

UN'APPLICAZIONE DI ALGORITMI DI MACHINE LEARNING NEI PROCESSI DI EARLY WARNING: COMPARAZIONE DELLE PERFORMANCE RISPETTO ALLA MODELLIZZAZIONE TRADIZIONALE

DANIELE VERGARI*
CRISTINA CAPRARA*

Abstract

Le tecniche di machine learning non hanno ancora conosciuto un'ap-prezzabile diffusione nelle prassi del risk management Italiano, principal-mente per la scarsa trasparenza degli algoritmi non parametrici. Obiettivo di questo studio è presentare un'inda-gine comparativa dei risultati ottenuti nell'ambito dell'early warning, con-frontando l'applicazione di un modello logistico, di una random forest e di una rete neurale, per mettere in luce le po-tenzialità di queste nuove soluzioni e i vantaggi generabili dalla loro applica-zione nei processi di gestione della filie-ra creditizia.

Il contesto di riferimento

Il ricorso a tecniche di machine learning per la gestione dei processi di controllo e prevenzione del rischio di credito non ha conosciuto finora

* Crif

un'applicazione estensiva in ambito bancario per ragioni ormai ben note: la modellizzazione essendo di tipo non parametrico presenta alcune difficoltà interpretative; richiede competenze tecniche altamente specifiche; può presentare delle serie limitazioni nei sistemi AIRB soggetti a vigilanza regolamentare per la scarsa trasparenza degli algoritmi sottostanti alla misurazione del merito creditizio delle controparti.

Le tecniche parametriche classiche rendono infatti immediatamente interpretabili i singoli coefficienti, o pesi, assegnati a ciascuna variabile predittiva in relazione all'evento target del modello, tipicamente il passaggio a default. La possibilità di quantificare il peso informativo di ciascuna variabile nell'algoritmo di scoring è considerata spesso una caratteristica imprescindibile dai risk manager che hanno la responsabilità dello sviluppo e dell'applicazione dei modelli ed è anche una delle principali ostacole all'applicazione di soluzioni non parametriche.

D'altro canto le tecniche di machine learning hanno potenzialità ri-

levanti e possono utilmente supportare i processi di gestione del credito offrendo performance pari o superiori a quelle delle tecniche tradizionali e, soprattutto, caratterizzandosi per tempistiche di sviluppo e di aggiornamento dei modelli più veloci, fattore non trascurabile nella re-ingegnerizzazione dei processi creditizi.

La potenza di queste nuove tecnologie consente di abbattere i costi di analisi e di massimizzare il potere informativo dei big data disponibili sui sistemi informativi delle banche incrociandoli a quelli forniti dai provider esterni: anche il targeting della clientela può essere ottimizzato attraverso la costruzione di indicatori basati su tecniche non parametriche per supportare la personalizzazione dell'offerta e della gestione della relazione. Tra le nuove sfide tecnologiche del sistema bancario vi è oggi la capacità di fare innovazione sfruttando la PSD2: le tecniche di machine learning sono uno strumento assolutamente idoneo per raggiungere l'obiettivo e sostenere la competizione a livello globale.

L'obiettivo del case study

L'obiettivo di questo studio è quello di indagare comparativamente le diverse risultanze ottenute nell'ambito dell'early warning applicando una tecnica parametrica tradizionale – il logit – e due tecniche non parametri-

che alternative – la random forest e la rete neurale. L'analisi comparativa viene effettuata su un ampio campione di controparti business, estratte in forma anonima da Eurisc, il Sistema di Informazioni Creditizie di CRIF.

Su queste controparti, che al momento della rilevazione campionaria mostrano un comportamento creditizio regolare, si rileva il fenomeno di early warning, definito dalla tensione nella gestione del credito che si manifesta nei sei mesi successivi alla rilevazione attraverso il mancato adempimento degli obblighi contrattuali, come ad esempio il rimborso degli oneri finanziari già contratti oppure attraverso il ricorso alle linee di credito rotative in misura superiore ai massimali previsti. Il tasso di default che sintetizza a livello di singola impresa il manifestarsi di questi eventi di delinquency viene registrato sul campione su un orizzonte di sei mesi dalla data di rilevazione.

