta il cuore del paese, che ha subito gravi danni nel corso della crisi. Lo scopo quindi è anche quello di vedere se è possibile evidenziare la presenza di indici "critici" ai quali sia attribuibile il fallimento post-crisi.

La scelta del campione è proseguita in due tempi; inizialmente l'attenzione è stata posta sulle aziende fallite, secondo i criteri di ricerca sopra specificati. Sono state selezionate 50 società fallite, prese in ordine di ricavi delle vendite, e da queste ne sono state scelte 35 al fine di ottenere la completezza dei dati ed andare ad escludere quelle con i dati più eterogenei. Allo stesso modo sono state poi selezionate 35 aziende attive.

Successivamente, sono stati scelti gli attributi che specificheranno ogni azienda all'interno dell'*information table*. Non volendo avventurarsi nella ricerca di indici "esotici" o nuovi, è stato scelto di seguire l'esempio di Dimitras et al. (1998) in cui sono stati utilizzati 12 indici ricavati dai dati di bilancio e che rispondono a 3 precise istanze: Indici che rappresentano la profittabilità, la performance manageriale e la solvibilità. Questi indici sono:

1. Net Profit / Gross Profit: calcolato come Utile Netto / (Valore della Produzione – Costi della Produzione); rappresenta il rapporto tra il profitto effettivo e quello lordo.

2. $\text{Gross Profit} / \text{Total Assets}$: calcolato come $(\text{Valore della Produzione} - \text{Costi della Produzione}) / \text{Totale Attività}$; esprime quanto profitto lordo un'azienda sta producendo in base alle sue attività di bilancio.
3. $\text{Net Profit} / \text{Total Assets}$: calcolato come $\text{Utile Netto} / \text{Totale Attività}$; come l'indice precedente, l'attenzione viene però posta al risultato finale d'esercizio.
4. $\text{Net Profit} / \text{Net Worth}$: calcolato come $\text{Utile Netto} / \text{Patrimonio Netto}$; indica in che grado una società sta utilizzando i propri *asset* per trarre profitto.
5. $\text{Current Assets} / \text{Current Liabilities}$: calcolato come $\text{Attivo Circolante (al netto di crediti e attività finanziarie esigibili oltre l'esercizio)} / \text{Debiti entro l'esercizio}$; esprime la liquidità e quindi la capacità dell'azienda di onorare i propri debiti a breve periodo con le attività di breve periodo.
6. $\text{Quick Assets} / \text{Current Liabilities}$: calcolato come $\text{Current Assets (al netto delle Rimanenze)} / \text{Debiti entro l'esercizio}$; simile all'indice precedente, rappresenta però delle differenze sostanziali nel caso di società manifatturiere in cui le rimanenze costituiscono un'importante voce di bilancio.
7. $(\text{Long Term Debt} + \text{Current Liabilities}) / \text{Total Assets}$: calcolato come $\text{totale dei debiti entro e oltre l'esercizio} / \text{totale attività}$; anche questo un indice di liquidità, rappresentante la capacità di onorare la totalità dei debiti aziendali con la totalità degli *asset*, sia di breve che di lungo periodo.
8. $\text{Net Worth} / (\text{Net Worth} + \text{Long Term Debt})$: calcolato come $\text{Patrimonio Netto} / (\text{Patrimonio Netto} + \text{Debiti oltre l'esercizio})$; questo indice rappresenta sempre la solvibilità, con un'attenzione particolare però per i debiti oltre l'esercizio.
9. $\text{Net Worth} / \text{Net Fixed Assets}$: calcolato come $\text{Patrimonio Netto} / \text{Totale Immobilizzazioni}$; mostra in che grado le immobilizzazioni sono frutto del finanziamento dei soci.
10. $\text{Inventories} / \text{Working Capital}$: calcolato come $\text{Rimanenze} / (\text{Current Assets} - \text{Current Liabilities})$; questo indice raffigura la porzione di rimanenze sostenuta dalla presenza di liquidità.

11. Current Liabilities / Total Assets: calcolato come Debiti entro l'esercizio / Totale Attività; anche questo è un indicatore di solvibilità che mostra in che proporzione le attività dell'azienda sono composte da debiti entro l'esercizio.

12. Working Capital / Net Worth: calcolato come (Current Assets – Current Liabilities) / Patrimonio Netto;

Quelli appena citati rappresentano gli attributi condizionali che andranno a formare le colonne della tabella d'informazione; verranno nominati in ordine crescente da *a1* ad *a12* e a questi verrà aggiunto un tredicesimo attributo Decisionale *d* a descrivere lo Stato Giuridico dell'azienda: 0 sarà per fallita, 1 per attiva.

Dopo avere descritto le modalità di reperimento dei dati, verrà ora condotta una breve analisi descrittiva di questi.

3.2 Analisi Descrittiva

Prima di procedere con l'analisi Rough-Set, risulta opportuno uno studio dei dati per poter essere poi in grado di comprendere meglio i risultati, date le informazioni di partenza.

La Tabella 1 riporta la media e le statistiche di sintesi dei 12 attributi per il gruppo delle 35 aziende fallite. Come si può notare alcuni indici presentano dei valori molto diversi fra di loro, in particolare *a4*, *a9*, *a10* e *a11*.

Tabella 1 – Statistiche di sintesi Aziende Fallite

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
a1 Net Income/ Gross Profit	35	.05	3.58	1.2305	.59651
a2 Gross Profit/ Total Asset	35	-2.11	.09	-.4841	.50003
a3 Net Income/ Total Asset	35	-2.15	.04	-.5824	.54827
a4 Net Income/ Net Worth	35	-5.13	10.20	.9467	2.12941
a5 Current Assets/ Current Liab	35	.07	1.95	.6670	.47111
a6 Quick Assets/ Current Liab	35	.02	1.55	.4411	.34281
a7 Tot Debt/ Total Asset	35	.68	5.71	1.5555	1.09703
a8 Net Worth/ Net Worth+LT Debt	35	-.57	7.13	1.2458	1.28026
a9 Net Worth/ Net Fixed Asset	35	-11.41	8.44	-1.8506	3.74899
a10 Inventories/ Working Capital	35	-2.87	30.32	1.8914	6.31686
a11 Current Liab/ Total Asset	35	.30	5.71	1.3952	1.12120
a12 Working Capital/ Net Worth	35	.03	9.56	1.7254	2.04664
Valid N (listwise)	35				

Elaborazione personale dei dati dal Dataset AIDA

Risulta interessante il comportamento della deviazione standard di *a10*, dovuta probabilmente al fatto che trattandosi di imprese manifatturiere, le voci dell'inventario possono avere grandi oscillazioni, in base anche alle scelte di gestione effettuate.

Tabella2 – Statistiche di sintesi Aziende Attive

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
a1 Net Income/ Gross Profit	35	.17	3.29	.6465	.63022
a2 Gross Profit/ Total Asset	35	.00	.38	.0899	.08174
a3 Net Income/ Total Asset	35	.01	.25	.0556	.06112
a4 Net Income/ Net Worth	35	.02	.81	.1938	.16951
a5 Current Assets/ Current Liab	35	1.02	4.67	1.5971	.80072
a6 Quick Assets/ Current Liab	35	.50	4.45	1.2550	.79511
a7 Tot Debt/ Total Asset	35	.20	.85	.6348	.17798
a8 Net Worth/ Net Worth+LT Debt	35	.33	1.00	.7407	.22719
a9 Net Worth/ Net Fixed Asset	35	.38	17.41	2.7799	3.66003
a10 Inventories/ Working Capital	35	.00	16.70	1.9408	3.41902
a11 Current Liab/ Total Asset	35	.20	.83	.5400	.15515
a12 Working Capital/ Net Worth	35	.04	2.88	.8258	.52423
Valid N (listwise)	35				

Elaborazione personale dei dati dal Dataset AIDA

Per quanto riguarda le aziende in stato di attività (Tabella 2) la deviazione standard appare essere più accentuata sugli attributi *a9* e *a10*, in tendenza con il comportamento delle aziende fallite, facendo pensare che siano dati attendibili dal settore di riferimento. In entrambi i gruppi, ma decisamente in maggior misura tra le aziende fallite, l'attributo *a1*

presenta dei valori fuori norma a prima vista; dipende dalla presenza di perdite. Allo stesso modo a_9 avente una distribuzione della media così evidentemente diversa tra i due gruppi, è spiegabile per via della presenza di valori del Patrimonio Netto negativi dovuti al fatto che le aziende si trovano nell'anno prima del fallimento.

Può essere interessante anche andare a osservare le correlazioni tra gli attributi del gruppo delle aziende Fallite, al fine di indagare sulla presenza di correlazioni. L'analisi di correlazione viene svolta tramite l'indice di correlazione di Pearson con il seguente sistema di ipotesi:

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

I valori riportati in tabella 3 seguiti da un asterisco rappresentano un livello di significatività inferiore allo 0,05 mentre quelli seguiti da due asterischi inferiore allo 0,01.

