

# Verifica di un'ipotesi d'impiego delle Reti Neurali per la previsione dell'Economic Value Added di un'impresa

*Corrado Boschioli, Damiano Carrara, Enrico Cavalli*

## 1. Introduzione

L'uso delle reti neurali (R.N. nel seguito) per prevedere le insolvenze aziendali attraverso l'analisi di bilancio ha dato esiti molto promettenti a livello sperimentale [Atiya 2001], [Cavalli 2001]. In particolare, un lavoro condotto dagli autori [Boschioli et al. 2001], ha prodotto risultati incoraggianti nella prospettiva di applicare questa analisi alla revisione periodica degli affidamenti: il compito della rete sarebbe quello di eseguire uno screening automatico di ampi volumi di crediti, e di segnalare agli operatori le situazioni divenute potenzialmente a rischio, sulle quali effettuare supplementi di indagine e di valutazione.

Si pensa che i risultati ottenuti potrebbero essere ulteriormente migliorati ampliando il campo dell'analisi per comprendere, oltre alle informazioni quantitative desumibili dai bilanci d'esercizio, anche dati sull'andamento del rapporto creditizio (per esempio: numero, frequenza, durata, entità degli sconfinamenti) oppure informazioni di Centrale Rischi, per cogliere tempestivamente ulteriori elementi segnalatici di deterioramento della solvibilità dell'azienda.

Un approccio alternativo per migliorare la previsione della futura solvibilità aziendale potrebbe essere quello di stimare la capacità dell'azienda di creare valore attraverso la previsione dell'indicatore EVA™ (*Economic Value Added*), (indicato con EVA nel seguito) [SternStewart]. In tal caso le segnalazioni ottenute con l'analisi di bilancio potrebbero essere integrate con quelle della previsione dell'EVA. Se l'azienda in futuro creerà valore (ovvero se si prevede un indice EVA maggiore di 0) la previsione di solvibilità aziendale risulterà rafforzata. Diversamente, se si prevede un indice EVA minore di 0, e quindi si prevede che l'azienda distruggerà valore, si rafforza la convinzione di possibili situazioni di insolvenza in futuro.

Un possibile ulteriore sviluppo di questo approccio potrebbe essere quello di prevedere le performance attese di un macro settore, per esempio meccanico, tessile o chimico, o di un'area geografica o di una classe dimensionale di imprese – in termini di creazione/distruzione di valore sulla base dell'EVA – attraverso l'analisi dei rispettivi dati congiunturali, includendo dati esogeni, come i tassi di interesse e di cambio, la domanda interna ed esterna, ed endogeni, quali investimenti e occupazione, forniti da agenzie specializzate. Questa indagine, oltre a rispondere a finalità di approfondimento dei periodici rapporti di indagine congiunturale, potrebbe essere poi utilizzata nell'ottica di un ulteriore affinamento delle analisi sull'insolvenza aziendale, attraverso l'individuazione delle imprese sopra o sotto performanti rispetto all'aggregato di riferimento.

Il lavoro descritto in questo rapporto s'inquadra in questo filone d'indagine ed ha lo scopo di verificare l'ipotesi di prevedere l'andamento dell'EVA al tempo  $t+1$ , noti il valore al tempo  $t$  di alcuni indicatori di bilancio che rappresentano la situazione economico finanziaria dell'impresa esaminata, oltre a indicatori macroeconomici di sintesi per rappresentare il contesto economico-finanziario nel quale opera l'impresa. L'indagine è stata condotta su un campione di 750 imprese manifatturiere della provincia di Bergamo, i cui bilanci sono stati forniti dalla locale Camera di Commercio. In prima istanza si è ipotizzato di prevedere, a livello di singola azienda, semplicemente il segno dell'Economic Value Added e, valutati i risultati, di proseguire la sperimentazione per stimare non solo il segno ma anche l'intensità della performance.

La sperimentazione è stata condotta usando il sistema NeuroSolutions ver 5.0 per simulare via software le reti neurali feed forward back propagation usate in questo lavoro [NeuroSolutions].

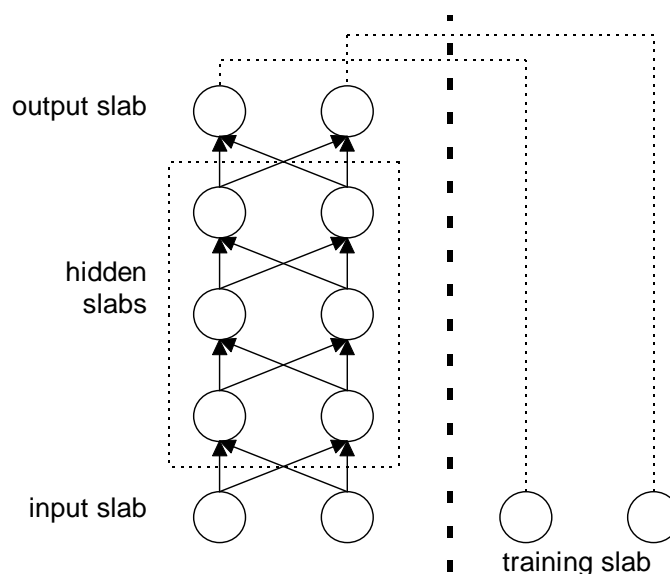
NeuroSolutions è un software per lo sviluppo di R.N. prodotto da: NeuroDimensions Inc.<sup>1</sup>. Di questo prodotto esistono diverse versioni: quella usata è una versione avanzata, interfacciata con Microsoft Excel, che permette di preparare e manipolare i dati necessari all'addestramento e al testing della rete con un foglio elettronico.

## 2. Le Reti Neurali

La ricerca è stata condotta nell'ipotesi che la tecnologia delle reti neurali si adattasse al problema esaminato. In particolare si è ritenuta essenziale la capacità delle R.N. di risolvere problemi per i quali non si conosce una soluzione algoritmica. Le reti usate nella sperimentazione sono infatti reti *feed forward supervisionate* che sono addestrate alla soluzione di un dato problema sottoponendo ripetutamente alla rete un insieme di esempi, detto *training set*, formato da coppie  $(\mathbf{V}, \mathbf{T})$ , dove  $\mathbf{V}$  è un possibile input e  $\mathbf{T}$  rappresenta il corrispondente output corretto. Per ogni  $\mathbf{V}$ , la rete produce un output  $\mathbf{O}$ . L'errore  $(\mathbf{T}-\mathbf{O})$  è usato per modificare il comportamento della rete, variando i pesi delle connessioni tra i nodi, detti *pesi sinattici*, in modo da minimizzare gli errori futuri. Più in particolare: le reti usate sono reti *multistrato feed forward*, istruite con l'algoritmo di apprendimento noto come *backward propagation*. Per una discussione approfondita delle architetture di rete neurale si vedano [Hertz et al. 1991] oppure [Lawrence 1993].

In una rete multistrato feed forward i neuroni sono suddivisi in un certo numero di strati (*slabs*). Tra questi gli strati di input e di output, o *input slab* e *output slab*, servono per la comunicazione con il mondo esterno alla rete. Tra lo strato di input e quello di output ci sono uno o più strati interni (o nascosti), detti appunto *hidden slabs*, che ricevono segnali da tutti i neuroni dello strato precedente e, dopo averli elaborati, li inviano a tutti e soli i neuroni dello strato successivo (da qui il nome di reti feed forward). In ogni neurone gli impulsi in ingresso sono pesati secondo i pesi sinattici e quindi sommati. La somma è infine trasformata secondo una *funzione di trasferimento* il cui output, che costituisce lo stato del neurone, è il segnale in uscita del neurone stesso.

La natura del problema che si vuole risolvere definisce implicitamente la quantità di neuroni presenti nello strato di input e nello strato di output mentre è scopo della sperimentazione scegliere il numero di strati nascosti e il numero di neuroni in ciascuno di essi, ovvero la completa definizione dell'*architettura* della rete.



**Struttura di una rete backpropagation**

<sup>1</sup> NeuroDimension, Inc. - 3701 NW 40th Terrace, Suite 1 - Gainesville, FL 32606 - www.nd.com

Com'è già stato accennato, un generico *vettore di input*  $\mathbf{V}$  definisce lo stato dei neuroni dello strato di input. Questi valori influenzeranno l'output dei neuroni degli strati intermedi fino a raggiungere lo strato di output nel quale lo stato dei neuroni costituisce la risposta della rete  $\mathbf{O}$ . La risposta, nella fase di apprendimento, viene confrontata con la risposta corretta  $\mathbf{T}$ , che è fornita da un apposito strato chiamato *training slab*, attivo esclusivamente in questa fase, in modo che la rete possa calcolare l'errore commesso ( $\mathbf{T}-\mathbf{O}$ ) e minimizzarlo modificando i pesi sinattici (questa volta in direzione opposta, da ciò il nome *backpropagation*). Iterando il procedimento, quando l'errore scende al di sotto del limite stabilito come accettabile si dirà che la rete "ha appreso". Questo significa che la rete, dopo avere esaminato una serie di dati per un sufficiente numero di volte, o epoche, si abitua a fornire risposte simili a quelle mostrate negli esempi. La caratteristica saliente di questo tipo di reti è data dal meccanismo di aggiornamento, che modifica i pesi sinattici in modo da minimizzare l'errore globale commesso dalla rete.