Per garantire una piena compar-

bilità dei risultati, le stime del logit, della rete neurale e della random forest vengono condotte ceteris paribus sullo stesso campione, adottando lo stesso set informativo ovvero la stessa lista di variabili predittive nonché la stessa definizione di default.

Per gli approfondimenti sulle tre soluzioni metodologiche si rimanda alla bibliografia in appendice; la regressione logistica è infatti ormai assunta a uno standard nelle pratiche di risk management mentre le random forest e la rete neurale, introdotte soltanto a partire dagli anni 2000, hanno potenzialità ancora poco esplorate nel risk management, come la capacità di cogliere le relazioni non lineari insite nei dati.

I principali risultati

La prima evidenza che emerge dal case study è che le tre tecniche offrono performance statistiche assolutamente in linea con le best practice della scoring industry. La tabella 1 mostra gli indici di Kolmogorff Smirnov, l'Accuracy Ratio e l'Area di Roc sul campione di validazione:

Tabella 1 – Analisi comparativa dei KPI

	Ks	Ar	Roc
Logit	53.5	67.3	83.7
Random forest	57.6	71.6	85.8
Neural Network	56.2	69.2	84.6

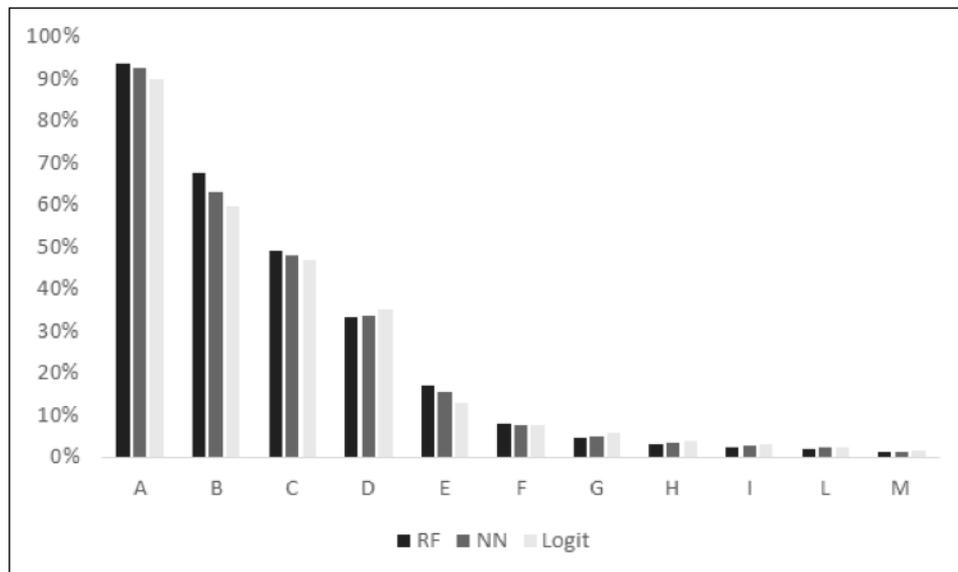
Nello specifico si evidenzia che la random forest presenta le performance migliori, con un accuracy ratio pari al 71.6%; a seguire la rete neurale, con un accuracy ratio pari al 69.2, e infine il logit, che si attesta a 67.3.

Il Grafico 1 evidenzia la distribuzione del default rate percentuale per classi di scoring di pari densità prodotte dai tre modelli di valutazione. Per garantire la massima comparabilità dei risultati si sono definite delle classi di scoring che hanno la stessa distribuzione percentuale sul campione per tutti e tre i modelli (dette classi iso-densità): l'analisi conferma

che la miglior capacità discriminante è offerta dalla random forest e, a seguire, dalla rete neurale ma si confer-

ma decisamente competitivo anche il logit.

Grafico 1 – Distribuzione del default rate per classi di rating



Volendo comparare l'approccio tradizionale a quello di machine learning, si nota che le valutazioni prodotte rispettivamente dal modello logit e dal modello random forest risultano complessivamente coerenti,

come rappresentato dalla matrice di transizione dei due modelli in Figura 2, che presenta una buona concentrazione sulla diagonale principale soprattutto sulle classi estreme.