Tabella 3 – Correlazioni Aziende Fallite

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
a1	1,0	0,0	0,0	-0,1	-0,2	-0,2	0,0	0,0	0,0	.393*	0,0	-0,2
a2	0,0	1,0	.983**	-0,1	.679**	.589**	-.693**	-0,2	.607**	0,3	-.707**	0,1
a3	0,0	.983**	1,0	-0,1	.712**	.617**	-.710**	-0,2	.621**	0,3	-.721**	0,2
a4	-0,1	-0,1	-0,1	1,0	-0,3	-0,2	0,0	0,2	-0,1	-0,3	0,0	.593**
a5	-0,2	.679**	.712**	-0,3	1,0	.829**	-.635**	-0,3	.599**	0,3	-.633**	0,0
a6	-0,2	.589**	.617**	-0,2	.829**	1,0	-.573**	-0,3	.574**	0,2	-.546**	0,0
a7	0,0	-.693**	-.710**	0,0	-.635**	-.573**	1,0	0,0	-.676**	-0,2	.993**	-0,2
a8	0,0	-0,2	-0,2	0,2	-0,3	-0,3	0,0	1,0	0,0	-0,3	0,0	-0,1
a9	0,0	.607**	.621**	-0,1	.599**	.574**	-.676**	0,0	1,0	0,2	-.696**	0,2
a10	.393*	0,3	0,3	-0,3	0,3	0,2	-0,2	-0,3	0,2	1,0	-0,2	-0,2
a11	0,0	-.707**	-.721**	0,0	-.633**	-.546**	.993**	0,0	-.696**	-0,2	1,0	-0,1
a12	-0,2	0,1	0,2	.593**	0,0	0,0	-0,2	-0,1	0,2	-0,2	-0,1	1,0

Elaborazione personale dei dati dal Dataset AIDA

Come è possibile osservare, l'ipotesi viene accettata in molti casi, ma si riscontrano anche numerosi casi di correlazioni importanti tra attributi simili, aventi ad esempio in comune il denominatore dell'indice.

Questo è del tutto normale ed in linea con i risultati di altri articoli in letteratura aventi utilizzato variabili simili; comunque in letteratura l'attenzione viene posta sul fatto che non ci deve essere correlazione tra almeno alcune delle variabili, in modo da garantire che vengano rappresentati aspetti aziendali eterogenei. Nel nostro caso, come detto poco sopra, appare non vi siano problemi di correlazione e in particolare gli attributi $a1$, $a4$, $a8$ e $a12$ risultano non correlati con gran parte degli altri attributi.

Si prosegue con un'ultima analisi al fine di indagare la possibile connessione tra ogni singola variabile e lo stato di Attività o Fallimento dell'azienda. A tale fine viene utilizzato l'*Independent sample t-Test* secondo il seguente sistema di ipotesi:

- $H_0: \mu_1^1 = \mu_1^2$
- $H_1: \mu_1^1 \neq \mu_1^2$

In questa elaborazione le variabili vengono suddivise in due gruppi in base al valore dell'attributo "Stato Giuridico", e di questi viene comparata la media, rispettivamente μ_1^1 e μ_1^2 nel sistema di ipotesi riportato poco sopra. L'ipotesi viene rifiutata in tutti gli attributi, solo $a10$ presenta valori simili in entrambi i gruppi, portando a pensare che non sia utilizzabile quindi in caso di analisi Univariata. Questo comunque non ne pregiudica la capacità discriminante, in quanto questa analisi fa riferimento unicamente ad ogni attributo condizionale in connessione all'attributo decisionale, e non è dunque considerabile l'influenza degli altri attributi. Per quanto riguarda il resto delle variabili, presentando valori considerevolmente diversi a seconda del gruppo di riferimento, è possibile affermare che hanno capacità discriminante, anche se non è chiaramente possibile stabilirne il grado. Nella Tabella 4 sono rappresentati gli esiti dell'analisi:

Tabella 4- Independent sample t-Test

		N	Mean	Std. Deviation	Std. Error Mean
a1 Net Income/ Gross Profit	Fallita	35	1,2305	,59651	,10083
	Attiva	35	,6465	,63022	,10653
a2 Gross Profit/ Total Asset	Fallita	35	-,4841	,50003	,08452
	Attiva	35	,0899	,08174	,01382
a3 Net Income/ Total Asset	Fallita	35	-,5824	,54827	,09268
	Attiva	35	,0556	,06112	,01033
a4 Net Income/ Net Worth	Fallita	35	,9467	2,12941	,35994
	Attiva	35	,1938	,16951	,02865
a5 Current Assets/ Current Liab	Fallita	35	,6670	,47111	,07963
	Attiva	35	1,5971	,80072	,13535
a6 Quick Assets/ Current Liab	Fallita	35	,4411	,34281	,05795
	Attiva	35	1,2550	,79511	,13440
a7 Tot Debt/ Total Asset	Fallita	35	1,5555	1,09703	,18543
	Attiva	35	,6348	,17798	,03008
a8 Net Worth/ Net Worth+LT Debt	Fallita	35	1,2458	1,28026	,21640
	Attiva	35	,7407	,22719	,03840
a9 Net Worth/ Net Fixed Asset	Fallita	35	-1,8506	3,74899	,63369
	Attiva	35	2,7799	3,66003	,61866
a10 Inventories/ Working Capital	Fallita	35	1,8914	6,31686	1,06774
	Attiva	35	1,9408	3,41902	,57792
a11 Current Liab/ Total Asset	Fallita	35	1,3952	1,12120	,18952
	Attiva	35	,5400	,15515	,02623
a12 Working Capital/ Net Worth	Fallita	35	1,7254	2,04664	,34595
	Attiva	35	,8258	,52423	,08861

Elaborazione personale dei dati dal Dataset AIDA

3.3 Applicazione della Rough-Set Theory

Dopo aver analizzato i dati, si passa ora all'analisi vera e propria, seguendo gli *step* dell'analisi Rough-Set.

Per prima cosa verrà ricercata una discretizzazione che possa permettere una lettura e catalogazione più semplice dei dati, portandoli da valori continui ad un intervallo di numeri interi. Dopo di ciò si cercherà la presenza di attributi nel *Core*, ovvero quegli attributi la cui eliminazione comporterebbe una perdita di informazioni, e contestualmente verranno calcolati tutti i sottoinsiemi dei Reduct, ovvero gruppi di attributi ristretti ma senza perdita di informazioni. Da questi ne verrà scelto uno, e dalla Tabella di Informazione così ridotta verranno indotte le regole decisionali che saranno poi validate.

Questa analisi sarà condotta utilizzando il software ROSE2, sviluppato dal Professore Roman Słowiński dell'University of Technology di Poznań in Polonia.

3.3.1 Discretizzazione degli Attributi

La scelta delle regole per la discretizzazione degli attributi gioca un ruolo fondamentale in quelli che saranno poi i risultati, o meglio le regole decisionali indotte. La prima tipologia di discretizzazione provata è stata la seguente:

A1 : (-inf, 0.00) <0.00, 0.25) <0.25, 1.00) <1.00, +inf)

A2 : (-inf, -0.00) <-0.00, 0.25) <0.25, 0.50) <0.50, +inf)

A3 : (-inf, -0.05) <-0.05, 0.05) <0.05, 0.20) <0.20, +inf)

A4 : (-inf, 0.00) <0.00, 0.25) <0.25, +inf)

A5 : (0.00, 0.70) <0.70, 1.00) <1.00, 1.50) <1.50, 2.00) <2.00, +inf)

A6 : (0.00, 0.50) <0.50, 0.80) <0.80, 1.00) <1.00, 1.20) <1.20, +inf)

A7 : (0.00, 0.667) <0.667, 0.80) <0.80, 1.00) <1.00, +inf)

A8 : (-inf, 0.00) <0.00, 0.50) <0.50, +inf)

A9 : (-inf, 0.00) <0.00, 0.30) <0.30, +inf)

A10 : (-inf, 0.00) <0.00, 0.50) <0.50, 0.75) <0.75, 1.00) <1.00, +inf)

A11 : (-inf, 0.25) <0.25, 0.50) <0.50, 0.75) <0.75, 1.00) <1.00, +inf)

A12 : (-inf, 0.00) <0.00, 0.50) <0.50, +inf)

Questa suddivisione proposta da Dimitras et al. (1999) ha portato ad ottenere dei dati fin troppo normalizzati, dovuto al fatto che la discretizzazione era stata tarata sul loro Dataset. In questo modo il software non riusciva a discriminare con risultati soddisfacenti ed è stato necessario provare un altro metodo.

ROSE2, oltre che alle funzioni sopra enunciate, è anche in grado di operare discretizzazioni creando degli intervalli basati sull'analisi dei dati stessi. È così che è stata utilizzata questa metodologia di discretizzazione:

A1 : (-inf, 0.12) <0.12, 0.87) <0.87, 2.615) <2.615, +inf)

A2 : (-inf, -0.005) <-0.005, 0.085) <0.085, 0.095) <0.095, +inf)

A3 : (-inf, 0.005) <0.005, 0.025) <0.025, 0.045) <0.045, +inf)

A4 : (-inf, 0.015) <0.015, 0.19) <0.19, 0.885) <0.885, +inf)

A5 : (-inf, 1.015) <1.015, 1.06) <1.06, 1.955) <1.955, +inf)

A6 : (-inf, 0.485) <0.485, 0.805) <0.805, 0.865) <0.865, +inf)

A7 : (-inf, 0.675) <0.675, 0.705) <0.705, 0.865) <0.865, +inf)

A8 : (-inf, 0.405) <0.405, 0.995) <0.995, 1.015) <1.015, +inf)

A9 : (-inf, 0.34) <0.34, 0.535) <0.535, 0.71) <0.71, +inf)

A10 : (-inf, -0.005) <-0.005, 0.255) <0.255, 7.96) <7.96, +inf)

A11 : (-inf, 0.44) <0.44, 0.515) <0.515, 0.855) <0.855, +inf)

A12 : (-inf, 0.835) <0.835, 0.91) <0.91, 1.35) <1.35, +inf)

I risultati portati dalla scelta di questo metodo sono stati di gran lunga migliori; per ogni intervallo si intende una numerazione che va da 0 a 3.

3.3.2 Reduct e Core

Una volta discretizzati i valori all'interno dell' *Information Table*, è possibile con maggiore semplicità calcolare il Core degli attributi e le relative Reduct.

È interessante notare come utilizzando la prima norma di discretizzazione apparisse un attributo in Core, in particolare *a8*. Con le seconde norme invece, nessun attributo rientra nel Core, a significare quindi che nessuno di essi è indispensabile per spiegare lo stato di attività o fallimento. Questo è in linea con i risultati iniziali di Dimitras et al. (1999) ed anche per questo si è deciso di proseguire con il secondo metodo di discretizzazione.

I Reduct estratti dall' *Information Table* sono in tutto 36, vengono di seguito rappresentati i primi 15 in ordine di lunghezza nella tabella 5:

Tabella 5 – I reduct

#	Reduct	Lunghezza
1	a6, a7, a10, a11, a12	5
2	a2, a3, a4, a8	4
3	a1, a2, a8, a10	4
4	a1, a2, a4, a8	4
5	a3, a4, a8, a9	4
6	a1, a2, a5, a8	4
7	a1, a8, a9, a10	4
8	a6, a8, a9, a12	4
9	a6, a7, a9, a12	4
10	a5, a6, a7, a12	4
11	a6, a9, a10, a12	4
12	a2, a3, a5	3
13	a2, a4, a5	3
14	a1, a2, a6	3
15	a2, a3, a6	3

Elaborazione personale da Dataset AIDA

Tabella 6 – Frequenza degli attributi nel reduct

#	Attributo	Frequenza	% Frequenza
1	a2	15	41.67%
2	a6	14	38.89%
3	a9	14	38.89%
4	a4	12	33.33%
5	a3	10	27.78%
6	a8	9	25.00%
7	a1	9	25.00%
8	a10	7	19.44%
9	a11	7	19.44%
10	a7	7	19.44%
11	a5	6	16.67%
12	a12	6	16.67%

Elaborazione personale da Dataset AIDA

Dalla Tabella 6 è possibile osservare la frequenza degli attributi all'interno dei *Reduct*. Essendo ogni *Reduct* equivalente agli altri per quanto riguarda l'assenza di perdita di informazioni, è stato scelto di proseguire nell'elaborazione dei dati con il *Reduct* numero 3. Questo è stato scelto perché comprendente indicatori della redditività quali *a1* e *a2*, oltre che l'attributo *a8* che rappresentava l'attributo in Core nella prima norma di

discretizzazione utilizzata e si focalizza sulla presenza di debiti a lungo termine, e infine *a10*, che si era dimostrato l'unico attributo a non avere connessione con lo Stato Giuridico aziendale. Una volta eseguita la riduzione degli attributi, si ottiene la seguente Tabella d'Informazione:

Tabella 7 – Tabella di informazione ridotta

a1	a2	a8	a10	D		a1	a2	a8	a10	D
2	0	0	3	0		1	1	1	2	1
2	0	0	3	0		1	3	2	1	1
2	0	3	0	0		1	3	1	2	1
2	0	3	0	0		1	3	1	3	1
2	0	3	0	0		1	1	1	1	1
2	0	3	0	0		1	1	1	2	1
2	0	3	0	0		1	1	1	2	1
2	0	2	1	0		1	1	1	2	1
2	0	3	0	0		1	1	1	3	1
1	2	1	1	0		1	1	1	2	1
2	0	0	0	0		1	1	1	2	1
2	0	3	0	0		1	3	1	2	1
2	0	1	1	0		1	1	1	2	1
2	0	2	0	0		1	1	1	2	1
2	0	3	0	0		1	3	1	1	1
2	0	0	2	0		1	3	2	1	1
0	1	0	2	0		1	1	1	2	1
2	0	3	0	0		1	1	0	2	1
2	0	0	0	0		1	1	1	2	1
3	0	0	3	0		1	1	1	2	1
2	0	3	0	0		1	3	1	2	1
2	0	3	0	0		1	3	0	2	1
2	0	3	0	0		1	3	2	2	1
2	0	2	0	0		1	1	1	2	1
2	0	3	0	0		1	1	2	2	1
2	0	2	0	0		1	1	2	2	1
2	0	2	0	0		1	3	1	2	1
0	1	0	2	0		1	1	2	2	1
2	0	3	0	0		1	1	2	2	1
1	1	2	3	0		1	1	1	2	1
2	0	3	0	0		1	3	1	1	1
2	0	3	0	0		3	1	2	1	1
2	0	3	0	0		1	1	1	2	1
2	0	3	0	0		1	1	1	2	1
2	0	3	0	0		3	1	1	2	1

Elaborazione personale da Dataset AIDA

La tabella rappresenta la parte finale della preparazione dei dati; è stata divisa in due colonne per semplificarne la visualizzazione in una pagina. A partire dai dati sopra esposti, verranno ora estrapolate le regole decisionali tramite gli algoritmi induttivi di ROSE2.

3.3.3 Induzione delle Regole decisionali

La generazione delle regole avviene tramite l'utilizzo di una serie di algoritmi induttivi mirati ad ottenere una "descrizione discriminante" o parzialmente discriminante di entrambe le classi decisionali, ovvero un insieme di regole di tipo *if $a_1 = "x"$; then $D = "y"$*

Tramite Rose 2, si ha la possibilità di utilizzare i seguenti 3 diversi metodi per la generazione di regole decisionali:

- Basic Minimal Covering
- Extended Minimal Covering
- Satisfactory Description

Tralasciando l'Extended Minimal Covering, il quale fornisce indicazioni troppo generali e di scarso interesse, ovvero l'insieme di tutte le possibili regole, l'attenzione viene posta sui restanti due. Il Basic Minimal Covering consiste nell'insieme minimo di regole che comprendono tutti gli oggetti dell'Information Table, mentre la Satisfactory Description fornisce regole decisionali "forti", anche parzialmente discriminanti, ma non obbligatoriamente comprendenti tutti gli oggetti. La differenza principale tra i due metodi sta quindi nel fatto che mentre il Basic Minimal Covering offre esclusivamente regole di carattere certo, non si può dire lo stesso per la Satisfactory Description, che genera regole si forti, ma anche incerte. Per regole "forti", si intende la frequenza di oggetti facenti parte del set oggetto della regola, ovvero per quanti oggetti si ripete un dato valore dell'attributo che condiziona il valore dell'attributo decisionale. Il fatto che una regola sia "forte" non comporta il fatto che sia certa; infatti tale regola può rappresentare un set di oggetti esteso,

ma non è detto che tutti questi oggetti abbiano lo stesso valore in d . In tal caso la regola viene definita parzialmente discriminante.

Per semplificare, viene rappresentata la modalità di visualizzazione di una regola:

rule 1. (A2 in {0, 2}) => (D = 0); [32, 32, 91.43%, 100.00%][32, 0]

{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 31, 32, 33, 34, 35}, {}

Questa regola, facente parte del gruppo di regole minime, è stata ricavata selezionando la modalità “estendi condizioni”, in tal modo è possibile aumentare la forza della regola. In questo caso il significato è che se l’attributo a_2 ha valore compreso tra 0 e 2, l’azienda sarà da considerarsi fallita. Questo avviene nel 91.43% degli oggetti facenti parte delle aziende fallite, il che rappresenta la forza della regola, ovvero 32, e ha capacità discriminante del 100% riferita agli oggetti medesimi.

Viene ora raffigurato in tabella 8 il set di regole minime:

Tabella 8 – Regole minime indotte dal reduct3

# Regola	Condizioni				Decisione	
	a1	a2	a8	a10	D	Forza
1		0-2			0	32
2	0				0	2
3			2	3	0	1
4	1-3			2	1	27
5		3			1	4
6		1		1	1	2
7			1	3	1	2

Elaborazione personale da Dataset AIDA

Come si può facilmente notare le regole con la maggior forza sono quelle estese a più valori; le regole indicanti lo stato di fallimento sono inferiori rispetto a quelle indicanti lo stato di attività, il che può far pensare vi sia una maggiore difficoltà per gli algoritmi di rilevare i sintomi di cattiva salute aziendale piuttosto che quelli di un andamento ordinario e non fallimentare. In ogni caso, date le dimensioni ristrette del campione e la non omogeneità massima dei dati, è impossibile attendersi la generazione di un gran numero di regole minime.

Il secondo set di regole generato tramite la *Satisfactory Description*, prevede, prima del lancio degli algoritmi induttori, la taratura di questi. Dopo alcune prove è stato deciso di impostare la lunghezza massima della regola a 2, con una copertura relativa degli esempi della classe del 10% e un livello di capacità discriminante minima del 90%. Il risultato è il seguente set di 8 regole:

Tabella 9 – Regole forti indotte dal reduct3

# Regola	Condizioni				Decisione	
	a1	a2	a8	a10	D	Forza
1	2				0	30
2		0			0	31
3			3		0	19
4				0	0	25
5	1				1	33
6		3			1	11
7			1		1	25
8				2	1	27

Elaborazione personale da Dataset AIDA

Come si può vedere nella tabella 9, le regole sono aumentate di un' unità ma soprattutto la forza che le caratterizza è aumentata. Questo è dovuto ai criteri impostati da un lato più stringenti che permettono di ottenere un modesto numero di set di regole con una

rappresentatività giudicabile buona, anche se si crea il rischio di discriminare erroneamente.

Una volta ottenute le regole decisionali, sia minime che “forti”, si effettua la validazione.

3.3.4 Validazione

La validazione delle regole decisionali avviene tramite un procedimento deduttivo. Gli algoritmi di ROSE2 operano quella che è una classificazione sugli oggetti in esame utilizzando le regole decisionali indotte tramite il test *k-fold cross validation* a 10 passaggi. I risultati vengono poi confrontati con i valori reali della variabile decisionale d e sorgono così due possibili tipi di errore:

- Errore del tipo I: Aziende fallite classificate come attive
- Errore del tipo II: Aziende attive classificate come fallite

Nel dettaglio in tabella 10 sono raffigurati i risultati della validazione delle regole minime e delle regole forti (le percentuali esprimono il valore medio della precisione all'interno di un intervallo).

Tabella 10 – Errori di classificazione del I e II tipo dal reduct3

Classificazione in base alle regole minime				Classificazione in base alle regole forti			
	Fallite	Attive	Nessun a		Fallite	Attive	Nessun a
Aziende Fallite	32	2	1	Aziende Fallite	31	4	0
Aziende Attive	0	33	2	Aziende Attive	0	35	0
	Corrett o	Incorrett o	Nessun o		Corrett o	Incorrett o	Nessun o
Totale	92,86%	2,86%	4,29%	Totale	94,29%	5,71%	0,00%
Aziende Fallite	90,83%	5,83%	3,33%	Aziende Fallite	87,17%	13,33%	0,00%
Aziende Attive	95,00%	0,00%	5,00%	Aziende Attive	100,00 %	0,00%	0,00%

Elaborazione personale da Dataset AIDA

Come si vede nella tabella 10, le regole minime portano ad una precisione di classificazione totale inferiore alle regole forti. Tuttavia andando a guardare nel dettaglio gli errori del I e del II tipo, si può vedere come le aziende fallite vengano riconosciute con maggior precisione con l'utilizzo delle regole minime; al contrario le aziende attive vengono classificate con perfezione con le regole forti. Entrambi questi set di regole sono considerabili solo parzialmente discriminanti e vanno quindi intesi come tali; possono fornire un'indicazione sull'appartenenza delle aziende a uno dei due gruppi, ma mantenendo un margine di errore non del tutto trascurabile.

Al fine di completezza, è stato selezionato per la generazione delle regole un secondo *Reduct*, il numero 1, composto dagli attributi *a6*, *a7*, *a10*, *a11*, *a12*. Questo è stato scelto principalmente per due motivi: rappresenta il *Reduct* con lunghezza maggiore, pari a 5, ed è composto da variabili diverse da quelle del *Reduct* precedente, escluso *a10*. In tal modo è possibile evidenziare la capacità discriminante dalle regole generate da entrambi i *Reduct*, andando a cogliere le differenze in base agli attributi utilizzati. Le regole così generate vengono esposte nella tabella 11 per quanto riguarda le regole minime, e nella tabella 12

per le regole forti; i criteri di induzione sono stati mantenuti uguali a quelli utilizzati nel Reduct numero 3.

Tabella 11 – regole decisionali minime indotte da reduct1

# Regola	Condizioni					Decisione	
	a6	a7	a10	a11	a12	D	Forza
1		3-1				0	31
2					1	0	1
3	2					0	1
4			3	0		0	1
5	1				3	0	2
6	3		2			1	19
7			2		0	1	6
8		0				1	2
9				2	2	1	2
10				1		1	1
11		2		2	0	1	8

Elaborazione personale da Dataset AIDA

Tabella 12 – Regole decisionali forti indotte da reduct1

# Regola	Condizioni					Decisione	
	a6	a7	a10	a11	a12	D	Forza
1	0					0	23
2		3				0	28
3			0			0	25
4				3		0	25
5					1	0	4
6					3	0	15
7	3					1	25
8		0				1	17
9			2			1	27
10				1		1	9
11		2			2	1	8
12		2			0	1	9
13				2	2	1	8

Elaborazione personale da Dataset AIDA

Tramite lo stesso processo di validazione sopra esposto, si ottengono i seguenti risultati di classificazione del test *k-fold*:

Tabella 13 – Errori di classificazione del I e II tipo da *reduct1*

Classificazione in base alle regole minime				Classificazione in base alle regole forti			
	Fallite	Attive	Nessuna		Fallite	Attive	Nessuna
Aziende Fallite	31	4	0	Aziende Fallite	28	7	0
Aziende Attive	2	33	0	Aziende Attive	0	35	0
	Corretto	Incorretto	Nessuno		Corretto	Incorretto	Nessuno
Totale	91,43%	8,57%	0,00%	Totale	90,00%	10,00%	0,00%
Aziende Fallite	89,17%	10,83%	0,00%	Aziende Fallite	80,00%	20,00%	0,00%
Aziende Attive	94,17%	5,83%	0,00%	Aziende Attive	100,00%	0,00%	0,00%

Elaborazione personale da Dataset AIDA

È interessante notare dalla tabella 13 come entrambi i set di regole presentino una precisione inferiore rispetto alle regole generate dal *Reduct* numero 3, in particolar modo il set di regole forti si dimostra inappropriato nel riconoscere le aziende fallite.

Dall'osservazione dei risultati, risulta come la presenza di indicatori di redditività tra gli attributi (*Reduct* numero 3) semplifichi il lavoro degli algoritmi nella determinazione delle regole, rendendole maggiormente discriminanti. Per questo motivo viene scelto di utilizzare il *Reduct* 3 per procedere all'ultima parte di analisi. Lo scopo è quello di esaminare la qualità della classificazione tra aziende fallite e attive in un arco temporale che va dall'anno prima del fallimento fino a 4 anni prima, ovvero gli anni -1, -2, -3, -4. Purtroppo ROSE2 non permette di classificare un set di oggetti se non con le regole indotte dal set stesso, quindi non è possibile applicare le regole del *Reduct* 3 riferite all'anno -1 agli altri insiemi di dati. Per questo motivo l'analisi che viene condotta ora è eseguita per ogni anno con delle regole

decisionali diverse, ricavate sempre dal *Reduct a1, a2, a8, a10* ma su di una base di dati differente. Per le aziende fallite vengono usati i dati relativi agli anni precedenti al fallimento, in particolare fino al quarto anno prima. Per le aziende attive invece, viene usato lo stesso campione corrispondente all'anno -1. Questo perché non si vuole rischiare di andare oltre avventurandosi al di là delle proprie competenze in un'analisi sulle dinamiche temporali. Lo scopo quindi è quello di verificare la capacità di creare regole discriminanti tra aziende fallite e attive negli anni prima che queste dichiarassero fallimento. I risultati sono rappresentati nella tabella 14 e 15.

Tabella 14 – Test di classificazione delle regole forti da reduct3

Classificazione in base alle regole forti				
	Anno -1	Anno -2	Anno -3	Anno -4
Totale	94,29%	90,00%	75,71%	64,29%
Aziende Fallite	87,17%	83,33%	86,83%	77,57%
Aziende Attive	100,00%	97,50%	66,00%	49,17%

Elaborazione personale da Dataset AIDA

Tabella 15 – Test di classificazione delle regole minime da reduct3

Classificazione in base alle regole minime				
	Anno -1	Anno -2	Anno -3	Anno -4
Totale	92,86%	94,29%	94,29%	84,29%
Aziende Fallite	90,83%	96,67%	90,83%	79,17%
Aziende Attive	95,00%	92,50%	97,50%	89,17%

Elaborazione personale da Dataset AIDA

Come si nota, il risultato ottenuto dalla classificazione sulla base delle regole minime è decisamente più interessante di quello ottenuto dalle regole forti. Tralasciando l'anno -1 del quale si è già parlato prima, si analizzano i risultati nei restanti anni. Nella classificazione in base alle regole forti si nota un calo generale della qualità delle regole; è facilmente spiegabile dal fatto che utilizzando regole incerte, le approssimazioni tendono ad essere sempre maggiori allontanandosi dall'anno del fallimento. Ci si aspetta che un'azienda fallita nell'anno 0 possa presentare nell'anno -3 o -4 dei dati di bilancio non necessariamente indicanti il prossimo fallimento. Proprio per questo l'induzione di regole incerte e parzialmente discriminanti porta ad un'erronea classificazione. Per quanto riguarda la classificazione adoperata tramite le regole minime, si registra invece una buona qualità nell'induzione delle regole fino all'anno -3. A differenza della classificazione precedente le regole utilizzate sono solamente quelle certe; si verifica addirittura una maggior precisione nell'anno -2 rispetto all'anno -1. La capacità discriminatoria perde efficacia nell'anno -4 come è facile attendersi.

Si tiene a ricordare che la classificazione per ogni anno avviene tramite l'utilizzo di *set* di regole diverse; quelle esposte in questa Tesi nelle tabelle 8, 9, 11 e 12 fanno unicamente riferimento al set dell'anno -1. I dati esposti sopra dunque non intendono arrogarsi alcuna capacità predittiva, bensì fungono a dimostrazione di come, dati due insiemi di aziende fallite e attive di cui si conosce a priori lo stato Giuridico, gli algoritmi alla base dell'induzione delle regole della Rough-Set Theory riescano a generare delle regole in grado di discriminare al livello del 94,29% anche sui dati precedenti 3 anni al fallimento. Oltre a questa distanza, i dati dei due gruppi diventano troppo omogenei e non è più possibile un'induzione efficace delle regole, né di tipo minimo né forti.

Conclusione

I risultati ottenuti in questa Tesi dimostrano come l'utilizzo della teoria dei Rough-Sets possa condurre a dei buoni risultati di classificazione delle società nel panorama degli studi sul fallimento aziendale. In particolare la metodologia appare particolarmente indicata per questo tipo di analisi poiché si concentra su quel tipo di dati incerti caratteristici dei bilanci delle aziende. Questa analisi è stata condotta in maniera differente da ciò che è stato riscontrato in letteratura sotto due principali aspetti: innanzitutto l'intervallo temporale tramite il quale sono state selezionate le aziende è stato posto tra il 2011 ed il 2015, quindi successivamente alla recente crisi finanziaria. Tale scelta è stata apportata al fine di indagare l'effettiva validità dell'analisi Rough-Set anche nei confronti di dati ulteriormente imprevedibili quali quelli di aziende uscenti dal periodo di crisi globale. In seconda istanza l'analisi di classificazione è stata eseguita accoppiando ai dati di bilancio delle aziende fallite degli anni -1, -2, -3, -4 unicamente i dati delle aziende attive nell'ultimo anno di bilancio disponibile. Questo è stato fatto per concentrare l'attenzione sulle aziende fallite, senza dover considerare possibili variazioni dei dati nelle aziende attive; inoltre prendendo l'ultimo bilancio disponibile delle aziende attive, si evita di analizzare dati eccessivamente imprecisi e fuorvianti derivanti da bilanci redatti in concomitanza alla crisi finanziaria, che potrebbero far apparire le società attive troppo simili a quelle fallite.

I risultati mostrano come tramite l'applicazione del set di regole forti e parzialmente discriminanti (incerte) sia possibile ottenere una precisione totale media di classificazione del 94,29% nell'anno -1 e del 90,00% nell'anno -2, andando poi a perdere capacità discriminanti nell'anno -3 attestandosi al 75,71%. A sua volta il set di regole minime e certe presenta risultati comparabili nei primi due anni, mantenendo però una buona precisione anche nell'anno -3 attestandosi al 94,29% e al 84,29% nell'anno -4.

Una delle prime limitazioni di questo lavoro è l'impossibilità di applicare i set di regole ad un *testing sample* al fine di esaminarne la capacità predittiva. La causa è la limitazione del

software ROSE2, il quale conclude il proprio campo di applicazione all'induzione delle regole e alla classificazione dei set secondo le regole estratte, senza poter applicare queste ad altri set di dati. Per quanto riguarda le limitazioni dell'analisi, si considera la scelta operata in riguardo alla discretizzazione degli attributi: dalle due modalità utilizzate, si evince come questo *step* abbia importanti ripercussioni sui risultati finali; non essendovi un modello generale da seguire, in questa tesi è stato scelto di utilizzare la funzione di discretizzazione offerta da ROSE2; utilizzando una diversa norma di discretizzazione, i risultati possono essere profondamente differenti. Un'altra importante limitazione sta nell'ampiezza del campione: si è qui scelto di seguire quel ramo di letteratura che predilige un campione di ridotte dimensioni e avente al suo interno in egual misura aziende attive e fallite. Nulla garantisce che questo sia il metodo migliore, infatti guardando a possibili sviluppi futuri del lavoro sarebbe interessante raccogliere un campione di maggiori dimensioni in modo tale da comprendere maggiori aspetti relativi al fallimento delle aziende e poter così generare, si pensa, un maggior numero di regole decisionali. Infine per quanto riguarda le variabili utilizzate, si ritiene che queste siano adatte allo scopo, come visto nell'analisi descrittiva eseguita e confermato dalla letteratura, dove ne viene fatto un uso massiccio. Si conclude ribadendo l'effettiva efficacia del metodo qui proposto, sia in merito ai risultati osservabili nella letteratura, sia per i risultati raggiunti in questo lavoro.

Bibliografia

- E. I. ALTMAN, *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, in "The Journal of finance", 4, 1968, pp. 589-609.
- M. A. AZIZ, H. A. DAR, *Predicting corporate bankruptcy: where we stand*, in "Corporate Governance: The international journal of business in society", 6, 2006, pp. 18-33.
- S. BALCAENA, H. OOGHE, *35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems*, in "The British Accounting Review", 38, 2006, pp. 63-93.
- W. BEAVER, *Financial ratios as predictors of failure*, in "Journal of Accounting Research", 4, 1967, pp. 71-111
- W. BEAVER, *Alternative accounting measure predictors of failure*, in "The accounting review", 43, 1968, pp. 113-122.
- M. BLUM, *Failing company discriminant analysis*, in "Journal of Accounting Research", 12, 1974, pp. 1-25.
- J.E. BORITZ, D.B. KENNEDY, *Effectiveness of neural network types for prediction of business failure*, in "Expert Systems With Applications", 9, 1995, pp. 503-512.
- E.B. DEAKIN, *Discriminant analysis of predictors of business failure*, Wiley on behalf of Accounting Research Center, Booth School of Business, University of Chicago, in "Journal of Accounting Research", 10, 1972, pp 167-179.
- A.I. DIMITRAS, R. SLOWINSKI, R. SUSMAGA, C. ZOPOUNIDIS, *Business failure prediction using rough sets*, in "European Journal of Operational Research", 114, 1999, pp. 263-280.
- E. K. LAITINEN, T. LAITINEN, *Cash management behavior and failure prediction*, in "Journal of Business Finance&Accounting", 25, 1998, 893-919.
- H. LI, J. SUN, *Majority voting combination of multiple case-based reasoning for financial distress prediction*, in "Expert Systems with Applications", 36, 2009, pp. 4363-4373.
- G. S. MADDALA, *A perspective on the use of limited-dependent and qualitative variables models in accounting research*, in "The Accounting Review", 66, 1991, pp. 788-807.
- D. MOSES, S.S. LIAO, *On developing models for failure prediction*, in "Journal of Commercial Bank Lending", 69, 1987, pp. 27-38.
- J. A. OHLSON, *Financial Ratios and the probabilistic prediction of bankruptcy*, in "Journal of Accounting Research", 18, 1980, pp. 109-131.
- Z. PAWLAK, *Rough-sets*, in "International Journal of Computer and Information Sciences", 11, 1982, pp. 341-356.
- Z. PAWLAK., *Rough-sets. Theoretical Aspects of Reasoning about Data*, Kluwer, Dordrecht, 1991.
- Z. PAWLAK, A. Skowron, *Rudiments of rough sets*, in "Information Sciences", 177, 2007, pp. 3-27.

B. PREDKI, R. SLOWINSKI, J.STEFANOWSKI, R.SUSMAGA, SZ.WILK, *ROSE - Software Implementation of the Rough Set Theory*, in: L.Polkowski, A.Skowron, eds. *Rough Sets and Current Trends in Computing*, Lecture Notes in Artificial Intelligence, 1424, 1998, pp. 605-608.

B. PREDKI, SZ. WILK, *Rough Set Based Data Exploration Using ROSE System*, in: Z.W.Ras, A.Skowron, eds. *Foundations of Intelligent Systems*, Lecture Notes in Artificial Intelligence, 1609, 1999, pp. 172-180

J. SCOTT, *The probability of bankruptcy a comparison of empirical predictions and theoretical models*, in "Journal of Banking and Finance", 5, 1981, pp. 317-344.

M. TAMARI, *Financial Ratios as a Means of Forecasting Bankruptcy*, in "Management International Review", 6, 1966, pp. 15-21.

L. TOMASELLA, *La teoria dei rough-set per la modellazione della scelta modale*, Tesi in dottorato di ricerca in ingegneria delle infrastrutture, delle strutture e dei trasporti, 2008

J. B. WARNER, *Bankruptcy Costs: Some Evidence*, in "The Journal of Finance", 32, 1977, pp. 337-347.

Elenco delle Formule

- (1) Funzione Discriminante del modello MDA lineare
- (2) Funzione dello *z-score* ottenuta da Altman (1968)
- (3) Funzione Logit
- (4) Relazione di Indiscernibilità tra gli oggetti nella Rough-Set Theory
- (5) Approssimazione inferiore del Rough-Set
- (6) Approssimazione superiore del Rough-Set
- (7) Regione dell'incertezza del Rough-Set
- (8) Il Rough-Set
- (9) Accuratezza di approssimazione del set
- (10) Qualità di approssimazione del set
- (11) Probabilità di un oggetto di appartenere al set
- (12) Core del set ridotto

Elenco delle figure e tabelle

Figura 1) Approssimazione dei Rough-Sets

Tabella 1) Statistiche di sintesi aziende fallite

Tabella 2) Statistiche di sintesi aziende attive

Tabella 3) Correlazioni aziende fallite

Tabella 4) Independent sample t-test

Tabella 5) I reduct

Tabella 6) Frequenza degli attributi nel reduct

Tabella 7) Tabella di informazione ridotta

Tabella 8) Regole minime indotte dal reduct3

Tabella 9) Regole forti indotte dal reduct3

Tabella 10) Errori di classificazione del I e II tipo dalle regole del reduct3

Tabella 11) Regole minime indotte dal reduct1

Tabella 12) Regole forti indotte dal reduct1

Tabella 13) Errori di classificazione del I e II tipo da reduct1

Tabella 14) Test di classificazione delle regole forti da reduct3

Tabella 15) Test di classificazione delle regole minime da reduct3

Appendice

Dati grezzi delle aziende fallite e attive nell'anno -1 dal fallimento

d	0	0	0	0	0	0	0	0	0
a12	0,03	0,88	0,90	1,00	0,86	0,66	1,00	1,21	
a11	0,43	0,59	1,55	2,75	1,59	1,04	1,44	1,33	
a10	30,32	11,35	-0,20	-0,03	-0,02	-0,49	-0,02	0,00	
a9	0,28	0,09	-3,92	-4,31	-2,18	-2,02	-2,45	-2,45	
a8	0,32	0,11	1,22	1,04	1,47	1,52	1,16	1,00	
a7	0,76	0,85	1,71	2,83	2,06	1,14	1,54	1,33	
a6	0,71	0,53	0,37	0,14	0,17	0,72	0,47	0,63	
a5	1,01	1,05	0,47	0,17	0,19	0,81	0,48	0,63	
a4	-0,52	-2,45	1,54	0,63	0,34	1,48	0,73	1,02	
a3	-0,08	-0,08	-1,40	-1,45	-0,50	-0,44	-0,54	-0,42	
a2	-0,05	-0,03	-1,32	-1,33	-0,30	-0,30	-0,51	-0,37	
a1	1,49	2,43	1,06	1,09	1,65	1,47	1,06	1,14	
Oggetto	x1	x2	x3	x4	x5	x6	x7	x8	

d	0	0	0	0	0	0	0	0	0	0	0	0	0
a12	1,52	0,65	9,56	2,03	2,08	2,19	1,50	1,87	1,21				
a11	1,08	0,52	0,82	1,22	0,55	1,48	0,98	0,76	0,30				
a10	-0,22	0,07	-0,13	-0,05	0,25	-0,05	-0,64	2,43	1,18				
a9	-1,21	0,68	-0,10	-0,65	8,44	-0,80	-0,66	0,30	0,57				
a8	2,66	0,54	-0,57	1,85	0,61	1,00	7,13	0,38	0,37				
a7	1,35	0,70	0,98	1,45	0,68	1,48	1,30	0,83	0,70				
a6	0,25	1,25	0,27	0,14	1,55	0,21	0,09	0,84	0,82				
a5	0,58	1,28	0,47	0,19	1,74	0,25	0,45	1,12	1,95				
a4	1,02	0,21	2,34	0,96	-0,54	1,41	1,96	-5,13	0,01				
a3	-0,45	0,04	-0,13	-0,47	-0,10	-0,72	-0,72	-0,24	0,00				
a2	-0,31	0,09	-0,15	-0,45	-0,08	-0,48	-0,58	-0,21	0,02				
a1	1,43	0,50	0,89	1,05	1,30	1,50	1,24	1,13	0,07				
Oggetto	x9	x10	x11	x12	x13	x14	x15	x16	x17				

d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a12	9,45	1,45	0,25	1,27	1,10	1,51	0,85	1,65	1,11									
a11	0,88	0,55	0,67	2,14	1,34	1,11	1,15	1,29	5,71									
a10	-0,12	-0,68	17,58	-0,01	-0,27	-0,14	-0,47	-0,07	-0,06									
a9	-0,07	-0,24	0,29	-2,70	-2,18	-0,83	-5,29	-0,82	-11,26									
a8	1,59	-0,49	0,29	1,09	1,17	2,79	1,00	1,86	1,00									
a7	0,89	0,96	0,88	2,26	1,41	1,40	1,15	1,60	5,71									
a6	0,65	0,39	0,45	0,18	0,45	0,30	0,75	0,09	0,02									
a5	0,69	0,77	1,03	0,21	0,56	0,41	0,83	0,14	0,07									
a4	10,20	2,34	-0,42	1,03	2,07	1,92	1,23	1,98	0,32									
a3	-0,30	-0,32	-0,04	-1,41	-1,10	-0,87	-0,28	-1,32	-1,53									
a2	-0,25	-0,25	-0,01	-1,19	-0,94	-0,78	-0,18	-0,81	-1,15									
a1	1,18	1,30	3,58	1,18	1,17	1,12	1,54	1,63	1,33									
Oggetto	x18	x19	x20	x21	x22	x23	x24	x25	x26									

d	0	0	0	0	0	0	0	0	0	0	0	0	0
a12	1,09	3,29	1,39	0,17	2,28	1,65	0,61	1,02	1,09	1,02	1,09	1,09	1,09
a11	3,67	0,61	1,87	0,89	1,02	0,95	1,31	4,20	1,07	4,20	1,07	1,07	1,07
a10	-0,06	1,87	-0,23	10,12	-0,52	-0,50	-1,09	-0,02	-2,87	-0,02	-2,87	-2,87	-2,87
a9	-6,46	3,28	-8,70	1,05	-0,49	-0,57	-11,41	-5,96	-2,04	-5,96	-2,04	-2,04	-2,04
a8	1,00	0,37	1,03	1,00	2,62	2,11	1,13	1,10	1,15	1,10	1,15	1,15	1,15
a7	3,67	0,79	1,90	0,89	1,15	1,09	1,38	4,55	1,09	4,55	1,09	1,09	1,09
a6	0,10	0,50	0,12	0,83	0,29	0,31	0,40	0,05	0,42	0,05	0,42	0,42	0,42
a5	0,15	1,58	0,33	1,02	0,53	0,54	0,72	0,08	0,85	0,08	0,85	0,85	0,85
a4	0,75	0,01	1,22	0,30	1,22	1,36	1,16	0,20	1,26	0,20	1,26	1,26	1,26
a3	-2,15	0,00	-1,17	0,03	-0,26	-0,36	-0,71	-0,75	-0,19	-0,75	-0,19	-0,19	-0,19
a2	-2,11	0,03	-1,10	0,07	-0,18	-0,33	-0,55	-0,72	-0,16	-0,72	-0,16	-0,16	-0,16
a1	1,02	0,05	1,07	0,39	1,44	1,09	1,30	1,03	1,14	1,03	1,14	1,14	1,14
Oggetto	x27	x28	x29	x30	x31	x32	x33	x34	x35	x34	x35	x35	x35

d	1	1	1	1	1	1	1	1	1	1	1	1
a12	1,23	1,00	0,66	0,07	1,26	0,22	0,40	0,34	0,04			
a11	0,51	0,20	0,46	0,55	0,33	0,58	0,54	0,49	0,47			
a10	0,41	0,06	1,64	16,70	0,00	0,58	1,16	3,14	11,72			
a9	0,78	8,64	1,08	0,47	1,76	0,50	1,08	0,57	1,04			
a8	0,43	1,00	0,76	0,44	0,78	0,52	0,83	0,57	0,88			
a7	0,77	0,20	0,56	0,80	0,43	0,75	0,58	0,67	0,53			
a6	1,27	4,45	0,70	0,62	2,32	1,03	0,97	0,65	0,61			
a5	1,46	4,67	1,47	1,02	2,32	1,07	1,17	1,16	1,04			
a4	0,04	0,23	0,28	0,37	0,13	0,05	0,08	0,07	0,02			
a3	0,01	0,16	0,09	0,07	0,05	0,01	0,02	0,02	0,01			
a2	0,03	0,22	0,14	0,14	0,08	0,03	0,04	0,06	0,03			
a1	0,28	0,73	0,66	0,48	0,58	0,38	0,41	0,28	0,34			
Oggetto	x36	x37	x38	x39	x40	x41	x42	x43	x44			

d	1	1	1	1	1	1	1	1	1	1	1	1
a12	0,52	1,18	1,08	1,20	0,97	0,79	0,75	0,26	2,88			
a11	0,58	0,70	0,60	0,70	0,78	0,29	0,61	0,53	0,56			
a10	0,26	0,62	0,48	1,85	2,05	0,11	0,06	1,54	1,13			
a9	0,40	4,51	4,53	1,51	1,80	5,23	1,94	0,47	1,47			
a8	0,44	0,75	0,92	0,72	0,57	0,99	1,00	0,47	0,36			
a7	0,74	0,77	0,62	0,76	0,85	0,29	0,61	0,75	0,78			
a6	1,08	1,14	1,29	0,77	0,89	2,47	1,35	0,95	0,92			
a5	1,11	1,36	1,56	1,27	1,11	2,65	1,38	1,10	1,60			
a4	0,23	0,07	0,81	0,36	0,12	0,25	0,22	0,04	0,09			
a3	0,03	0,01	0,25	0,06	0,01	0,15	0,07	0,01	0,01			
a2	0,07	0,04	0,38	0,08	0,06	0,21	0,11	0,04	0,02			
a1	0,39	0,35	0,65	0,70	0,18	0,70	0,59	0,21	0,43			
Objetto	x45	x46	x47	x48	x49	x50	x51	x52	x53			

d	1	1	1	1	1	1	1	1	1	1	1
a12	1,31	1,10	1,00	0,93	1,31	1,27	0,92	0,82	0,07		
a11	0,43	0,67	0,24	0,60	0,50	0,59	0,72	0,26	0,49		
a10	0,60	0,47	0,27	1,38	0,26	0,49	2,28	0,33	5,71		
a9	2,55	0,74	5,51	0,38	12,38	0,95	6,61	17,41	0,63		
a8	0,64	0,54	0,97	0,33	1,00	0,45	1,00	1,00	0,66		
a7	0,61	0,78	0,26	0,82	0,50	0,80	0,72	0,26	0,64		
a6	1,38	1,11	2,94	0,94	1,70	1,20	0,59	2,37	0,79		
a5	1,96	1,21	3,65	1,17	1,95	1,38	1,32	3,03	1,05		
a4	0,04	0,09	0,27	0,36	0,50	0,30	0,08	0,04	0,18		
a3	0,01	0,01	0,17	0,04	0,18	0,05	0,02	0,02	0,05		
a2	0,03	0,05	0,25	0,10	0,25	0,08	0,03	0,06	0,10		
a1	0,46	0,22	0,69	0,40	0,71	0,68	0,80	0,41	0,58		
Objetto	x54	x55	x56	x57	x58	x59	x60	x61	x62		

d	1	1	1	1	1	1	1	1	1
a12	0,63	0,76	0,73	0,67	0,11	0,73	0,72	0,94	
a11	0,83	0,51	0,71	0,72	0,51	0,76	0,42	0,45	
a10	5,80	0,85	2,57	0,22	0,00	0,33	2,04	0,84	
a9	0,92	3,24	1,33	1,34	0,96	1,31	0,74	2,51	
a8	1,00	1,00	0,84	0,93	1,00	0,81	0,49	0,82	
a7	0,83	0,51	0,74	0,74	0,51	0,80	0,71	0,54	
a6	0,66	1,10	0,71	1,14	1,09	1,10	0,50	1,14	
a5	1,07	1,66	1,19	1,17	1,09	1,15	1,48	1,85	
a4	0,14	0,08	0,16	0,44	0,35	0,24	0,04	0,02	
a3	0,01	0,04	0,03	0,08	0,14	0,04	0,01	0,01	
a2	0,04	0,04	0,05	0,12	0,05	0,06	0,06	0,00	
a1	0,32	0,85	0,55	0,72	2,80	0,63	0,17	3,29	
Objetto	x63	x64	x65	x66	x67	x68	x69	x70	

Dati grezzi delle aziende fallite nell'anno -2 dal fallimento

a12	0,45	-0,35	-0,66	1,46	0,89	2,43	-0,88	-26,82
a11	0,44	0,58	0,59	1,14	1,19	0,68	0,71	0,88
a10	1,25	-5,90	-0,91	-0,38	-0,04	0,42	-1,26	-0,57
a9	0,43	0,27	0,41	-1,16	-1,46	0,93	0,25	0,03
a8	0,40	0,36	0,72	1,17	2,35	0,36	0,55	0,15
a7	0,73	0,78	0,68	1,18	1,76	0,83	0,78	0,92
a6	0,95	0,54	0,53	0,47	0,24	1,18	0,76	0,68
a5	1,20	0,93	0,75	0,62	0,26	1,31	0,89	0,84
a4	0,01	0,05	0,01	1,58	1,34	-0,51	0,02	2,96
a3	0,00	0,01	0,00	-0,48	-1,33	-0,04	0,00	0,02
a2	0,02	0,06	0,03	-0,38	-0,62	-0,02	0,04	0,10
a1	0,07	0,08	0,04	1,26	2,15	2,00	0,03	0,21
Oggetto	x1	x2	x3	x4	x5	x6	x7	x8

a12	49,38	-6,46	-4,76	0,57	1,31	-0,75	0,18	-0,45	0,20
a11	0,78	0,76	0,64	0,65	0,60	0,58	0,53	0,77	0,49
a10	-0,68	-0,20	-0,22	-0,24	0,32	-1,48	7,88	-8,78	5,38
a9	-0,01	0,08	0,13	-0,63	13,18	0,33	0,58	0,28	0,58
a8	-0,03	0,21	0,24	3,75	1,00	0,44	0,51	0,62	0,62
a7	0,97	0,88	0,87	0,79	0,60	0,80	0,74	0,81	0,66
a6	0,41	0,67	0,32	0,79	1,41	0,45	0,49	0,65	0,54
a5	0,70	0,73	0,58	1,06	1,62	0,80	1,18	0,97	1,11
a4	4,80	-1,17	-1,42	2,45	0,12	-0,13	0,06	0,28	0,01
a3	-0,03	-0,04	-0,11	-0,47	0,03	-0,02	0,01	0,02	0,00
a2	-0,03	0,00	-0,11	-0,45	0,07	0,04	-0,02	0,06	0,02
a1	0,95	-20,28	1,01	1,06	0,47	-0,52	-0,62	0,28	0,10
Oggetto	x9	x10	x11	x12	x13	x14	x15	x16	x17

a12	0,02	-0,36	-0,36	-19,51	-0,18	1,05	1,69	-0,34	1,23
a11	0,66	0,38	0,76	0,64	0,59	0,48	0,89	0,33	1,75
a10	7,50	-3,63	-7,99	-0,18	-2,01	1,56	2,34	-1,37	-0,26
a9	0,64	0,18	0,38	0,03	0,77	0,75	1,97	0,37	-4,03
a8	0,94	0,18	0,47	0,06	0,96	0,44	1,00	0,43	1,00
a7	0,67	0,81	0,87	0,93	0,61	0,75	0,89	0,69	1,75
a6	0,95	0,58	0,57	0,40	0,70	0,73	0,88	0,32	0,28
a5	1,01	1,17	0,95	0,54	0,90	1,48	1,09	0,72	0,43
a4	-0,31	-1,83	0,05	-0,57	-0,19	-0,68	0,11	-0,30	0,68
a3	-0,06	-0,18	0,01	-0,01	-0,06	-0,15	0,00	-0,08	-0,56
a2	-0,04	-0,22	0,00	0,03	-0,06	-0,13	0,13	-0,06	-0,44
a1	1,52	0,78	1,35	-0,30	1,14	1,13	0,04	1,40	1,26
Objetto	x18	x19	x20	x21	x22	x23	x24	x25	x26

a12	2,16	3,74	-0,78	0,29	-5,38	-2,30	1,14	0,98	1,25
a11	1,18	0,54	0,58	0,90	0,88	0,77	0,74	2,76	0,90
a10	-0,32	1,68	-2,69	5,27	-1,32	-1,19	4,25	-0,04	17,30
a9	-0,71	2,59	1,40	1,12	0,13	0,20	1,09	-3,84	0,52
a8	1,06	0,36	0,37	0,98	0,55	0,39	0,46	1,29	0,44
a7	1,19	0,74	0,83	0,90	0,92	0,89	0,84	3,33	0,94
a6	0,37	0,49	0,30	0,88	0,36	0,48	0,57	0,06	0,39
a5	0,52	1,74	0,88	1,03	0,72	0,77	1,19	0,10	1,04
a4	1,53	0,01	0,02	0,07	0,60	-0,02	-0,01	1,13	0,02
a3	-0,40	0,00	0,00	0,01	0,03	0,00	0,00	-2,87	0,00
a2	-0,37	0,03	0,03	0,08	0,05	0,03	0,00	-1,02	0,02
a1	1,09	0,04	0,08	0,08	0,49	-0,06	-0,90	2,81	0,02
Objetto	x27	x28	x29	x30	x31	x32	x33	x34	x35

Dati grezzi delle aziende fallite nell'anno -3 dal fallimento

a12	0,55	-0,41	-0,72	-0,56	-1,74	1,08	-3,41	4,19
a11	0,42	0,54	0,61	0,90	0,62	0,69	0,76	0,91
a10	0,93	-0,52	-0,70	-23,08	-0,15	0,18	-0,43	-0,88
a9	0,43	0,33	0,43	0,35	0,30	1,09	0,15	-0,14
a8	0,39	0,47	0,85	0,59	0,66	0,47	0,60	-1,69
a7	0,73	0,71	0,65	0,92	0,71	0,81	0,81	0,96
a6	1,02	0,83	0,52	0,59	0,44	1,13	0,59	0,75
a5	1,26	0,89	0,72	0,98	0,51	1,16	0,73	0,87
a4	0,01	0,16	0,00	0,02	0,00	-0,15	0,10	1,27
a3	0,00	0,02	0,00	0,00	0,00	-0,02	0,01	-0,04
a2	0,05	0,11	0,03	0,03	0,03	-0,02	0,08	0,04
a1	0,02	0,22	0,01	0,02	0,01	0,80	0,08	-0,90
Oggetto	x1	x2	x3	x4	x5	x6	x7	x8

a12	-10,46	-2,63	-1,11	-0,11	1,22	-0,61	0,53	1,61	0,07
a11	0,75	0,73	0,50	0,51	0,57	0,54	0,38	0,85	0,54
a10	-0,86	-0,32	-0,37	-2,21	0,48	-1,60	2,75	4,18	12,18
a9	0,05	0,17	0,34	0,88	9,36	0,35	0,39	0,91	0,61
a8	0,10	0,33	0,39	0,63	1,00	0,46	0,33	0,54	0,73
a7	0,95	0,86	0,78	0,67	0,58	0,77	0,77	0,89	0,64
a6	0,46	0,69	0,44	0,80	1,32	0,42	0,54	0,74	0,59
a5	0,71	0,78	0,80	1,32	1,65	0,79	1,36	1,08	1,04
a4	-2,09	0,03	-0,08	-0,33	0,01	0,00	0,16	0,19	0,00
a3	-0,04	0,00	-0,01	-0,09	0,00	0,00	0,03	0,01	0,00
a2	-0,04	0,07	0,00	-0,08	0,03	0,04	0,03	0,04	0,01
a1	1,01	0,03	7,03	1,18	0,05	0,01	1,03	0,20	0,11
Objetto	x9	x10	x11	x12	x13	x14	x15	x16	x17

a12	0,27	0,36	0,51	-12,16	0,11	0,86	1,62	-0,10	1,66
a11	0,65	0,30	0,69	0,62	0,54	0,44	0,90	0,28	1,12
a10	0,50	1,14	4,90	-0,19	2,58	1,44	1,51	-3,46	-1,35
a9	0,83	0,43	0,52	0,04	1,00	1,13	1,74	0,49	-1,24
a8	0,90	0,39	0,49	0,08	0,88	0,60	1,00	0,52	1,00
a7	0,68	0,67	0,81	0,94	0,59	0,66	0,90	0,62	1,12
a6	1,05	0,96	0,65	0,38	0,88	0,73	0,96	0,40	0,43
a5	1,10	1,42	1,09	0,53	1,07	1,62	1,07	0,87	0,76
a4	0,00	-0,34	0,02	-0,24	0,00	0,00	0,28	-0,04	1,11
a3	0,00	-0,08	0,00	-0,01	0,00	0,00	0,01	-0,02	-0,18
a2	0,03	-0,17	0,02	0,02	0,05	0,03	0,11	0,00	-0,03
a1	0,00	0,48	0,08	-0,26	0,02	0,03	0,10	4,76	6,07
Objetto	x18	x19	x20	x21	x22	x23	x24	x25	x26

a12	-0,91	3,81	-2,25	0,37	-7,00	-2,35	-5,71	-0,96	2,13
a11	0,86	0,51	0,78	0,77	0,81	0,75	0,81	0,57	0,82
a10	-4,30	1,47	-1,04	3,83	-1,58	-1,09	6,95	-0,82	5,37
a9	0,41	1,83	1,22	0,94	0,09	0,20	-0,13	0,30	0,58
a8	0,81	0,33	0,70	0,74	0,22	0,41	-0,13	0,42	0,31
a7	0,88	0,73	0,83	0,83	0,92	0,87	0,92	0,79	0,91
a6	0,54	0,60	0,21	0,79	0,28	0,42	0,48	0,48	0,55
a5	0,92	1,84	0,63	1,08	0,72	0,73	1,10	0,72	1,10
a4	-0,07	0,01	0,07	0,07	-1,84	-0,09	9,69	0,07	-0,34
a3	-0,01	0,00	0,01	0,01	-0,06	-0,01	-0,12	0,01	-0,01
a2	0,01	0,02	0,04	0,05	-0,11	0,02	-0,12	0,06	0,01
a1	-0,40	0,05	0,22	0,21	0,53	-0,35	0,97	0,20	-1,45
Objetto	x27	x28	x29	x30	x31	x32	x33	x34	x35

Dati grezzi delle aziende fallite nell'anno -4 dal fallimento

a12	0,26	1,98	-3,30	0,68	-1,53	1,59	-0,59	-22,10
a11	0,41	0,39	0,71	0,86	0,55	0,69	0,64	0,93
a10	1,14	0,21	-0,59	16,94	-0,18	0,60	-0,71	-0,62
a9	0,59	0,58	0,12	0,28	0,30	1,22	0,49	0,03
a8	0,57	0,33	0,35	0,35	0,64	0,50	0,90	0,49
a7	0,63	0,68	0,82	0,91	0,67	0,80	0,67	0,94
a6	0,97	1,61	0,57	0,60	0,34	1,10	0,64	0,70
a5	1,19	1,76	0,73	1,03	0,44	1,26	0,81	0,82
a4	0,01	-0,20	-0,12	0,01	-0,08	0,05	-0,16	4,24
a3	0,00	-0,03	-0,01	0,00	-0,02	0,01	-0,04	0,03
a2	0,03	-0,02	0,00	0,04	0,02	0,05	0,06	0,11
a1	0,13	1,70	-2,81	0,01	-0,71	0,11	-0,65	0,29
Oggetto	x1	x2	x3	x4	x5	x6	x7	x8

a12	-6,22	-3,78	-0,65	0,15	1,10	-0,36	-0,73	2,81	0,05
a11	0,84	0,75	0,44	0,36	0,54	0,48	0,61	0,81	0,52
a10	-0,55	-0,13	-0,56	1,48	0,72	-2,30	-2,33	1,99	16,54
a9	0,11	0,14	0,42	1,11	7,69	0,36	0,33	0,69	0,60
a8	0,39	0,31	0,41	0,67	0,99	0,43	0,44	0,39	0,73
a7	0,92	0,87	0,75	0,56	0,55	0,75	0,81	0,87	0,62
a6	0,42	0,68	0,50	0,92	1,19	0,49	0,39	0,85	0,60
a5	0,63	0,73	0,94	1,74	1,68	0,85	0,87	1,15	1,03
a4	-3,40	0,01	0,01	-0,03	0,00	-0,11	0,26	0,13	0,00
a3	-0,17	0,00	0,00	-0,01	0,00	-0,02	0,04	0,01	0,00
a2	-0,13	0,07	0,01	0,01	0,01	0,01	-0,03	0,03	0,02
a1	1,37	0,01	0,24	-1,25	0,08	-4,08	-1,46	0,17	0,02
Oggetto	x9	x10	x11	x12	x13	x14	x15	x16	x17

a12	0,23	0,67	0,39	-8,67	0,25	0,84	0,05	0,00	-51,13
a11	0,67	0,32	0,74	0,58	0,45	0,39	0,91	0,30	0,96
a10	0,37	0,44	5,08	-0,19	0,93	1,09	36,51	96,73	-1,97
a9	0,90	0,52	0,80	0,05	1,03	1,04	0,85	0,59	0,02
a8	0,97	0,44	0,72	0,08	0,85	0,58	1,00	0,60	1,00
a7	0,68	0,62	0,80	0,94	0,53	0,63	0,91	0,57	0,96
a6	1,05	1,28	0,71	0,44	1,01	0,93	0,95	0,54	0,45
a5	1,08	1,60	1,07	0,58	1,22	1,74	1,00	1,00	0,81
a4	0,01	-0,34	-0,01	-0,26	0,02	0,02	0,41	0,00	-39,20
a3	0,00	-0,08	0,00	-0,01	0,01	0,01	0,01	0,00	-0,14
a2	0,04	-0,09	0,06	0,02	0,06	0,04	0,12	0,00	-0,07
a1	0,09	0,88	-0,02	-0,43	0,16	0,17	0,10	0,28	1,86
Objetto	x18	x19	x20	x21	x22	x23	x24	x25	x26

a12	-1,02	3,88	-1,27	0,53	-18,14	-2,09	-1,83	-1,54	0,90
a11	0,82	0,44	0,74	0,59	0,91	0,74	0,83	0,67	0,84
a10	-2,67	1,20	-1,39	0,38	-1,35	-1,07	8,45	-0,49	3,90
a9	0,38	1,25	1,43	1,16	0,05	0,23	-0,46	0,29	0,77
a8	0,76	0,31	0,75	0,74	0,23	0,48	-0,53	0,58	0,50
a7	0,85	0,71	0,80	0,69	0,95	0,85	0,93	0,79	0,90
a6	0,54	0,79	0,26	1,16	0,44	0,40	0,44	0,44	0,83
a5	0,89	2,07	0,69	1,26	0,76	0,71	1,10	0,64	1,06
a4	-0,20	0,00	0,07	0,03	0,25	0,00	3,55	0,08	0,22
a3	-0,02	0,00	0,01	0,01	0,00	0,00	-0,12	0,01	0,01
a2	-0,02	0,02	0,02	0,04	0,05	-0,01	-0,09	0,05	0,04
a1	1,34	0,02	0,50	0,19	0,07	-0,01	1,34	0,28	0,31
Oggetto	x27	x28	x29	x30	x31	x32	x33	x34	x35