Per formalizzare il comportamento di questo tipo di reti indichiamo con  $\mathbf{j}$  un generico neurone di un qualsiasi strato successivo allo strato di input. Esso riceve in ingresso  $\mathbf{m}$  segnali  $\mathbf{y}_i$  provenienti dagli  $\mathbf{m}$  neuroni dello strato che lo precede e, come detto sopra, li compone con una somma pesata secondo i pesi sinattici  $\mathbf{w}_{ji}$  e, successivamente, li elabora con una trasformazione non lineare  $\mathbf{f}$  per generare l'output  $\mathbf{y}_j$  che il neurone  $\mathbf{j}$  invia a tutti i neuroni dello strato successivo:

$$\mathbf{y}_j = \mathbf{f}(\sum_i \mathbf{y}_i \mathbf{w}_{ji})$$

La *back propagation*, per minimizzare l'errore commesso  $\mathbf{E}$ :

$$\mathbf{E} = 1/2 \sum_j (\mathbf{T}_j - \mathbf{O}_j)^2 = 1/2 \sum_j (\mathbf{T}_j - \mathbf{f}(\sum_k \mathbf{y}_k \mathbf{w}_{jk}))^2$$

usa il metodo del gradiente, pertanto la funzione di trasferimento  $\mathbf{f}$  deve essere differenziabile, e genera un processo iterativo nel quale i valori delle connessioni tra i neuroni sono modificati con la *delta rule*:

$$\Delta \mathbf{w}_{ji} = \alpha \delta_j \mathbf{y}_i \quad \delta_j = \mathbf{e}_j \mathbf{f}'(\mathbf{V}_j)$$

dove la variazione dei pesi sinattici dipende da una costante  $\alpha$ , detta *learning rate*, che determina la velocità di apprendimento, e quindi la rapidità con la quale è cercato il minimo dell'errore. Il termine  $\delta_j$  è proporzionale alla derivata della funzione di trasferimento attraverso il termine  $\mathbf{e}_j$ , che misura l'errore commesso e assume forme diverse a seconda che il nodo  $\mathbf{j}$  appartenga allo strato di output o a uno strato interno:

$$\begin{aligned} \mathbf{e}_j &= \mathbf{T}_j - \mathbf{O}_j && \text{dove } \mathbf{j} \text{ varia sui nodi di output} \\ (*) \quad \mathbf{e}_j &= \sum_s \delta_s \mathbf{w}_{sj} && \text{dove } \mathbf{j} \text{ varia sui nodi di uno strato interno} \end{aligned}$$

Nella (\*)  $\mathbf{w}_{sj}$  rappresenta, naturalmente, il valore dei pesi sinattici delle connessioni tra il nodo  $\mathbf{j}$  e i nodi  $\mathbf{s}$  dello strato successivo a  $\mathbf{j}$ . In questo modo la variazione dei pesi si propaga all'indietro partendo dallo strato di output. Per una analisi dettagliata dell'algoritmo di back propagation si veda, per esempio, [Rojas 1996].

Il software NeuroSolutions, per ottimizzare le prestazioni dell'algoritmo back propagation, aggiorna i pesi con la tecnica del *momentum*. Secondo questo metodo i pesi sono aggiornati con la relazione:

$$\Delta \mathbf{w}_{ji}^{\text{New}} = \alpha \delta_j \mathbf{y}_i + \beta \Delta \mathbf{w}_{ji}^{\text{Old}}, \quad |\beta| < 1$$

dove, con il coefficiente  $\beta$ , detto *momentum*, è introdotto un meccanismo di "memoria" che ha lo scopo di smorzare le possibili oscillazioni nella variazione dei pesi, in quanto la precedente variazione viene rafforzata o ridotta a seconda che il nuovo termine sia dello stesso segno o di segno opposto alla variazione precedente.

In generale questa tecnica di aggiornamento dei pesi assicura la convergenza verso il minimo dell'errore. Importante, in questo meccanismo, è il ruolo giocato dai due parametri *learning rate*  $\alpha$  e

*momentum*  $\beta$ , essi infatti influenzano la velocità dell'apprendimento (talvolta eccessivamente lenta), tramite  $\alpha$ , e le oscillazioni del processo tramite  $\beta$ . In particolare il *momentum* permette di usare *learning rate* più elevati senza incorrere in oscillazioni nella variazione dei pesi.

Con il sistema NeuroSolutions l'apprendimento può essere di tipo *data-adaptive* oppure *block-adaptive* (detto anche *batch*). I due metodi differiscono per il fatto che, nel primo caso, l'aggiornamento dei pesi avviene dopo l'esame di ogni singolo dato del training set, mentre nell'apprendimento block adaptive l'aggiornamento avviene dopo ogni *epoca*, cioè dopo avere sottoposto alla rete tutti i dati del training set. I comportamenti che ne conseguono sono molto differenti: il primo metodo è più sensibile ai casi-limite e ai dati rumorosi e si comporta in maniera più "aggressiva" nell'aggiornare i pesi. Esso è anche più sensibile all'ordine di presentazione degli input ed è raccomandato in applicazioni nelle quali questo ordine è rilevante. Il secondo metodo aggiorna i pesi utilizzando la media degli errori commessi su tutto il blocco di dati, comportandosi quindi in modo più "prudente", ed è preferito quando non si vuole che il modello sia troppo sensibile ai casi-limite.

Normalmente l'apprendimento *data-adaptive* ha tempi di apprendimento minori mentre il *block-adaptive* è più lento ma ha migliori capacità di generalizzazione. La capacità di *generalizzare* è una importante caratteristica della rete. Con questo termine si suole indicare infatti la capacità della rete di fornire risposte corrette a input generici che non sono stati usati nella fase di addestramento. Per misurare la abilità della rete a generalizzare, e quindi il suo livello di apprendimento, è stata usata la tecnica della *independent validation* (o *cross validation*). Secondo questa metodologia i dati a disposizione sono suddivisi in tre sottoinsiemi: *training set*, *validation set* e *test set*. Il training set è usato per addestrare la rete, mentre il validation set è usato per valutare la capacità di generalizzare della rete appena addestrata. Pertanto dopo ogni epoca di apprendimento, o dopo un numero prefissato di epoche, si misura l'errore commesso  $E_T$  sui casi del training set e su quelli del validation set  $E_V$ . L'addestramento della rete prosegue sino al raggiungimento del minimo del valore  $E_T + E_V$  (oppure del solo  $E_V$ ). Questa tecnica si basa sul presupposto che  $E_T$  diminuisca con il crescere delle epoche di addestramento ma che, con il proseguire dell'addestramento, la rete sia soggetta ad overfitting e impari a memoria i casi del training set ma senza acquisire capacità di generalizzare. Il test set che, come il validation set, è composto da casi che non sono mai stati usati per addestrare la rete, serve, alla fine dell'addestramento, per misurare la capacità della rete a risolvere i problemi per i quali è stata progettata. Spesso, per ovviare alla limitatezza dei dati disponibili per addestrare la rete, test set e validation set coincidono.

Prima di cominciare ad istruire una rete con la back propagation bisogna inizializzare i pesi sinattici  $w_{ik}$  dei neuroni che la compongono. Questo avviene assegnando loro dei valori casuali scelti in un dato intervallo. Il sistema NeuroSolutions permette di scegliere, mediante un parametro indicato con **IWM** (*Initial Weight Maximum*), il valore massimo da assegnare ai pesi in questa fase.

La progettazione di una rete neurale richiede al progettista di definire il numero di strati nascosti e dei neuroni in ognuno di essi. Per scegliere l'architettura della rete esistono essenzialmente due metodologie: il *network pruning method* e il *network growing method* [Sietsma-Dow 1988]. La prima metodologia consiste nel partire con un elevato numero di neuroni per poi ridurlo gradatamente eliminando quei neuroni che meno contribuiscono all'ottenimento del risultato finale; la seconda parte da strutture semplici aggiungendo strati e neuroni fino a quando si raggiunge il necessario grado di precisione nelle risposte della rete. Per il nostro caso si è scelta la tecnica di *network growing* che consente una maggiore semplicità e velocità di lavoro.

### 3. La creazione di valore

Secondo il “mainstream” delle teorie economiche degli ultimi decenni, già anticipato nelle grandi corporation americane di inizio novecento e alimentato dai lavori di economisti come Modigliani, Miller, Friedman e Rappaport, compito del management è creare valore per gli azionisti. Diverse e più recenti acquisizioni ampliano la nozione di creazione di valore, incorporando gli interessi non dei soli azionisti, ma di tutti gli stakeholder dell’impresa [Freeman 1984]. Dapprima la creazione di valore è stata individuata nella realizzazione di risposte adeguate ai bisogni espressi dal mercato attraverso la migliore allocazione e organizzazione delle risorse disponibili, e la misurazione del valore avveniva tramite indicatori di natura contabile, quali l’utile d’esercizio in primis, e da indicatori quali ROI, ROE, ROA e altri ancora. Successivamente, il sempre maggior sviluppo delle componenti immateriali del patrimonio aziendale (es. marchi e altri elementi di capitale intellettuale, relazionale e organizzativo), nonché l’emergere di profili di rischio ulteriori rispetto ai tradizionali rischi operativi e di mercato, evidenziarono i limiti dei modelli contabili di misurazione del valore e indussero a ricercare nuovi modelli in grado di meglio catturare le componenti immateriali e finanziarie della creazione di valore.

Negli anni ’80 Bennet Stewart (Stern Stewart & Co ), a partire dagli studi dell’economista Alfred Marshall, [Marshall 1890], riprendendo le gli studi dell’italiano Zappa sul REI (Risultato Economico Integrato) [Zappa 1943], ha sviluppato e registrato un modello di creazione del valore denominato EVA (*Economic Value Added*). EVA è un marchio registrato di Stern Stewart & Co.<sup>2</sup>

L’EVA è un indicatore di performance che misura la differenza tra rendimento e costo del capitale, sottraendo al risultato operativo dopo le imposte (NOPAT) il costo del capitale utilizzato per produrre tale risultato. In altre parole: l’EVA misura il *valore creato*, definito come ciò che rimane dopo aver sottratto al reddito della gestione la remunerazione di tutto il capitale che finanzia l’attività, cioè i mezzi propri ed i debiti. L’EVA è un indicatore più “fine” rispetto ai precedenti, in quanto, assumendo un’ottica più finanziaria, riflette non solo la remunerazione del capitale di debito, ma anche le attese di remunerazione del capitale di rischio, e tiene conto dell’accrescimento del patrimonio aziendale di natura immateriale.

La sua formulazione corrente è la seguente:

$$\text{EVA} = \text{NOPAT} - \text{WACC} * \text{CINO}$$

dove:

**NOPAT** (*Net Operating Profit After Tax*) è il reddito netto prodotto dalla gestione operativa;

**WACC** (*Weighted Average Cost of Capital*) è il costo medio ponderato del capitale, che tiene conto sia del rendimento atteso da parte dell’azionista per il capitale di rischio, sia del costo del capitale di debito;

**CINO** (*Capitale Investito Netto Operativo*) è la dotazione di capitale (di rischio e di debito) assegnata al management per produrre il reddito operativo netto.

L’EVA è quindi pari alla differenza fra il reddito netto prodotto dalla gestione operativa e il costo del capitale impiegato per produrlo ed è possibile parlare di creazione di valore solo quando il reddito netto prodotto dalla gestione operativa, **NOPAT**, eccede il rendimento atteso (in relazione al livello di rischio dell’investimento) dagli azionisti e finanziatori, ossia coloro che hanno fornito il **CINO**.

Nella sperimentazione descritta in queste pagine si è utilizzata una formula di calcolo dell’EVA leggermente differente, in termini di fattori di composizione, da quella proposta dalla letteratura, introducendo alcune semplificazioni, peraltro abituali nella prassi operativa. In particolare:

---

<sup>2</sup> [www.sternstewart.com](http://www.sternstewart.com)

- nel calcolo del **NOPAT** si è applicata al risultato lordo di gestione, **RO**, l'aliquota teorica di tassazione corrispondente all'IRE, ossia il 34%, che può differire in taluni casi anche sensibilmente dall'aliquota effettiva pagata dalle singole imprese nel corso degli anni;
- il costo medio del capitale **WACC** è approssimato come media ponderata tra **T<sub>m</sub>**, ossia il costo medio del capitale di debito, **D**, e **K<sub>e</sub>** che rappresenta il rendimento atteso del capitale proprio, **PN**. **K<sub>e</sub>** è calcolato di norma aggiungendo al rendimento di investimenti privi di rischio, quali i BTP, uno spread per il rischio<sup>3</sup>. In questa sperimentazione **K<sub>e</sub>** è stato estrapolato dalla pubblicazione del *Rapporto sullo stato dell'Economia* pubblicato della Camera di Commercio di Macerata, in occasione della "Prima Giornata dell'Economia", che ha elaborato nel 2003 uno studio a livello nazionale sulle piccole e medie imprese, determinandolo a livello di sistema nella misura del 6,4% per l'anno 2000, del 6,7% per l'anno 2001 e del 6% per l'anno 2002.

Il costo medio del capitale diventa pertanto:

$$WACC = T_m * 0,66 \frac{D}{D + PN} + K_e \frac{PN}{D + PN}$$

Approssimando **T<sub>m</sub>** con il rapporto tra oneri finanziari, **OF**, e debiti verso terzi, **D**:

$$WACC = \frac{OF}{D} * 0,66 \frac{D}{D + PN} + K_e \frac{PN}{D + PN} = \frac{OF * 0,66 + K_e * PN}{D + PN}$$

- il **CINO** è approssimato dal valore del capitale investito, **CI**, definito come la somma del patrimonio netto **PN** e dei debiti verso terzi, **D**.

Di conseguenza, con riferimento ai dati a disposizione, le componenti dell'EVA sono state così definite:

$$\begin{aligned} NOPAT &= RO - (RO * 0,34) \\ WACC &= (OF * 0,66 + K_e * PN) / (D + PN) \\ CI &= D + PN \end{aligned}$$

#### 4. L'oggetto della sperimentazione

Il presente lavoro è finalizzato a sperimentare l'utilizzo delle reti neurali per prevedere la creazione di valore futura di un'impresa in funzione delle sue performance passate (indicatori di bilancio ed EVA) e di alcuni indicatori prospettici riferiti al contesto economico in cui essa opera.

Più precisamente: noti i valori di alcuni indicatori di bilancio e macro economici al tempo **t**, dell'EVA al tempo **t+1** ed il valore degli stessi indicatori al tempo **t+1** si vuole prevedere il valore dell'EVA al tempo **t+2**. Questo approccio si basa sul presupposto che la situazione economico-finanziaria dell'impresa in un dato anno sia indicativa della sua capacità di creare valore nell'anno successivo.

Si sono prese in considerazione 750 imprese manifatturiere della bergamasca, per ognuna delle quali si hanno a disposizione i dati di bilancio per il triennio 2000-2002. Tali imprese rappresentano un campione significativo, pari al 5% della popolazione delle società di capitali di Bergamo e provincia. Il campione, fornito dalla Camera di Commercio di Bergamo, rispecchia l'universo delle oltre 14.000 imprese per: tipologia (Srl e Spa), settore merceologico e dimensione (imprese piccolissime: con meno di 10 dipendenti, piccole: tra 10 e 49 dipendenti e medie: tra 50 e 249 dipendenti).

Per ogni impresa sono stati presi in considerazione i seguenti indicatori di bilancio:

<sup>3</sup> Rendimento BPT + spread rischio ± 0,5% in base all'andamento del settore economico

- **cost/income (C/I)**, calcolato come somma dei costi del personale e degli ammortamenti delle immobilizzazioni immateriali e materiali, fratto ricavi da vendite e prestazioni;
- **attivo immobilizzato / totale attivo (AI/TA)**;
- **leverage (LEV)**, inteso come rapporto tra capitale proprio e capitale di terzi;
- **tasso di rotazione dell'attivo** in due configurazioni:
  - **valore della produzione / totale attivo (VP/TA)**, dove il valore della produzione rappresenta l'output complessivo dell'attività produttiva cui si contrappone il complesso dei costi per la determinazione dell'Ebit (risultato aziendale prima delle imposte e degli oneri finanziari);
  - **fatturato / totale attivo (F/TA)**, al fine di depurare l'effetto di variazione del magazzino, che rappresenta una componente di valore non realizzato sul mercato.

Ai cinque indicatori di bilancio che rappresentano la situazione economico-finanziaria dell'impresa, sono stati aggiunti due indicatori di carattere macroeconomico, che rappresentano il contesto economico-finanziario entro il quale ciascuna impresa opera:

- **tasso di inflazione** rilevato dall'ISTAT: 2,7% per l'anno 2000, 2,4% per l'anno 2001 e 2,5% per l'anno 2002;
- **tasso di sconto**, calcolato come media mensile ponderata dei tassi di riferimento<sup>4</sup>: 4% per l'anno 2000, 4,37% per l'anno 2001 e 2,9% per l'anno 2002.

Gli indici usati nella sperimentazione sono indici di bilancio riferiti agli esercizi 2001 e 2002 e sono stati ripartiti in un training set composto da 500 imprese, delle quali due terzi creano valore (cioè presentano un EVA positivo) nel 2002 e un terzo lo distruggono, e un test set composto da 250 aziende, delle quali 139 creano valore e 111 lo distruggono, tutte scelte a caso in modo proporzionale ai diversi gruppi di appartenenza.

Il rapporto di 1 a 2 tra imprese con EVA negativo e imprese con EVA positivo è un valore di compromesso tra la necessità di rappresentare il campione, nel quale le imprese che distruggono valore sono in netta minoranza, e la necessità di comporre un training set con un numero di tali imprese sufficiente numeroso perché la rete eviti il grave errore di non riconoscere un'impresa che distrugge valore. Gli outliers, cioè i dati anomali o rumorosi, sono distribuiti proporzionalmente nei diversi set e le aziende che creano o distruggono valore sono state inserite nei set in modo casuale. In questo lavoro training set e validation set coincidono ovvero sono costituiti dai medesimi dati.

## 5. La preparazione dei dati di input per la rete

Sui dati selezionati è stata effettuata un'analisi di correlazione per individuare le eventuali dipendenze tra le diverse variabili considerate al fine di depurare la base dati da informazioni ridondanti. L'analisi effettuata, i cui risultati sono sintetizzati nella tabella sotto riportata, ha evidenziato che non esistono correlazioni tra gli indicatori identificati e pertanto tutti sono stati utilizzati nella sperimentazione.

	C/I	AI/TA	LEV	VP/TA	F/TA
C/I	-	0,240801419	0,11425	-0,216209	-0,01212
AI/TA	-	-	0,200158	-0,342259	0,024645
LEV	-	-	-	-0,214991	0,306181
VP/TA	-	-	-	-	-0,02322
F/TA	-	-	-	-	-

<sup>4</sup> La misura del tasso ufficiale di riferimento con decorrenza 1 gennaio 1999 è stata determinata sulla base del cessato tasso ufficiale di sconto alla data del 31/12/98 (art. 2 d.l. 213/98) dal 2004 determinato dalla **Banca Centrale Europea** (The European Central Bank)

Successivamente, come già fatto nelle precedenti sperimentazioni sulla previsione delle insolvenze, i dati destinati a comporre il set di valori di input per la rete neurale hanno subito alcuni trattamenti, per ottimizzarne l'utilizzo da parte della rete stessa.

Va preventivamente considerata la presenza di outliers, cioè valori anomali talmente elevati che provocherebbero la compressione dei valori "normali" in un intervallo molto ristretto attorno allo 0, riducendone drasticamente la significatività. Gli outliers non sono errori di rilevazione ma espressione di situazioni tutt'altro che rare nel mondo aziendale: eliminarli significherebbe escludere dal campione una porzione non trascurabile della popolazione. Pertanto, per non scartare le aziende che presentano outliers e nel contempo limitare l'effetto distorsivo di questi ultimi, prima di attuare la normalizzazione, i valori anomali sono stati sostituiti con opportuni valori limite come sotto precisato.

- **Analisi della frequenza di distribuzione e appiattimento dei valori di coda su valori limite.** Per ognuno degli indici considerati, con l'eccezione del PIL e dell'inflazione, che non dipendono dalle aziende analizzate e sono costanti per l'anno di analisi dei bilanci, è stato determinato un campo di variazione accettabile e i valori eccedenti sono stati riportati ai valori limite del normale campo di variazione. La tabella in figura riporta, per ogni indice, i valori limite accettati e, nell'ultima colonna, indicata con *Over*, è riportato il numero di casi nei quali l'indice esce dal range di valori compresi tra *Min* e *Max*.

	Min	Max	Over
<b>PIL</b>	-	-	0
<b>Inflazione</b>	-	-	0
<b>C/I</b>	0%	100%	3
<b>AI/TA</b>	0%	100%	0
<b>LEV</b>	-2	40	24
<b>VP/TA e F/TA</b>	0	9	7

- **Normalizzazione** degli indici, per ottenere una distribuzione con media  $\mu = 0$  e deviazione standard  $\sigma = 1$ , sottoponendoli alla trasformazione:  $Z = (x - \mu) / \sigma$ .

Gli indici utilizzati hanno natura profondamente diversa tra loro e per questo motivo assumono valori numerici entro intervalli molto diversi. Per evitare che la diversità dei campi di variazione dei diversi indicatori influenzi negativamente le capacità di apprendimento e generalizzazione della rete si esegue la *binarizzazione* degli indicatori, riportandone il campo di variabilità nell'intervallo [0,1] con la trasformazione:  $f(z) = (z - z_{\min}) / (z_{\max} - z_{\min})$ . Poiché agli outliers sono stati attribuiti i valori  $Z_{\min}$  o  $Z_{\max}$ , la binarizzazione ha l'effetto di proiettarli nei valori 0 e, rispettivamente, 1.

- La **binarizzazione** dei dati non è espressamente richiesta dalla versione del software Neurosolutions usata in questo lavoro, e non è stata applicata nella prima fase della sperimentazione, ma solo in una successiva fase di affinamento del processo.

## 6. La sperimentazione: fase 1

Obiettivo della prima fase di sperimentazione è stato la verifica della capacità di apprendimento della rete, attraverso il calcolo dello scarto quadratico medio quale misura dell'errore.

Alla rete sono stati forniti i seguenti dati in input:

- indicatori di bilancio e dati macroeconomici al tempo t (anno 2000) - 7 valori,
- EVA calcolato relativo al tempo t+1 (anno 2001) - 1 valore,



- indicatori di bilancio e dati macroeconomici al tempo t+1 (anno 2001) - 7 valori,

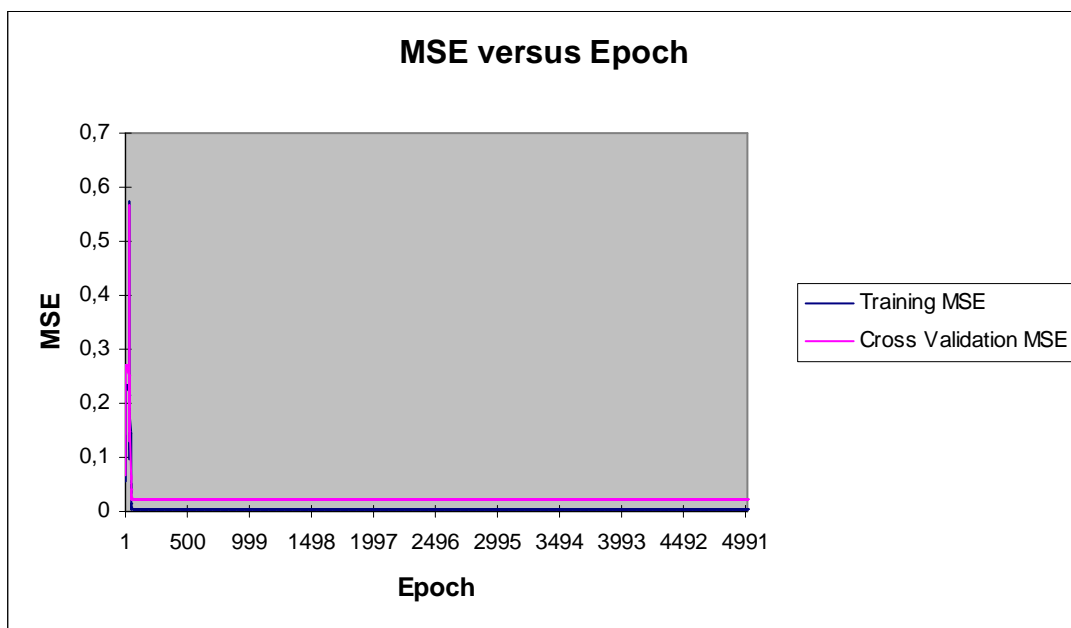
con lo scopo di ottenere una previsione dell'EVA al tempo t+2 (anno 2002). Conoscendo ex post i valori dell'EVA è possibile procedere con la verifica dei risultati forniti dalla Rete.

L'algoritmo di apprendimento impiegato è la Back Propagation, che modifica i pesi sinattici in modo da minimizzare l'errore globale commesso dalla rete, valutato con lo scarto quadratico medio (MSE) fra output della rete e risultati corretti.

Sono state sperimentate diverse R.N. con 15 neuroni nello strato d'ingresso, in corrispondenza dei 15 valori di input sopra specificati, e un neurone nello strato di output dove la rete consegna la previsione dell'EVA. Fissato il numero di neuroni negli strati d'ingresso e d'uscita, sono state sperimentate diverse reti, variando il numero degli strati nascosti e il numero di neuroni di ogni strato nascosto, i parametri  $\alpha$  (tasso di apprendimento) e  $\beta$  (momentum), l'error tolerance, la metodologia di aggiornamento dei pesi e il valore massimo iniziale dei pesi (IWM).

I migliori risultati sono stati evidenziati da due reti: una di architettura 15-8-4-1, cioè con due strati nascosti, e l'altra di tipo 15-8-1 con un solo strato nascosto con otto neuroni. Poiché le due reti hanno fornito performance praticamente identiche l'attenzione si è concentrata sulla rete con un solo strato nascosto, più semplice da addestrare e da analizzare.

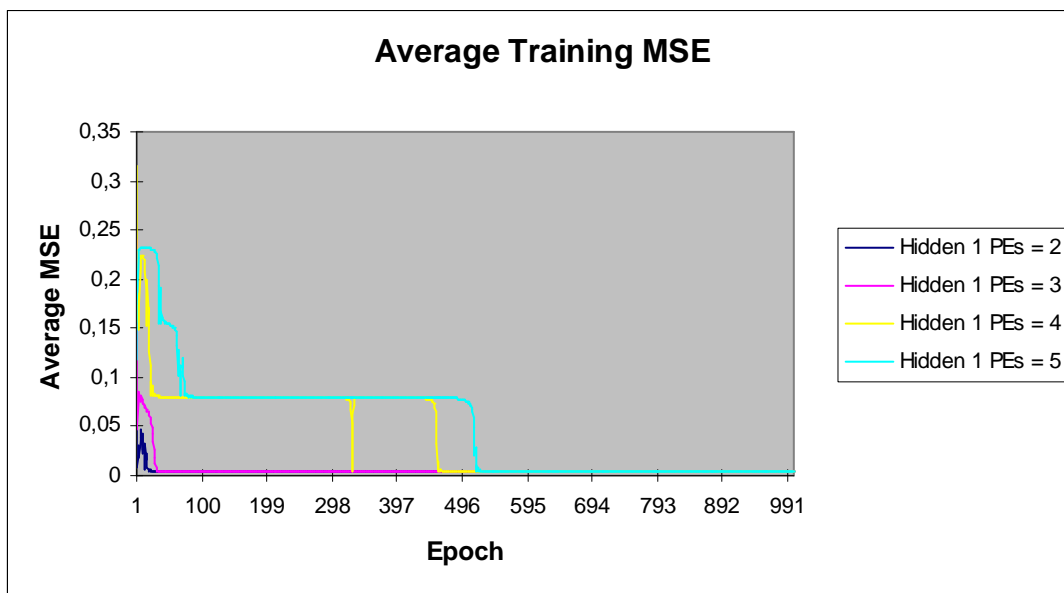
L'addestramento della rete è stato iterato per 5.000 epoche (ovvero tutti i record delle imprese sotto esame sono stati elaborati dalla rete per 5.000 volte), addestrando la rete con valori medi sia per i pesi iniziali (IWM) che per learning rate e momentum (rispettivamente 0,5 e 0,7), ed è stata individuata l'epoca in cui la somma dell'errore sul training set e sul test set risultava minore:  $\min(E_t + E_v)^5$ . Questa è definita epoca di congelamento, ovvero l'epoca in cui l'elaborazione viene terminata e i pesi vengono congelati. Secondo il modello utilizzato a questo punto la rete ha raggiunto il grado di addestramento ottimale in relazione alla capacità di "comprendere" il fenomeno in analisi. Dal grafico seguente si può notare che la misura dell'errore è diminuita drasticamente dopo poche decine di epoche, evidenziando una elevata capacità di apprendimento della rete.



<sup>5</sup> Nel caso in cui vi siano diverse epoche in cui si rilevi la medesima somma dell'errore, si considera la prima epoca in cui è minimo l'errore sul test set.

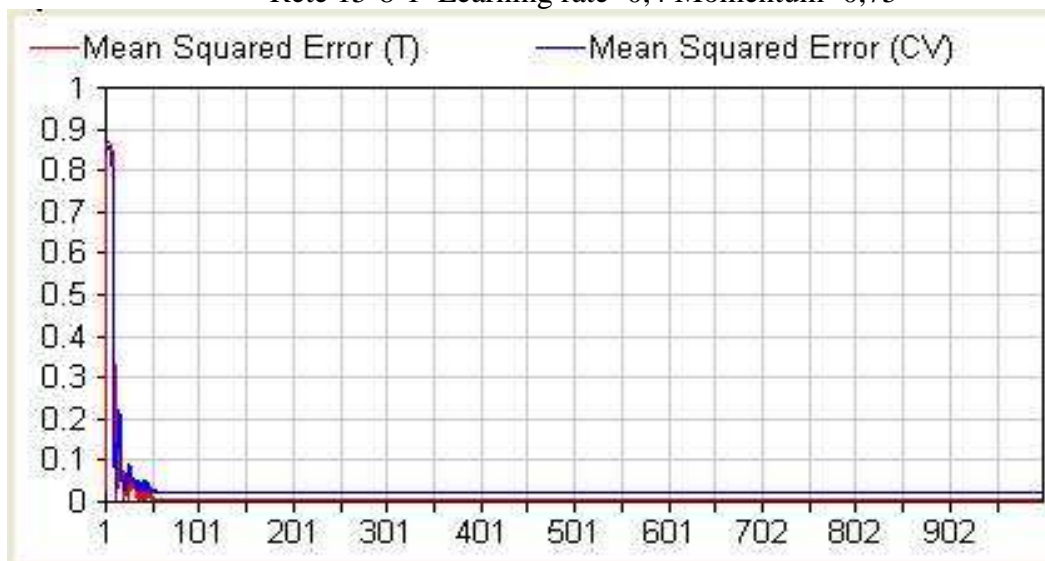
Parallelamente alla verifica di apprendimento, si è anche valutata la qualità degli indicatori in termini di contributo effettivo all'elaborazione dell'output, utilizzando una specifica funzionalità offerta dallo strumento Neurosolutions 5, che permette di controllare il valore attribuito ai pesi sinattici al termine dell'addestramento. Si è rilevato che il tasso di inflazione e il tasso di interesse, in quanto costanti per tutti i record da elaborare, sono stati di fatto inutilizzati dalla rete.

Il seguente grafico seguente mostra come l'iterazione della procedura con un differente numero di neuroni nello strato nascosto abbia influito sui tempi di apprendimento ma non abbia prodotto miglioramenti delle prestazioni.



Prestazioni molto buone, comunque in linea con le prime evidenze, sono state ottenute sfruttando una funzionalità del sistema Neurosolutions che modifica autonomamente i parametri della rete (struttura, coefficienti, numero di strati nascosti, ecc) in successive iterazioni dell'elaborazione. Partendo da un'architettura semplice, il software implementa il numero di strati nascosti e varia i parametri in funzione dei risultati raggiunti in termini di MSE calcolato sul training set. E' necessario comunque verificare se l'architettura proposta non presenti problematiche e perdita di performance sul validation set. Al termine dell'esecuzione il sistema ha selezionato la medesima architettura di rete individuata nelle precedenti analisi.

Rete 15-8-1 Learning rate=0,4 Momentum=0,75



I risultati di questa prima fase di sperimentazione, in termini di misura dello scarto quadratico medio dell'output effettivo rispetto a quello atteso sia nel training che nella cross validation, paiono dimostrare un elevato livello di apprendimento in tempi rapidi.

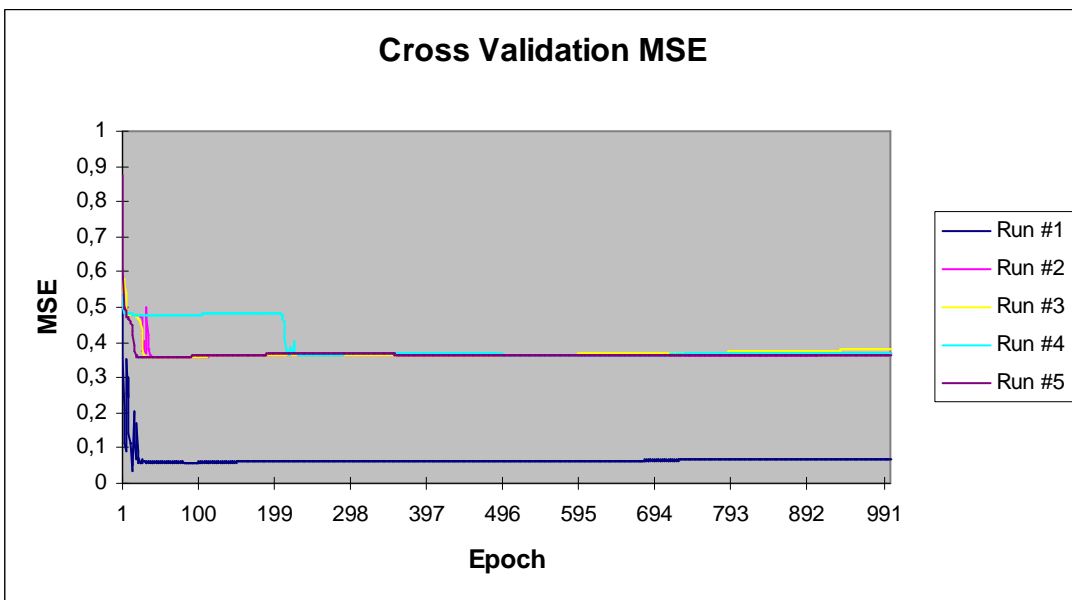
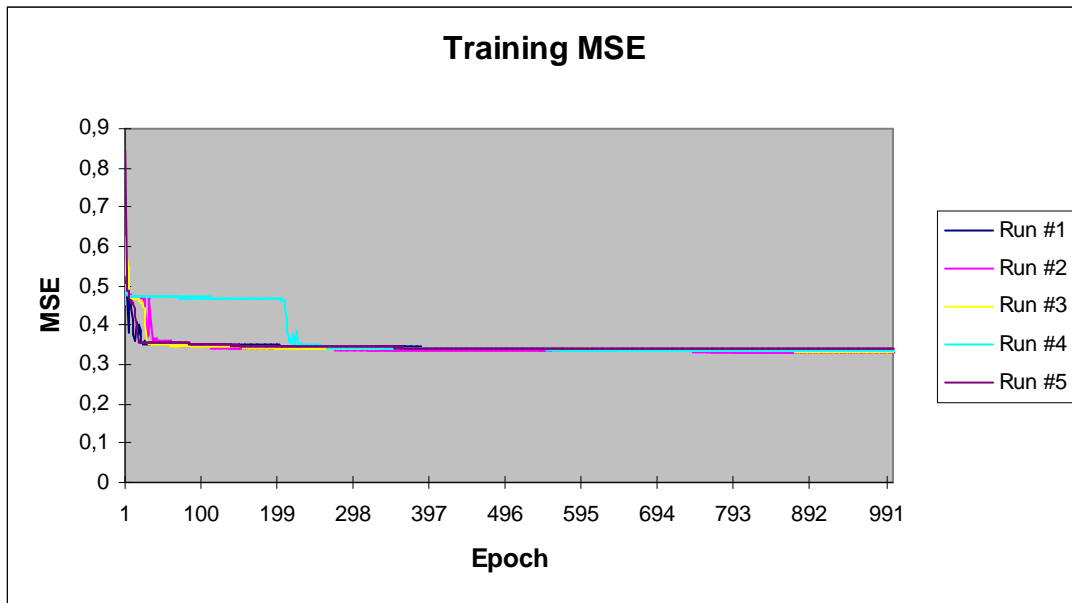
## 7. La sperimentazione: fase 2

Verificata la capacità di apprendimento della rete, in questa seconda fase si è indagata la sua capacità di generalizzare discriminando fra le imprese che creano valore e quelle che lo distruggono. Per farlo la rete viene addestrata a restituire in output una variabile binaria che associa il valore 0 agli EVA negativi (distruzione di valore) e il valore 1 agli EVA positivi (creazione di valore). Poiché gli indicatori utilizzati hanno natura profondamente diversa tra loro e quindi assumono valori numerici entro intervalli molto differenti, in questa fase, come è stato anticipato, si è ritenuto opportuno ridurre il campo di variabilità di ogni indicatore nell'intervallo tra 0 e 1 attraverso la binarizzazione dei dati, come descritto nel precedente paragrafo 5.

Nelle tabelle e nelle figure seguenti sono riportati i risultati ottenuti con una rete di architettura 15-8-1 che comprende tra i dati di input i tassi d'interesse e l'inflazione. Le didascalie Run #1, Run #2, ... , Run #5 nelle figure fanno riferimento a diverse esecuzioni ottenute sfruttando una specifica funzionalità di NeuroSolutions 5. Il software NeuroSolutions 5 dispone infatti di una funzione che, definita un'architettura di rete e scelto un opportuno training set, permette di automatizzare il processo di addestramento, ripetendo sino a cinque volte il processo, modificando il tasso di apprendimento e il momentum con un passo prefissato scelto dall'utente. Con questa procedura si è verificato che la modifica di questi parametri ha portato ad una variazione della velocità di apprendimento ma non del grado di apprendimento. La rete infatti alla fine di ogni elaborazione commette gli stessi errori (il MSE ottenuto è lo stesso). La fase di cross validation beneficia invece della scelta di un valore di *momentum* relativamente basso (0,5) mentre peggiora man mano che il valore si avvicina all'unità. Il MSE infatti dimostra un grado di generalizzazione superiore nelle prove con parametri più "aggressivi".

Run 1#	Learning rate=0,5	Momentum=0,7
Run 2#	Learning rate=0,55	Momentum=0,75
Run 3#	Learning rate=0,6	Momentum=0,8
Run 4#	Learning rate=0,65	Momentum=0,85
Run 5#	Learning rate=0,7	Momentum=0,90

La rete individua correttamente il 68% imprese che al tempo t+2 manifestano un EVA positivo (creano valore) e il 47% di quelle che al tempo t+2 evidenziano un EVA negativo (distruggono valore). Nei grafici e nelle tabelle seguenti sono evidenziate le prestazioni della rete. Le ultime due tabelle, in particolare, riportano i prospetti dei risultati ottenuti dalla rete e gli errori commessi.



All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	0,329948572	0,012016218	0,357773146	0,001635538
Average of Final MSEs	0,329958504	0,011997241	0,375233948	0,015606909

Best Networks	Training	Cross Validation
Run #	1	4
Epoch #	999	656
Minimum MSE	0,308652845	0,355180901
Final MSE	0,308727412	0,355424656

<b>Desiderato Ottenuto</b>	<b>EVA2002(1)</b>	<b>EVA2002(0)</b>
<b>EVA2002(1)</b>	<b>113 (67,7%)</b>	44 (53%)
<b>EVA2002(0)</b>	54 (32,3%)	<b>39 (47%)</b>

<b>Performance</b>	<b>EVA2002(1)</b>	<b>EVA2002(0)</b>
<b>MSE</b>	0,219237322	0,220134567
<b>NMSE</b>	0,988552962	0,992598692
<b>MAE</b>	0,433637968	0,435391021
<b>Min Abs Error</b>	0,260483291	0,255384273
<b>Max Abs Error</b>	0,739034027	0,73913255
<b>r</b>	0,120934913	0,120708299
<b>Percent Correct</b>	<b>67,66467065</b>	<b>46,98795180</b>

Avendo rilevato che il tasso di inflazione e il tasso di interesse sono stati, di fatto, inutilizzati dalla rete si è ipotizzato di poterne migliorare le prestazioni utilizzando, invece di questi due indicatori, un indice prospettico di costo del denaro specifico per ogni impresa. L'indice usato è stato calcolato sommando al costo del denaro rilevato al tempo t+1 (esercizio dei dati consuntivi di bilancio) la differenza tra il tasso di riferimento previsto al tempo t+2 e il tasso di riferimento rilevato al tempo t+1. Con ciò si è ritenuto di intercettare la diversa attitudine delle imprese a beneficiare di eventuali riduzioni dei tassi o a soffrire di eventuali incrementi in funzione della propria specifica situazione finanziaria e reddituale di partenza.

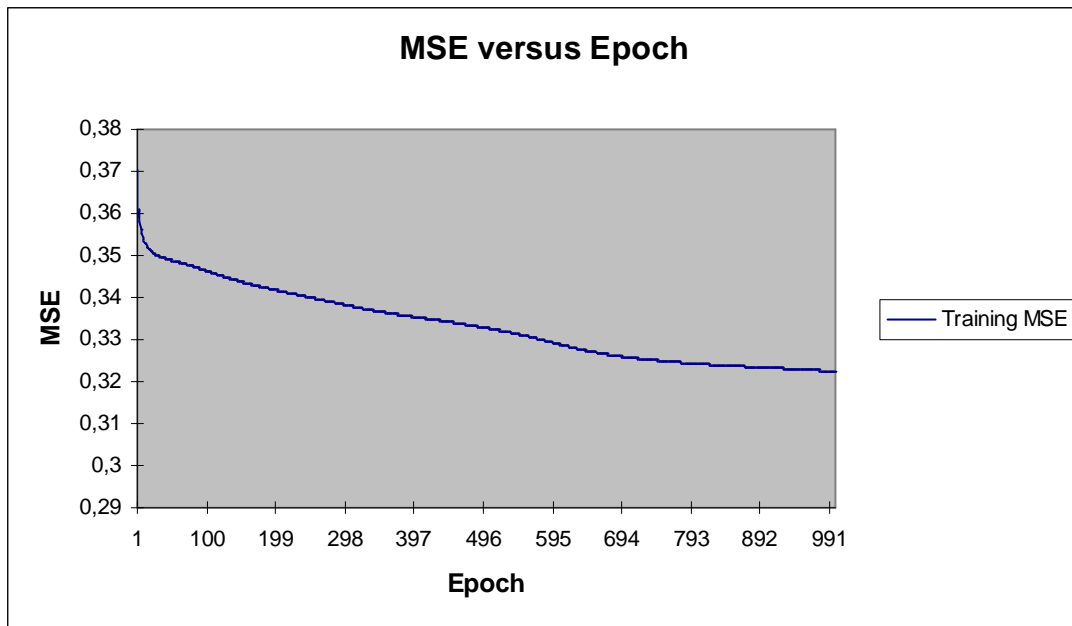
La nuova rete, di architettura 13-7-1, ha però prodotto risultati complessivamente peggiori, per via della minor capacità di individuare correttamente le imprese che distruggono valore. Come risulta dalle tabelle seguenti, essa individua infatti correttamente il 72% delle imprese che creano valore (rispetto al precedente 68%), ma solo il 40% delle imprese che distruggono valore (rispetto al precedente 47%).

<b>Desiderato Ottenuto</b>	<b>EVA2002(1)</b>	<b>EVA2002(0)</b>
<b>EVA2002(1)</b>	<b>126 (71,6%)</b>	45 (59,6%)
<b>EVA2002(0)</b>	50 (28,4%)	<b>29 (40,4%)</b>

<b>Performance</b>	<b>EVA2002(1)</b>	<b>EVA2002(0)</b>
<b>MSE</b>	0,235769395	0,223853088
<b>NMSE</b>	1,131417936	1,074233567
<b>MAE</b>	0,483804779	0,445750161
<b>Min Abs Error</b>	0,411738906	0,239209016
<b>Max Abs Error</b>	0,600738431	0,971965725
<b>r</b>	0,061731138	-0,039679978
<b>Percent Correct</b>	<b>71,59090909</b>	<b>40,405405405</b>

Si è anche provato a utilizzare una diversa misura dell'errore per valutare l'apprendimento della rete, ricercando non più la minimizzazione della somma degli errori in training e in test, ma la

minimizzazione del solo errore in addestramento. In questo caso tutte le aziende a disposizione sono state impiegate per il solo training set, potendo così impiegare per l'elaborazione 750 record, un terzo in più della precedente esperienza. Anche questo accorgimento non ha però prodotto risultati soddisfacenti, avendo migliorato sensibilmente la capacità individuare correttamente le imprese che creano valore, ma peggiorato altrettanto sensibilmente la capacità di individuare quelle che distruggono valore, come emerge dai grafici e tabelle seguenti.



Best Network	Training
Epoch #	1000
Minimum MSE	0,322349799
Final MSE	0,322349799

Desiderato Ottenuto	EVA2002(1)	EVA2002(0)
EVA2002(1)	<b>442 (88%)</b>	213 (87%)
EVA2002(0)	62 (12%)	<b>33 (13%)</b>

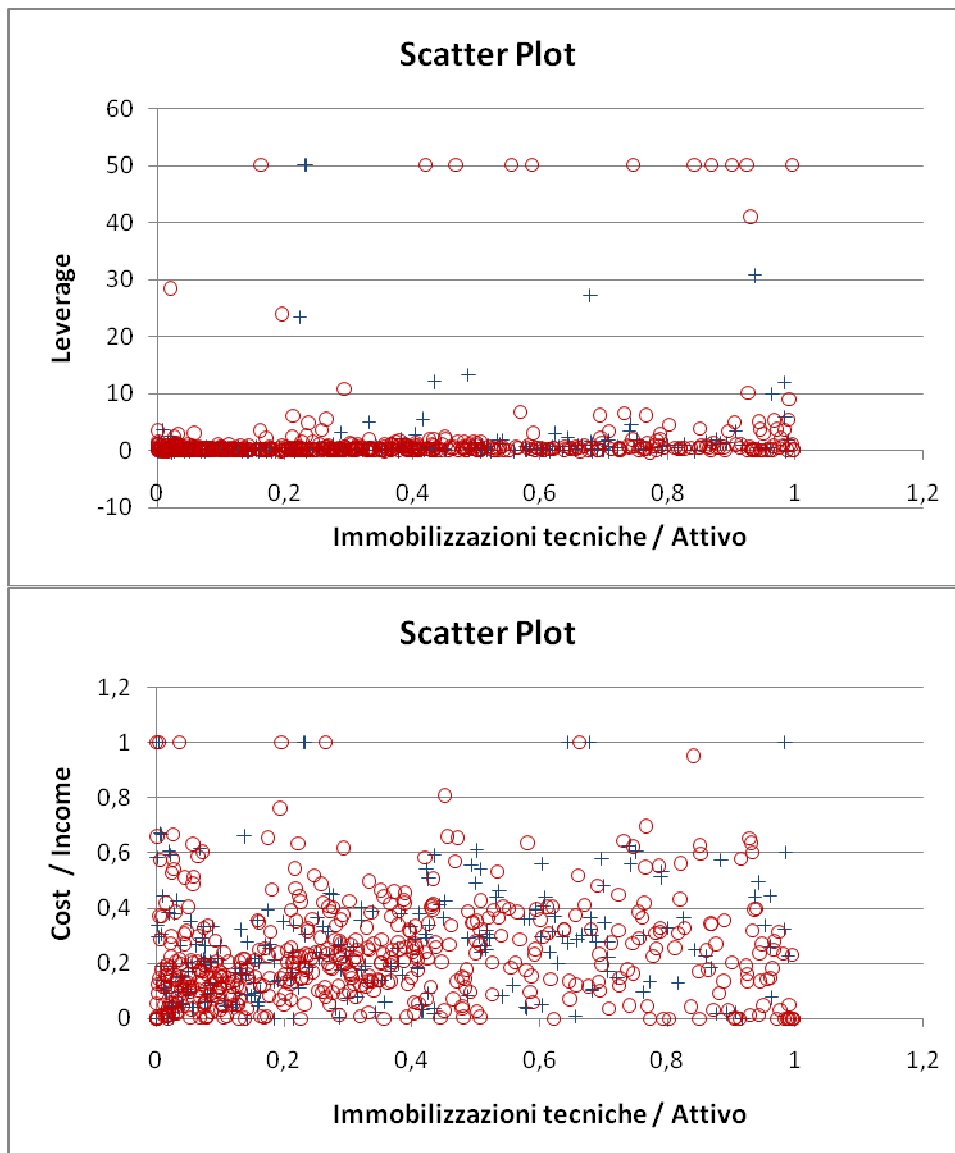
Performance	EVA2002(1)	EVA2002(0)
MSE	0,255091434	0,255988768
NMSE	1,224141173	1,228447328
MAE	0,459558787	0,461344487
Min Abs Error	0,022877173	0,022932829
Max Abs Error	1,049951385	1,048884345
R	-0,072213065	-0,072117157
Percent Correct	<b>87,6984127</b>	<b>13,41463415</b>

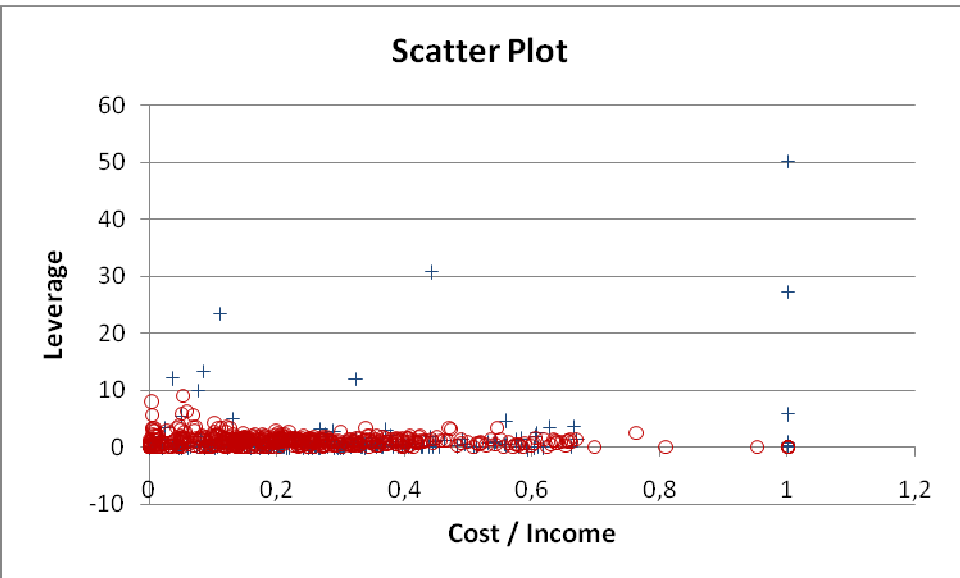
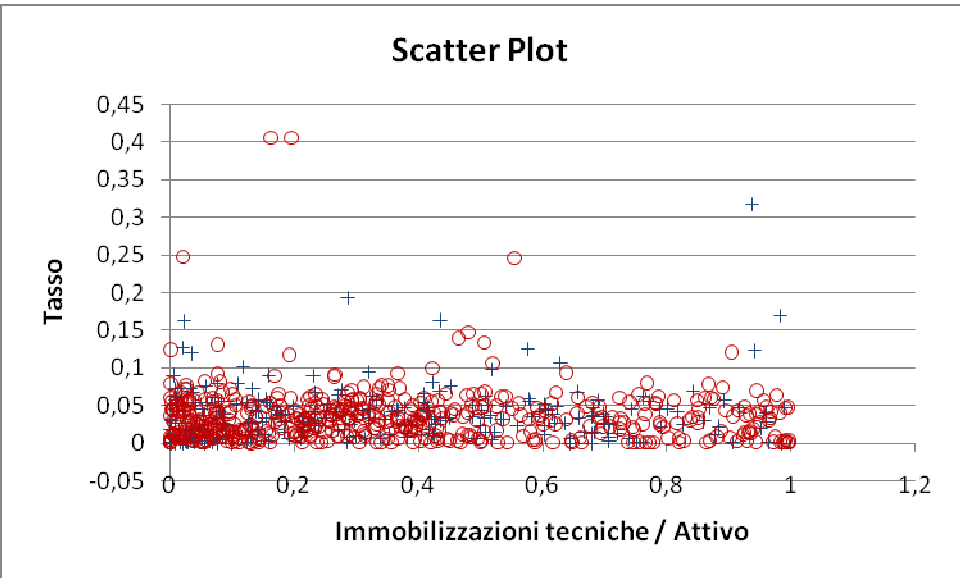
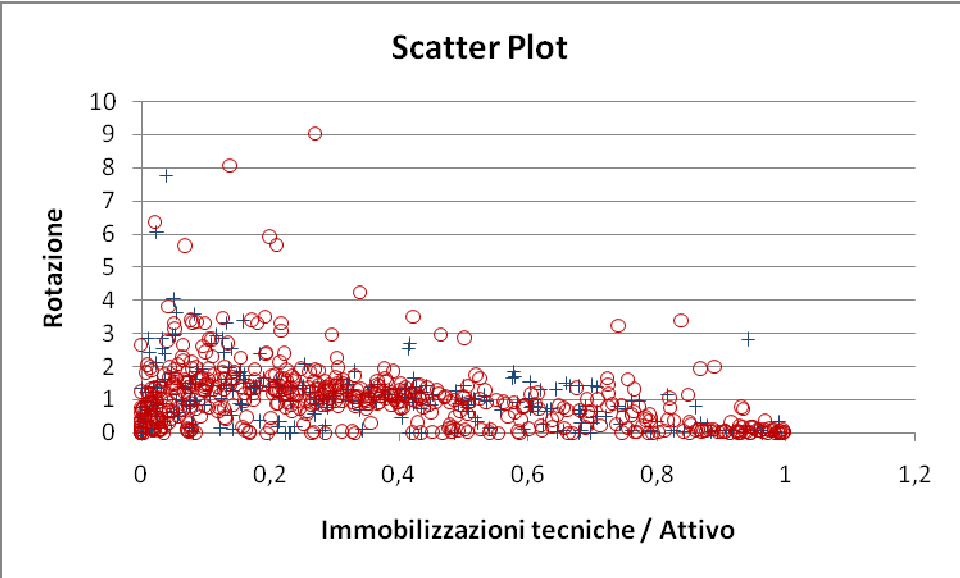
## 8. Conclusioni

Complessivamente, nonostante nella fase 1 le R.N. realizzate abbiano manifestato una buona capacità di apprendimento, arrivando a un basso valore dello scarto quadratico medio dell'errore, le prestazioni in termini di corretta previsione della creazione/distruzione di valore non paiono soddisfacenti. In particolare l'errore più grave, che consiste nella mancata individuazione delle imprese che distruggono valore, si mantiene sempre oltre il 50% dei casi.

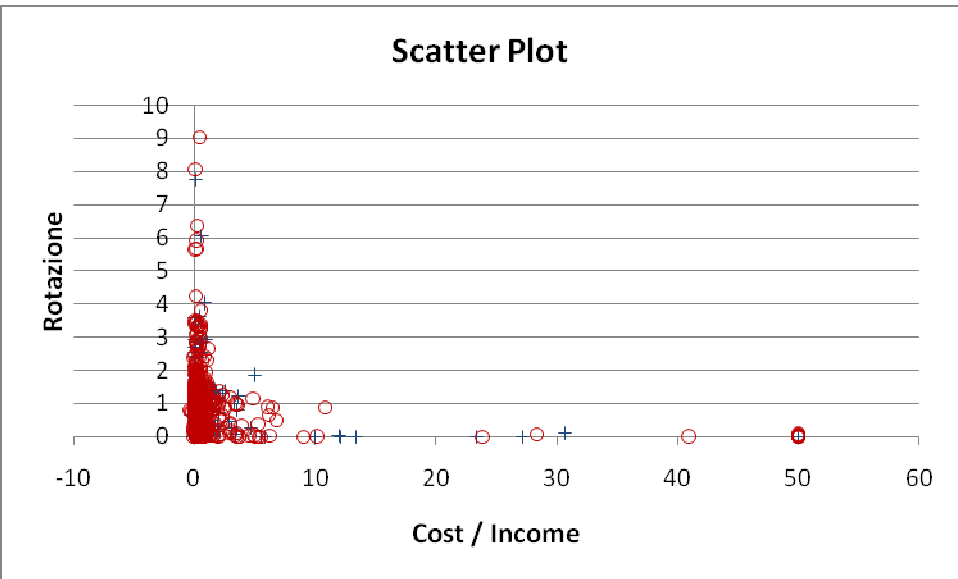
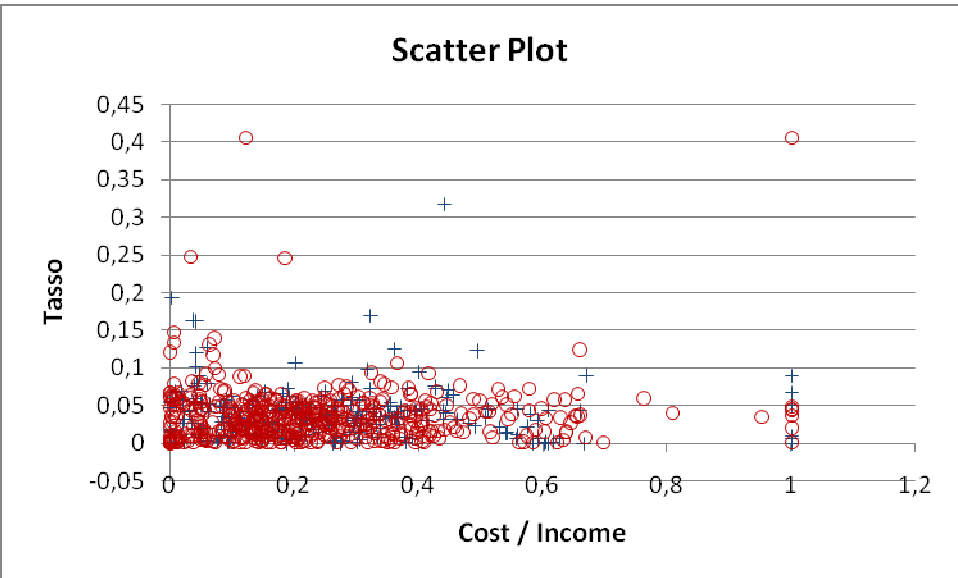
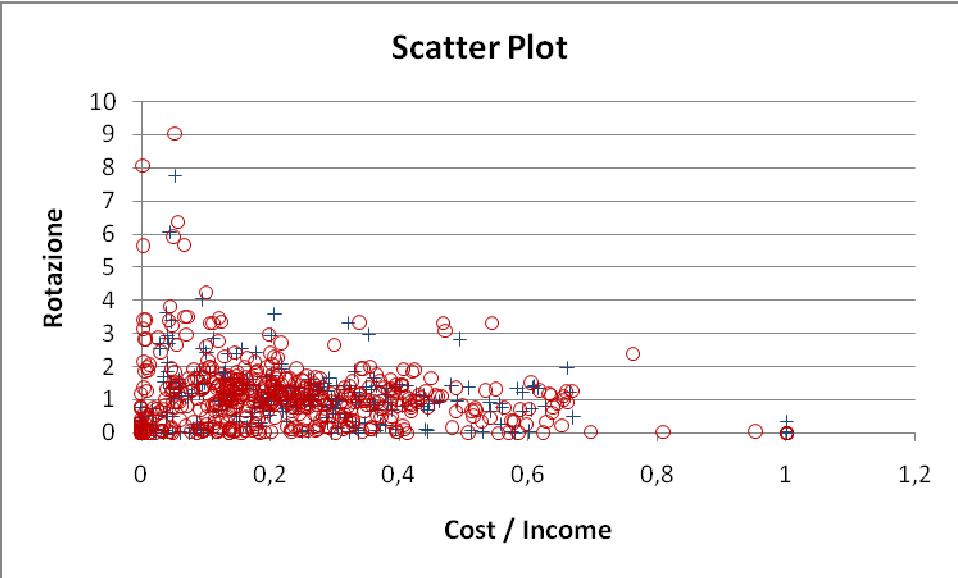
Ciò impone una riflessione sulla qualità dei dati utilizzati nella sperimentazione e in particolare sulla loro effettiva capacità di fornire informazioni utili per l'analisi. A tal fine sono state analizzate le distribuzioni degli indicatori, confrontando di volta in volta coppie degli stessi. In questi confronti l'indice indicato come *Tasso* è l'indice di costo del denaro introdotto nella fase 2 della sperimentazione. Dall'esame delle distribuzioni si nota che la quasi totalità delle imprese risulta concentrata lungo determinate direttrici - vicino ad un asse o talvolta vicino all'origine degli assi - e che i due cluster sono sovrapposti per tutte le coppie di indicatori considerate, suggerendo una scarsa capacità di differenziazione delle imprese rispetto alla creazione di valore. Si noti che nei grafici che seguono le imprese con EVA positivo o negativo sono rappresentate con i simboli:

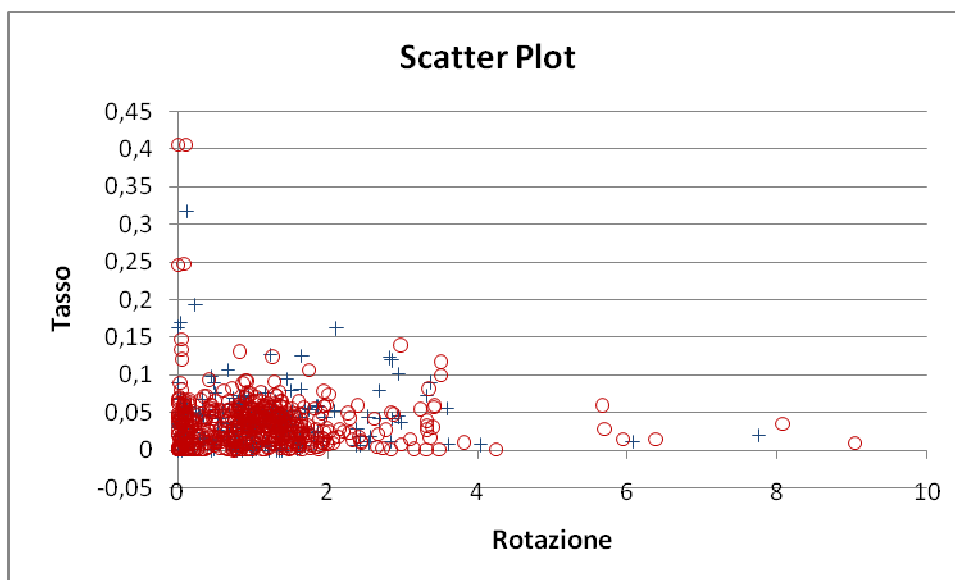
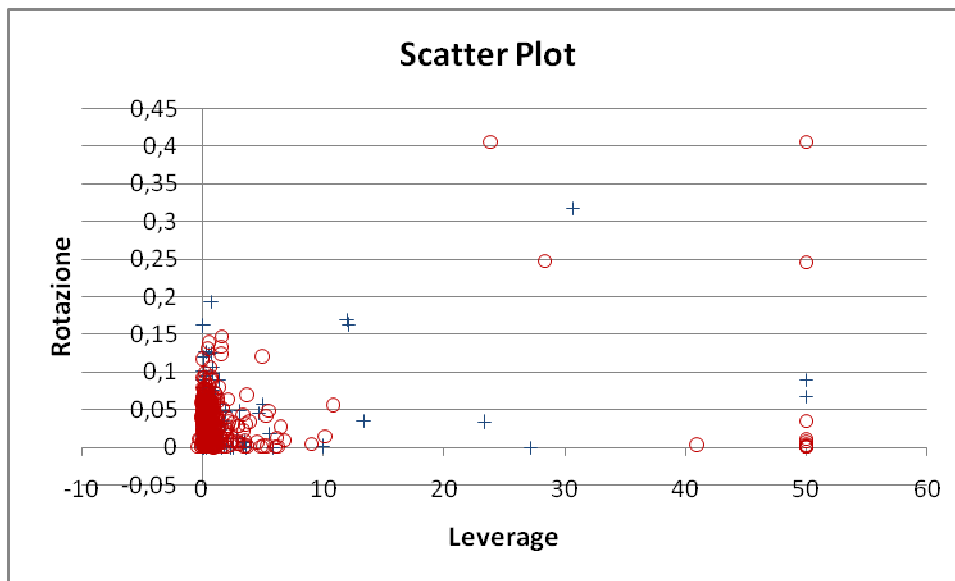
- per rappresentare le aziende che creano valore
- + per rappresentare le aziende che distruggono valore







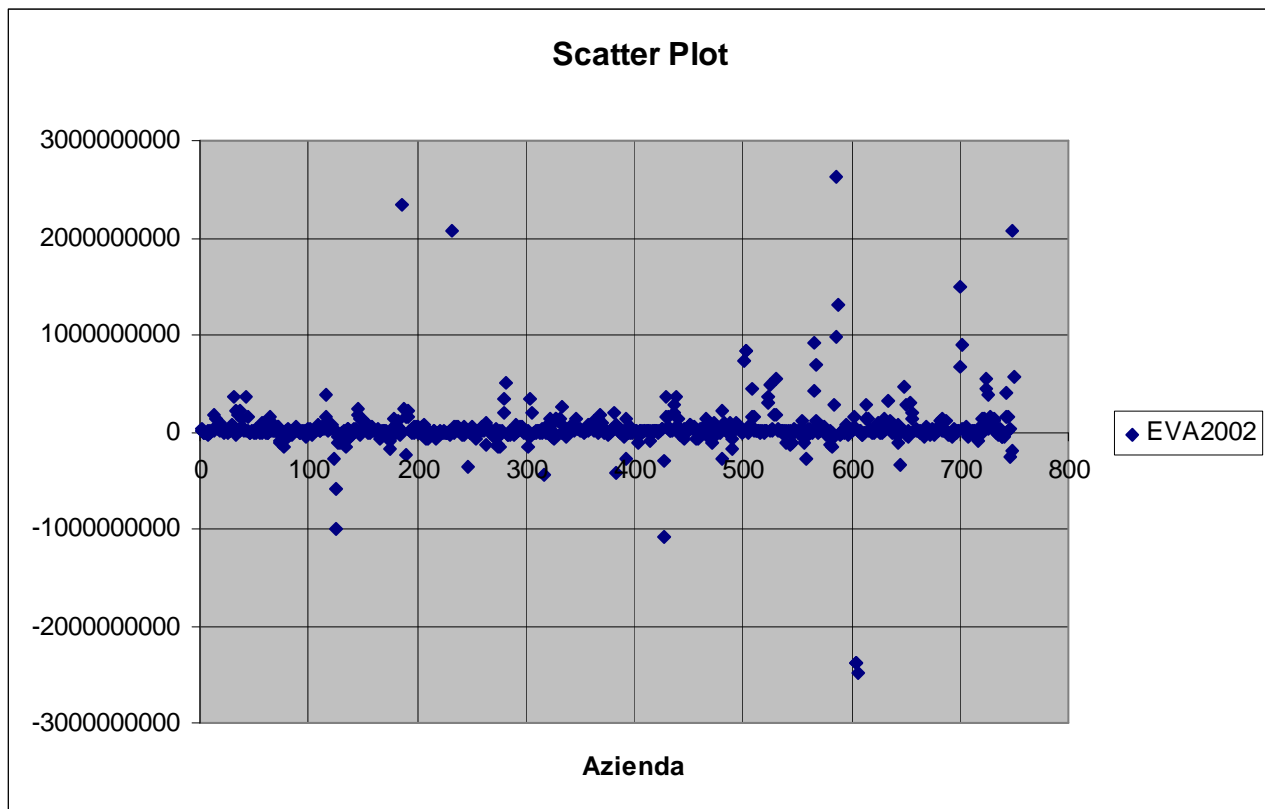




Il fenomeno appare ancor più evidente analizzando il valore effettivo dell'EVA delle imprese considerate: salvo casi particolari (outliers) che si possono discostare molto, sia in senso positivo che negativo dalla media degli EVA, la maggior parte dei valori si discosta poco dallo 0. Questa particolarità dell'EVA contribuisce probabilmente al fatto che la rete abbia difficoltà a classificare esattamente le imprese.

Nel grafico sottostante viene visualizzata la distribuzione del valore dell'EVA delle imprese disposte in base al loro numero identificativo. Per queste ragioni si ritiene che i "quadri clinici" offerti dagli indicatori di bilancio per imprese che creano o distruggono valore non siano così diversificati, soprattutto per valori di EVA (positivi o negativi) poco differenziati, come è il caso della maggior parte delle imprese del campione.

Per un'analisi più approfondita sulle determinanti della creazione di valore sarebbe necessario considerare anche indicatori più fini, e forse più appropriati per la soluzione del problema in oggetto, quali le misure di efficienza interna o di qualità delle strategie e dei processi di gestione, che, purtroppo, sono informazioni scarsamente disponibili a un analista esterno.



## Bibliografia

- [Atiya 2001] A.F.Atiya, *Bankruptcy prediction for credit risk using Neural Networks: a survey and new result*, IEEE Transaction on Neural Networks, vol.12, no. 4, July 2001.
- [Boschioli et al. 2001] C.Boschioli, D.Carrara, E.Cavalli, *Previsione di insolvenza e reti neurali: un'applicazione su bilanci d'esercizio di una banca dati operativi multisettoriale*, QDMSIA, Università degli Studi di Bergamo, Anno 2001, N. 32.
- [Cavalli 2001] E.Cavalli, *Experimenting Neural Networks to forecast business insolvency*, Neural Network World, Numero 4, 2001.
- [Freeman 1984] E.Freeman, *Strategic Management: A stakeholder approach*, Boston: Pitman 1984.
- [Hertz et al. 1991] J.Hertz-Z.Krogh-R.G.Palmer, *Introduction to the Theory of Neural Computation*, Addison Wesley, 1991.
- [Lawrence 1993] J.Lawrence, *Neural Network: design, Theory and Applications*, California Scientific Software, 1993.
- [Marshall 1890] A.Marshall, *Principle of Economics*, MacMillan&Co., New York 1890.
- [NeuroSolutions] <http://www.neurosolutions.com>
- [Rojas 1996] R.Rojas, *Neural Networks*, Springer-Verlag, Berlin, 1996.
- [Sietsma-Dow 1988] J.Sietsma-R.Dow, *Neural Net Pruning, why and how*, Proceeding IEEE ICNN, July 1988.
- [SternStewart] <http://www.sternstewart.com/?content=proprietary&p=eva>
- [Zappa 1943] G.Zappa, *"Il Reddito d'impresa"*, Giuffrè, Milano 1943.