Figura 2 - Confronto delle valutazioni prodotte dal logit (in colonna) e dalla random forest (in riga)

	A	B	C	D	E	F	G	H	I
A	82%	18%	0%	0%	0%	0%	0%	0%	0%
B	17%	70%	13%	0%	0%	0%	0%	0%	0%
C	0%	6%	49%	25%	16%	2%	1%	0%	0%
D	0%	0%	28%	35%	17%	8%	5%	3%	4%
E	0%	0%	12%	20%	22%	15%	9%	7%	15%
F	0%	0%	1%	12%	18%	26%	13%	10%	20%
G	0%	0%	0%	7%	13%	19%	15%	15%	31%
H	0%	0%	0%	3%	9%	14%	15%	20%	38%
I	0%	0%	0%	1%	4%	3%	9%	12%	72%

Quindi, se in termini di performance i tre modelli, seppur in diverso grado, possono essere considerati alternativi e competitivi, qualche differenza va segnalata per quanto concerne le modalità di sviluppo e fine tuning. A parità infatti di set informativo comune, una short list di circa 20 variabili ottenuta a partire da una long list di circa 200 già disponibili estratte dal Sistema di Informazioni Creditizie Eurisc, il modello logit è soggetto a maggiori vincoli imposti dalla metodologia. Le tecniche di machine learning, invece, in fase applicativa si caratterizzano per:

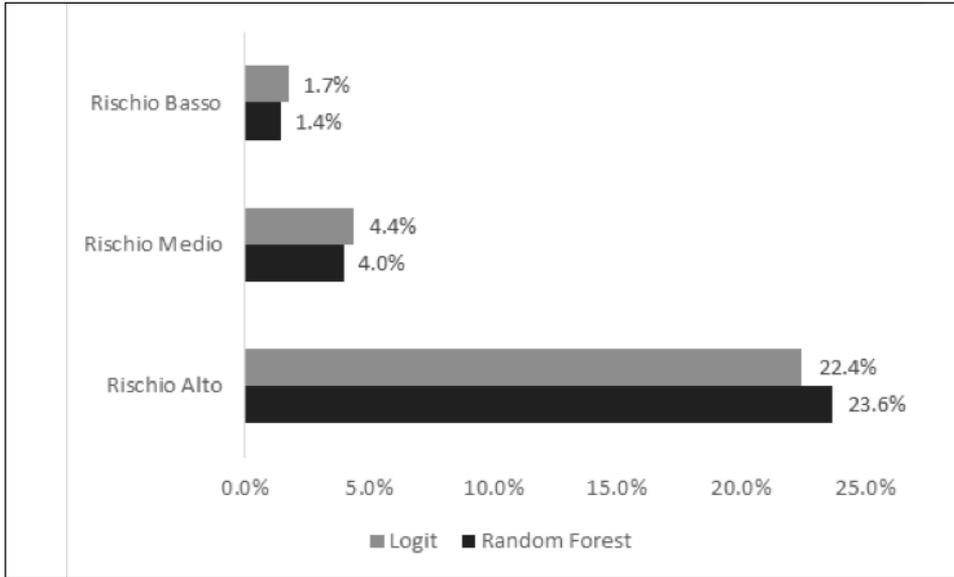
- Offrire maggiore flessibilità nel trattamento delle variabili
- Cogliere relazioni non lineari insite nei dati
- Massimizzare il valore di tutta l'informazione campionaria disponibile
- Velocizzare l'aggiornamento dei modelli su nuove basi campionarie

Nello specifico, la valutazione of-

ferta dalla random forest, che si basa su più di 100 alberi decisionali alternativi le cui valutazioni sulla singola impresa vengono mediate per produrre una valutazione finale complessiva, sembra essere la tecnica che offre i risultati migliori proprio perché non riducibile a un singolo modello statistico, bensì sintesi di una pluralità di alberi decisionali che sfruttano e massimizzano l'informazione portata da tutte le variabili predittive combinandole in modo alternativo, potenzialità che un singolo score non sempre può raggiungere.

In termini di controllo dell'esposizione residua e a rischio, i risultati risultano significativi e sono rappresentati nel grafico 2 che confronta il rapporto percentuale tra l'esposizione effettivamente insoluta nel periodo di performance e il valore dell'esposizione residua e a rischio al momento della rilevazione campionaria, divisi per macro fasce di rischio. La fascia a rischio alto intercetta il 23.6% di esposizione a rischio che sul periodo di performance passa a default, il logit ne intercetta il 22.4%.

Grafico 2 - Rapporto (%) tra l'esposizione insoluta osservata nel periodo di performance e l'esposizione residua e a rischio al momento della rilevazione per fasce di rischio

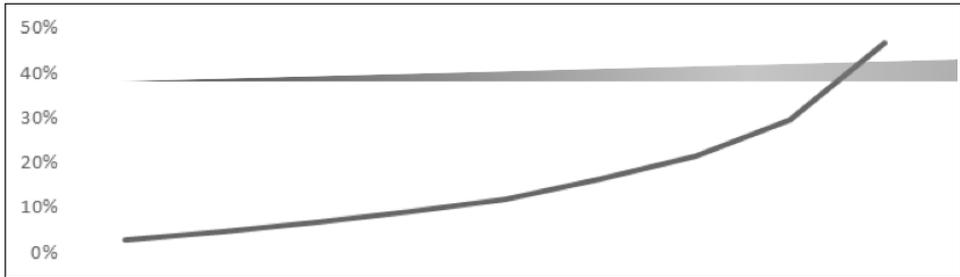


Simili risultati confermano i benefici potenzialmente derivabili dall'applicazione di tecniche di machine learning nei processi di gestione del credito e suggeriscono la ricerca di chiavi interpretative diverse, come ad esempio una reportistica descrittiva, che rendano il modello più trasparente e comprensibile agli utenti finali.

In tal senso un valido ausilio è dato dall'analisi descrittiva della probabilità sulle classi delle principali variabili predittive. La coerenza e la regolarità dell'andamento dei valori medi della

probabilità di default sulle classi delle variabili agevola notevolmente l'interpretazione e la comprensione delle caratteristiche del modello, in modo da rendere il risultato finale più intelligibile. Nello specifico, il Grafico 3 rappresenta la distribuzione della mediana della probabilità di default stimata dal modello random forest per grado di utilizzo delle linee di credito rotative: al crescere della tensione nel ricorso alle linee di fido, come atteso, il valore mediano della probabilità di default aumenta sensibilmente.

Grafico 3 – Andamento della mediana della probabilità di default (in ordinata) per intensità di utilizzo dei fidi (in ascissa)



Conclusioni

Lo studio realizzato da CRIF mette in luce le potenzialità delle tecniche di machine learning che offrono maggiore flessibilità in fase di sviluppo dei modelli, colgono relazioni non lineari insite nei dati, massimizzano il valore di tutta l'informazione campionaria disponibile, rendono molto rapido il fine tuning degli scoring su nuove basi campionarie.

Una diffusione più ampia e consapevole di questi strumenti nella gestione dei portafogli creditizi rappresenta un'opportunità che i risk manager possono cogliere per aumentare l'efficienza delle proprie

strutture, garantendo pari tempestività e accuratezza nella prevenzione del rischio di credito. Gli attuali vincoli normativi imposti dalla vigilanza regolamentare sulla trasparenza della modellistica AIRB non costituiscono di per sé un'ostacolo per limitare il ricorso a queste tecniche nella prassi gestionale di alcune particolari fasi del ciclo di vita del credito, come la collection, l'early warning, il targeting e, più in generale, per sostenere una competizione a livello globale basata sempre più sulla capacità di estrarre valore e conoscenza da grandi moli di dati.

Bibliografia

- Breiman, L. (2001). Random forests. *Mach Learn*, 45:5-32.
- Khandani, A.E., Kim, J., Lo, A.W. (2010). Consumer credit-risk models via machine-learning algorithms
- Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning scheme
- Kruppa, J., Schwarz, A., Arminger, G., Ziegler, A. (2013). Con-

- sumer credit risk : Individual probability estimates using machine learning
- Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G., & Ziegler, A. (2012). Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 51:74-81
 - Martin Riedmiller, Heinrich Braun (1993). A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. *IEEE International Conference On Neural Networks*. Vol. 16, pp.586—591
 - Nembrini, S., Koenig, I. R. & Wright, M. N. (2018). The revival of the Gini Importance? *Bioinformatics*
 - Tsai, C.F., Chen, M.L. (2010). Credit rating by hybrid machine learning techniques
 - Wright, M.N. & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R.
 - Zhao, Z., Xu, S., Kang, B.H., Kabir, M.J., Liu, Y., Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring