



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in Data Science per l’Economia e le Imprese

Previsione delle crisi finanziarie tramite
tecniche di Deep Learning

Prediction of financial crises using Deep
Learning techniques

Relatore:

Prof. Domenico Potena

Tesi di Laurea di:

Jacopo Bonanno

Correlatori:

Prof.ssa Claudia Pigni

Prof. Alex Mircoli

A.A. 2023/2024

Indice

Introduzione	1
Capitolo 1 Early Warning Systems.....	5
1.1 Definizione e Obiettivi degli Early Warning Systems	5
1.2 Applicazione degli EWS nel Contesto Finanziario	7
1.3 Componenti Chiave degli EWS Finanziari.....	9
1.4 Sfide e Limiti degli EWS	11
1.5 Letteratura e trend sugli Early Warning Systems.....	13
Capitolo 2 Dati e Metodologia.....	25
2.1 Definizione del Dataset	25
2.2 Analisi Esplorativa	28
2.3 Preprocessing	34
2.4 Reti Neurali.....	36
2.5 Reti Neurali Ricorrenti.....	42
2.6 Reti Neurali LSTM (Long-Short Term Memory).....	43
2.7 K-means.....	48
2.8 Problema dei dati sbilanciati	50
2.9 Oversampling	51
2.10 Undersampling	53
2.11 Metacost	54
Capitolo 3 Esperimenti e Risultati.....	58
3.1 Grid Search	58
3.2 Valutazione dei modelli.....	61
3.3 Risultati.....	65
Conclusione	76
Bibliografia.....	80

Introduzione

Nel corso degli ultimi decenni, il settore finanziario globale ha vissuto una serie di crisi che hanno avuto impatti devastanti sull'economia mondiale, causando perdite significative non solo per le istituzioni finanziarie, ma anche per cittadini e imprese. La crisi finanziaria del 2008, in particolare, ha messo in luce la vulnerabilità dei mercati finanziari e ha sottolineato l'importanza di strumenti adeguati al monitoraggio e previsione di eventi destabilizzanti. È in questo contesto che si inserisce la necessità di sviluppare tecniche avanzate per la previsione delle crisi finanziarie, allo scopo di anticipare e mitigare le conseguenze di tali eventi. L'obiettivo principale di questa tesi è di esplorare e implementare tecniche di Deep Learning per la previsione delle crisi finanziarie. Il Deep Learning, con la sua capacità di analizzare grandi quantità di dati e riconoscere pattern complessi, rappresenta una promettente soluzione per migliorare l'accuratezza delle previsioni e fornire avvertimenti tempestivi che possano ridurre l'impatto delle crisi.

La previsione delle crisi finanziarie è una sfida significativa a causa della complessità e della natura dinamica dei mercati finanziari. Gli approcci

tradizionali basati su modelli statistici ed econometrici hanno dimostrato dei limiti nella capacità di catturare le interazioni non lineari tra le variabili e di adattarsi rapidamente ai cambiamenti del mercato.

Nel primo capitolo di questa tesi verranno introdotti i Sistemi di Allerta Precoce (Early Warning Systems) e il loro ruolo nel prevedere eventi avversi in vari contesti, con un focus particolare sul settore finanziario. Gli EWS sono strumenti progettati per monitorare e analizzare una serie di indicatori economici e finanziari al fine di prevedere l'insorgere di crisi e fornire avvertimenti tempestivi. Il secondo capitolo si concentrerà sulle tecniche di Deep Learning, introducendo le reti neurali e, in particolare, le Long Short-Term Memory (LSTM), una specifica architettura di rete neurale ricorrente che si è dimostrata particolarmente efficace nella modellazione di dati sequenziali. Le LSTM sono in grado di mantenere informazioni per lunghi periodi e sono quindi adatte a catturare le dipendenze temporali nei dati finanziari, che sono cruciali per la previsione delle crisi. Verrà anche descritto il dataset utilizzato, che comprende dati storici finanziari, inclusi indicatori macroeconomici, prezzi delle azioni, tassi di interesse, tassi di cambio e altri indicatori rilevanti.

Nell'ultimo capitolo si presenteranno i risultati ottenuti dall'applicazione dei modelli ai dati storici e verrà condotta un'analisi approfondita delle

performance dei modelli. Verranno anche discussi i risultati in termini di accuratezza e di altre metriche.

Questo studio è di fondamentale importanza poiché mira a colmare il divario esistente tra le tecniche tradizionali di previsione finanziaria e le potenzialità offerte dal Deep Learning. L'integrazione di queste tecniche può non solo migliorare la capacità di prevedere crisi future, ma anche offrire strumenti più robusti per la gestione del rischio finanziario. L'accuratezza delle previsioni è essenziale per consentire alle istituzioni finanziarie, ai governi e agli investitori di prendere decisioni informate e tempestive, riducendo così il rischio di perdite significative.

Capitolo 1

Early Warning Systems

1.1 Definizione e Obiettivi degli Early Warning Systems

Gli Early Warning Systems (EWS), noti in italiano come Sistemi di Allerta Precoce, sono strumenti di monitoraggio e analisi progettati per prevedere l'insorgere di crisi, eventi avversi o situazioni di emergenza in vari contesti, tra cui quelli ambientali, sanitari e, in particolare, finanziari. Il loro scopo principale è identificare i segnali premonitori di potenziali crisi e fornire avvertimenti tempestivi, consentendo così alle autorità competenti, alle istituzioni e agli individui di adottare misure preventive o di mitigazione per ridurre l'impatto negativo di tali eventi. Questi sistemi si basano su una combinazione di tecnologie avanzate e metodologie analitiche per raccogliere e interpretare dati rilevanti provenienti da diverse fonti.

Ad esempio, nei contesti ambientali, gli EWS monitorano variabili climatiche e geologiche come precipitazioni, temperature, livelli dei fiumi e attività sismica per prevedere disastri naturali quali alluvioni, siccità,

terremoti e tsunami. Attraverso l'uso di sensori, satelliti e modelli climatici, è possibile ottenere previsioni accurate che possono salvare vite umane e proteggere infrastrutture critiche. L'applicazione di tali segnalatori nella creazione di un EWS per il cambiamento climatico viene descritta nel libro di Zinta Zommers e Ashbindu Singh, *Reducing Disaster: Early Warning Systems for Climate Change* (2014).

Nel settore sanitario, gli EWS analizzano dati epidemiologici, rilevamenti di laboratorio, e tendenze di malattie per anticipare epidemie o pandemie. L'integrazione di dati provenienti da ospedali, cliniche e organizzazioni sanitarie internazionali permette di identificare rapidamente focolai di malattie infettive, permettendo un intervento rapido ed efficace per contenere la diffusione di malattie. Questi sistemi sono stati cruciali, ad esempio, durante l'epidemia di Ebola e la pandemia di COVID-19, dimostrando l'importanza della sorveglianza sanitaria globale (Kamalrathne T, Amaratunga D, Haigh R, Kodituwakku L., 2023)).

Nel contesto finanziario, gli EWS sono progettati per individuare i segnali di instabilità economica e crisi bancarie (Casabianca et al., 2022)). Utilizzando modelli statistici e algoritmi di machine learning, questi sistemi analizzano indicatori economici chiave, quali tassi di interesse, inflazione, tassi di cambio e performance dei mercati azionari. L'obiettivo è rilevare

anomalie e tendenze che potrebbero indicare l'approssimarsi di una crisi finanziaria, permettendo ai governi e alle istituzioni finanziarie di intervenire con misure correttive prima che si verifichi un collasso.

1.2 Applicazione degli EWS nel Contesto Finanziario

Nel settore finanziario, gli EWS sono utilizzati per prevedere crisi economiche, instabilità dei mercati, e altri fenomeni che potrebbero avere ripercussioni significative sull'economia globale o su specifici segmenti di mercato. Questi sistemi analizzano una vasta gamma di dati economici, finanziari e geopolitici per individuare pattern e anomalie che possano indicare un imminente rischio di crisi.

L'importanza degli EWS in ambito finanziario è amplificata dalla complessità e dall'interconnessione dei mercati moderni, dove eventi in una parte del mondo possono rapidamente propagarsi e influenzare altre economie. La crisi finanziaria globale del 2008 ha evidenziato la necessità di strumenti più efficaci per monitorare e prevedere tali eventi, portando a un crescente interesse nello sviluppo e nell'implementazione di EWS avanzati.

Gli EWS finanziari si basano su modelli econometrici e algoritmi statistici che integrano dati provenienti da diverse fonti, tra cui mercati azionari, tassi di cambio, indici di produzione industriale, tassi di interesse e

indicatori macroeconomici come il PIL, il debito pubblico e il tasso di disoccupazione. Inoltre, questi sistemi possono includere analisi qualitative, come valutazioni del rischio politico e instabilità sociale, che potrebbero influenzare le dinamiche di mercato.

Uno degli aspetti chiave degli EWS nel settore finanziario è la loro capacità di fornire avvisi tempestivi, permettendo alle istituzioni finanziarie, ai governi e agli investitori di prendere decisioni informate per mitigare i rischi. Ad esempio, un EWS potrebbe segnalare un aumento anomalo del debito privato in un paese, suggerendo un potenziale rischio di crisi bancaria. In risposta, le autorità potrebbero intervenire con politiche monetarie per stabilizzare l'economia.

L'implementazione degli EWS finanziari non è priva di sfide. La complessità dei mercati finanziari e la rapidità con cui possono cambiare le condizioni economiche richiedono sistemi altamente adattabili e in grado di aggiornarsi continuamente con nuovi dati. Inoltre, la qualità dei dati utilizzati è cruciale: dati incompleti o inaccurati possono portare a previsioni errate e a decisioni subottimali.

Un altro aspetto critico è la comunicazione dei risultati degli EWS. Le previsioni e gli avvertimenti devono essere chiari e comprensibili per poter essere utilizzati efficacemente dai decision maker. Questo richiede non solo

un'interfaccia utente intuitiva, ma anche una formazione adeguata per gli utenti finali affinché possano interpretare correttamente le informazioni fornite e agire di conseguenza.

Gli EWS finanziari possono anche svolgere un ruolo importante nella regolamentazione e supervisione del sistema finanziario. Le autorità di regolamentazione possono utilizzare questi sistemi per identificare vulnerabilità strutturali e rischi sistemici, migliorando così la resilienza del sistema finanziario. Ad esempio, monitorando costantemente le istituzioni finanziarie, gli EWS possono aiutare a individuare e prevenire comportamenti rischiosi che potrebbero portare a una crisi.

1.3 Componenti Chiave degli EWS Finanziari

Gli EWS finanziari sono composti da diverse componenti essenziali che lavorano insieme per raccogliere, elaborare e interpretare i dati. Tra queste componenti troviamo:

1. **Raccolta Dati:** La raccolta di dati è la fase iniziale e fondamentale degli EWS. I dati possono provenire da una varietà di fonti, tra cui mercati finanziari, bilanci aziendali, indicatori macroeconomici, notizie economiche e politiche, e altri eventi rilevanti. La qualità e la tempestività dei dati raccolti sono cruciali per l'accuratezza delle previsioni. Oltre alle fonti tradizionali, gli EWS possono integrare

dati provenienti da nuove fonti come i social media e le piattaforme di notizie in tempo reale, che offrono una prospettiva immediata sulle percezioni del mercato e sugli sviluppi globali.

2. **Elaborazione dei Dati:** Una volta raccolti, i dati devono essere elaborati e analizzati. Questo processo può includere la pulizia dei dati per rimuovere errori e incongruenze, la normalizzazione per garantire coerenza e la trasformazione per adattare i dati alle esigenze dei modelli predittivi. Strumenti di data mining e tecniche di machine learning sono spesso impiegati in questa fase per estrarre informazioni rilevanti e individuare pattern nascosti. L'elaborazione dei dati può anche coinvolgere l'uso di algoritmi per l'aggregazione dei dati da diverse fonti e la loro integrazione in un formato unico e coerente.

3. **Modellizzazione Predittiva:** Il cuore degli EWS è rappresentato dai modelli predittivi utilizzati per analizzare i dati e generare previsioni. Questi modelli possono variare da approcci statistici tradizionali, come le regressioni e le analisi delle serie temporali, a tecniche di machine learning avanzate. Gli approcci più sofisticati permettono di catturare relazioni complesse e non lineari nei dati, migliorando l'accuratezza delle previsioni. La modellizzazione predittiva può

anche includere la validazione e il test dei modelli per assicurare che siano affidabili e robusti nelle diverse condizioni di mercato.

4. **Generazione di Allarmi:** Basandosi sulle analisi effettuate dai modelli predittivi, gli EWS generano allarmi o avvisi quando vengono identificati segnali di possibili crisi. Questi allarmi possono essere configurati per diversi livelli di gravità, fornendo indicazioni chiare sulle potenziali conseguenze e sulle azioni consigliate. Gli allarmi possono essere personalizzati in base alle esigenze degli utenti, permettendo una risposta tempestiva e adeguata.

1.4 Sfide e Limiti degli EWS

Nonostante i progressi tecnologici e metodologici, gli Early Warning Systems (EWS) affrontano numerose sfide e limitazioni. Tra le principali difficoltà vi sono:

Qualità dei Dati: La precisione delle previsioni dipende fortemente dalla qualità dei dati utilizzati. Dati incompleti, obsoleti o inaccurati possono portare a previsioni errate, compromettendo l'efficacia degli EWS. Inoltre, la raccolta di dati di alta qualità può essere costosa e richiedere risorse significative, soprattutto in contesti con infrastrutture di monitoraggio limitate.

Complessità dei Modelli: I modelli predittivi, specialmente quelli basati su machine learning e intelligenza artificiale, possono essere estremamente complessi e difficili da interpretare. Questa complessità può rappresentare un ostacolo alla loro adozione e utilizzo efficace. Gli utenti finali, come i decisori politici e gli operatori di mercato, potrebbero trovare difficoltà a comprendere e fidarsi delle previsioni se i modelli non sono trasparenti.

Rischio di Falsi Positivi e Negativi: Gli EWS possono generare falsi positivi, segnalando crisi inesistenti, o falsi negativi, mancando di prevedere eventi reali. Entrambi i casi possono avere conseguenze significative per le decisioni basate sugli allarmi generati. I falsi positivi possono causare allarmi inutili, con conseguenti costi economici e perdita di fiducia nel sistema, mentre i falsi negativi possono impedire interventi tempestivi, esponendo a rischi maggiori.

Adattabilità: I mercati finanziari e altri contesti di applicazione degli EWS sono dinamici e in continua evoluzione. Gli EWS devono essere costantemente aggiornati e adattati per tenere conto di nuove variabili e cambiamenti nelle dinamiche di mercato. Questa necessità di aggiornamento continuo richiede flessibilità nei modelli e nei sistemi di raccolta dati, oltre a una manutenzione regolare e a una revisione critica delle metodologie utilizzate.

Disponibilità delle Risorse: L'implementazione e la manutenzione di EWS efficaci richiedono risorse finanziarie, tecnologiche e umane significative. In molti paesi, specialmente quelli in via di sviluppo, la mancanza di queste risorse può limitare l'efficacia dei sistemi di allerta precoce.

Coordinamento e Cooperazione: Gli EWS spesso richiedono la collaborazione tra diverse agenzie governative, organizzazioni internazionali e settori privati. La mancanza di coordinamento e comunicazione tra questi attori può compromettere la raccolta e la condivisione dei dati, riducendo l'efficacia complessiva del sistema.

Privacy e Sicurezza dei Dati: La raccolta e l'analisi di grandi quantità di dati sensibili sollevano preoccupazioni riguardanti la privacy e la sicurezza dei dati. È essenziale garantire che i dati siano protetti da accessi non autorizzati e che vengano rispettate le normative sulla privacy, al fine di mantenere la fiducia degli individui e delle istituzioni coinvolte.

Interpretazione e Decision-Making: Anche con dati di alta qualità e modelli accurati, l'interpretazione dei risultati e la traduzione in decisioni concrete rappresentano una sfida. I decisori devono essere in grado di comprendere le implicazioni delle previsioni e agire rapidamente, bilanciando la necessità di intervento con i potenziali costi e benefici.

1.5 Letteratura e trend sugli Early Warning Systems

La letteratura sugli Early Warning System sebbene ampia, si è concentrata nel tempo su tre principali approcci al problema:

- **Metodo dell'estrazione di segnale:** questo metodo viene introdotto a fine anni 90, Kaminsky et al. (1998), consiste in un approccio non parametrico volto all'analisi di variabili macroeconomiche e del loro pattern nei momenti che precedono la crisi. L'obiettivo è quello di andare a definire delle soglie, che una volta superate segnalano il possibile avvento di una crisi nei prossimi 12-24 mesi. Gli indicatori utilizzati sono 105 e vengono categorizzati in più gruppi, tra cui: settore esterno, settore finanziario, settore reale, pubbliche finanze, variabili istituzionali e strutturali, variabili politiche e "effetti contagio". L'applicazione di tale modello consiste nel semplice conteggio degli indicatori che superano il proprio valore soglia, e sulla base di questo numero stabilire l'effettiva possibilità di una crisi futura. Il problema di tale approccio è che manca la possibilità di aggregazione dei segnali, andando a considerare tutte le variabili in maniera singola, questo potrebbe portare ad una situazione poco chiara in cui alcune variabili segnalano un pericolo mentre altre no (problema dei segnali contraddittori) in quanto verrebbe persa gran parte delle informazioni dovuta all'interazione di queste (proprietà olistica del sistema).

- Il **BCT** (Binary Crisis Tree) è un modello basato su alberi decisionali, ovvero strutture ad albero che estrapolano delle regole, tramite criteri di split, con le quali andare a dividere in modo ripetuto i dati (Duttagupta and Cashin, 2008). Questo approccio viene particolarmente utilizzato nel mondo bancario perché riesce a superare il problema del precedente metodo dei segnali, non andando più a considerare l'incidenza del singolo indicatore bensì andando ad osservare il quadro completo. Il BCT riconosce inoltre la natura non lineare dell'influenza degli indicatori economici sulla probabilità di crisi. In altre parole, non ogni aumento o diminuzione di un indicatore chiave porta necessariamente ad una crisi finanziaria. Solo se il valore di un indicatore supera una certa soglia critica, il rischio di crisi diventa significativo. I vantaggi di questo metodo sono quindi:
 - Approccio multifattoriale: Permette di considerare l'insieme delle vulnerabilità che possono portare ad una crisi bancaria, offrendo una visione più completa del fenomeno.
 - Non linearità: Riconosce che l'impatto degli indicatori economici sulla probabilità di crisi non è sempre lineare, fornendo un'analisi più accurata.

- Identificazione delle soglie di rischio: Individua i valori critici degli indicatori economici che superati aumentano significativamente il rischio di crisi

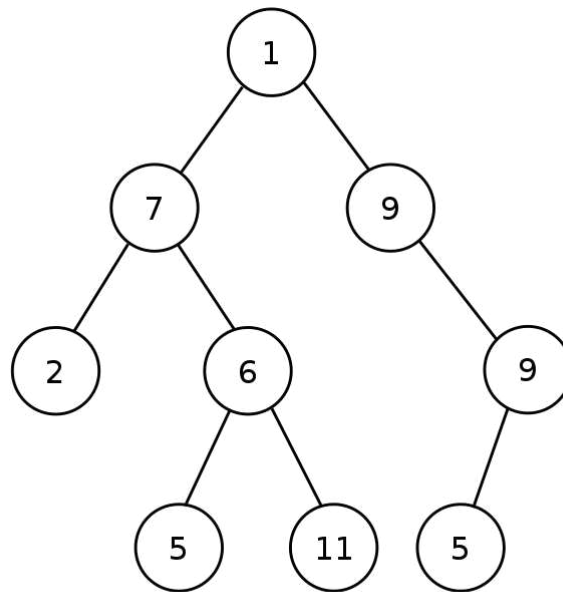


Fig 1.1 Struttura ad albero

- **Logit e Probit:** Nel campo degli EWS per la previsione delle crisi finanziarie, i modelli logit e probit sono approcci statistici ampiamente impiegati, che utilizzano strumenti della microeconometria per stimare la probabilità di eventi avversi. In questo ambito, la probabilità di una crisi viene modellata come funzione di vari predittori chiave. L'analisi dei coefficienti stimati e dei loro segni offre preziose informazioni sulle relazioni tra la variabile dipendente (crisi) e i regressori. Demirgüç-Kunt e

Detragiache (1998) hanno applicato con successo questa metodologia ad un ampio campione di paesi sviluppati e in via di sviluppo tra il 1980 e il 1994. I loro risultati mostrano che l'alta inflazione e la bassa crescita del PIL sono i principali determinanti delle crisi bancarie. Nonostante la natura non strutturale macroeconomica di questi modelli, essi permettono un'interpretazione economica significativa del legame tra la variabile dipendente e i predittori grazie all'analisi dei coefficienti stimati. Lo studio di Demirgüç-Kunt e Detragiache (1998) conferma l'efficacia predittiva dei modelli logit, dimostrando una percentuale di previsione corretta delle crisi pari al 70%.

Diversi studi hanno esplorato l'utilizzo di modelli econometrici al fine di migliorare l'accuratezza delle previsioni. Tra questi, Antunes et al. (2018) si sono concentrati sui modelli probit, introducendo l'utilizzo dei lag delle variabili. I risultati ottenuti dimostrano che l'inserimento di queste informazioni all'interno del modello permette di ottenere una migliore performance predittiva rispetto al modello probit standard, aumentando il numero di crisi correttamente previste.

Un approccio alternativo è stato proposto da Caggiano et al. (2016), che impiegano modelli multinomial logit. In questo caso, la variabile dipendente assume una struttura più complessa, includendo non solo gli stati di crisi e non crisi, ma anche l'anno successivo all'evento di crisi. Tale scelta metodologica consente di catturare informazioni più dettagliate sull'evoluzione del sistema finanziario, distinguendo tra la fase iniziale acuta della crisi e le sue fasi successive. Gli autori dimostrano che il modello multinomial logit risulta particolarmente efficace in contesti caratterizzati da crisi di durata media prolungata, evidenziando un miglioramento rispetto ai modelli logit binomiali tradizionali, che non distinguono tra le diverse fasi della crisi.

Altri contributi rilevanti in questo campo sono quelli di Schularick e Taylor (2012), che hanno sviluppato due tipologie di modelli: un modello pooled logit e un modello logit ad effetti fissi. I risultati ottenuti dagli autori mostrano che il modello ad effetti fissi offre performance predittive migliori rispetto al pooled logit. Tuttavia, quest'ultimo modello rimane più diffuso per ragioni econometriche legate alla robustezza delle stime.

I modelli econometrici sono ampiamente utilizzati per la loro interpretabilità e robustezza statistica, caratteristiche che li rendono

strumenti affidabili per analizzare le relazioni tra variabili economiche. Questi modelli, come i logit e probit, consentono agli economisti di derivare significati chiari dai coefficienti stimati, facilitando l'interpretazione delle dinamiche economiche sottostanti.

Tuttavia, questi modelli incontrano difficoltà significative quando si tratta di catturare relazioni non lineari, che sono frequentemente presenti negli EWS per la previsione delle crisi finanziarie. Le crisi economiche e finanziarie sono spesso il risultato di interazioni complesse e non lineari tra molteplici fattori, rendendo i modelli econometrici tradizionali meno efficaci in questi contesti.

Per affrontare queste complessità, si stanno sempre più adottando modelli di machine learning. Questi modelli offrono una maggiore flessibilità nel modellare relazioni non lineari e possono gestire grandi quantità di dati con molte variabili, adattandosi meglio alle dinamiche intricate dei mercati finanziari. Algoritmi come le reti neurali, le macchine a vettori di supporto (SVM) e le foreste casuali (random forests) sono particolarmente efficaci nel catturare pattern complessi e nel migliorare le capacità predittive degli EWS.

Nonostante la loro capacità di adattarsi meglio a dinamiche complesse, i modelli di machine learning presentano però una significativa limitazione:

la scarsa interpretabilità. Questa mancanza di trasparenza, spesso descritta come il problema della "black box", può rappresentare un ostacolo per la comprensione dei meccanismi sottostanti le previsioni e per la fiducia nelle decisioni basate su tali modelli. Gli economisti e i policy maker possono trovare difficile fidarsi di previsioni la cui logica interna non è facilmente comprensibile, limitando così l'adozione di questi strumenti in ambiti dove la trasparenza e l'interpretazione sono cruciali. Pertanto, mentre i modelli di machine learning rappresentano una promettente evoluzione nel campo degli EWS, è fondamentale sviluppare tecniche che migliorino la loro interpretabilità per rendere queste potenti metodologie più accessibili e utili nel contesto delle decisioni economiche e finanziarie.

Diversi studi hanno dimostrato la superiorità del machine learning rispetto ai modelli logit in termini di capacità predittiva out-of-sample, soprattutto in contesti caratterizzati da relazioni non lineari complesse tra le variabili economiche e finanziarie. Casabianca et al. (2022) hanno mostrato come i Random Forest, un algoritmo di machine learning, offrano una migliore accuratezza predittiva rispetto ai modelli logit, in particolare quando si analizzano dati di 55 crisi bancarie in 33 economie avanzate e 87 crisi in 67 economie emergenti. I risultati evidenziano che il machine learning è in grado di catturare meglio le relazioni non lineari che spesso caratterizzano i

fenomeni economici e finanziari, migliorando la capacità di prevedere eventi come le crisi bancarie.

Questo vantaggio è stato confermato anche da Holopainen e Sarlin (2017), che hanno dimostrato come il machine learning sia in grado di ridurre l'errore di classificazione, soprattutto per la classe minoritaria (le crisi bancarie). Questo aspetto è fondamentale per gli EWS, in quanto permette di identificare correttamente un'elevata percentuale di crisi, evitando i cosiddetti "falsi negativi" che possono avere conseguenze gravi per il sistema finanziario.

Tuttavia, alcuni studiosi sostengono che i modelli logit possano presentare una maggiore stabilità predittiva, soprattutto quando si utilizzano campioni di dati di grandi dimensioni. Beutel et al. (2019), analizzando dati di economie avanzate tra il 1970 e il 2016, hanno concluso che i modelli logit, se addestrati tramite recursive cross validation, offrono previsioni out-of-sample più stabili rispetto al machine learning, anche utilizzando diverse variabili e scenari. Questo risultato suggerisce che i modelli logit potrebbero essere più adatti in contesti in cui la stabilità delle previsioni è un requisito fondamentale, come ad esempio per la definizione di politiche di regolamentazione finanziaria.

Per superare il limite dell'interpretabilità di questi modelli, Bluwstein et al. (2013) hanno proposto un approccio che integra il machine learning con le teorie di Shapley (1953). Questo metodo permette di attribuire un'importanza relativa a ciascuna variabile utilizzata dal modello nella previsione di un evento, offrendo una maggiore interpretazione delle decisioni prese dal modello. Questo approccio potrebbe contribuire ad aumentare la fiducia e l'adozione del machine learning negli EWS.

Un ulteriore campo di ricerca nel campo degli EWS è lo sviluppo di modelli in grado di prevedere le crisi finanziarie con un anticipo significativo. Tölö (2020) ha sviluppato modelli basati su reti neurali ricorrenti (RNN), in particolare RNN-LSTM (Long Short Term Memory) e RNN-GRU (Gated Recurrent Unit), che sono in grado di prevedere crisi finanziarie sistemiche con un anticipo da uno a cinque anni. Le sue analisi dimostrano che le RNN superano significativamente le performance dei modelli di regressione logistica tradizionali nella previsione di crisi finanziarie. Barthelemy (2024) esplora l'implementazione di EWS per crisi valutarie utilizzando modelli LSTM e GRU, confrontandoli con regressioni logistiche e foreste casuali. I risultati dimostrano che le RNN avanzate superano i modelli tradizionali in termini di precisione e continuità dei segnali, avvertendo 20 delle 22 crisi del campione di test, inoltre cerca di

superare il problema della “black box” applicando gli SHAP values (SHapley Additive exPlanations). Fioramanti (2008) conferma la superiorità delle reti neurali artificiali (ANN) nella previsione delle crisi del debito sovrano, mostrando una maggiore capacità predittiva rispetto ai metodi parametrici e non parametrici. Ristolainen (2018) sottolinea l'importanza della non linearità e dell'eterogeneità regionale nell'EWS per le crisi bancarie, evidenziando come le ANN, supportate da dataset regionali, migliorino significativamente le previsioni rispetto alla regressione logit.

Questo risultato suggerisce che le RNN potrebbero essere una promettente tecnologia per lo sviluppo di sistemi di preavviso anticipato delle crisi bancarie. Successivamente Jin e Lin (2023), presentano un modello interpretabile di rete neurale ricorrente a memoria a lungo termine (IMV-LSTM) con focal loss (FL) per mitigare l'asimmetria tra le classi nel dataset. I risultati sono notevoli non solo per il tentativo di rendere comprensibile il modello, ma anche per il 100% di Recall ottenuto sul set di test. In questo set, oltre alla presenza di variabili economico-finanziarie, è stato incluso anche l'indicatore ESG (Environmental, Social and Governance) per ciascuna nazione.

Capitolo 2

Dati e Metodologia

2.1 Definizione del Dataset

Nel presente capitolo, verranno analizzate le tecniche di Machine Learning e Deep Learning applicate agli EWS, utilizzando il dataset dataset delle crisi finanziarie fornito da Laeven e Valencia (2018), abbinato ad indicatori macroeconomici disponibili come World Development Indicators (WDI) e International Financial Statistics (IFS), già utilizzato da Pignini (2021). I dati presentano una struttura panel, ovvero raccolgono informazioni relative a 129 paesi per un periodo di 34 anni, che va dal 1984 al 2017. In totale, il dataset comprende 4.386 osservazioni.

Variabile dipendente:

- **Crisi bancaria in corso (Current crisis):** Variabile binaria che assume valore 1 se nel paese, nell'anno considerato, si verifica una crisi bancaria, e 0 in caso contrario. La definizione di crisi bancaria

adottata segue quella di Laeven e Valencia (2018), che identificano due condizioni principali:

- evidenti segnali di difficoltà finanziarie nel sistema bancario, come massicci prelievi bancari, perdite significative o liquidazioni di banche;
- interventi di politica bancaria significativi in risposta a perdite rilevanti nel sistema stesso.

Le variabili esplicative:

- **Crescita del Deflatore del PIL (IFS):** Misura della crescita del deflatore del Prodotto Interno Lordo (PIL) basata sui dati dell'International Financial Statistics (IFS). Questa variabile rappresenta il tasso di crescita dei prezzi di beni e servizi in un'economia, depurato dagli effetti dell'inflazione.
- **Crescita del Tasso di Cambio del Dollaro USA (IFS):** Tasso di crescita del tasso di cambio del dollaro statunitense secondo IFS. Questa variabile riflette le variazioni nel valore del dollaro rispetto ad altre valute, influenzando il commercio internazionale e la competitività economica.
- **Crescita del Credito Privato Domestico (WDI):** Tasso di crescita del credito privato domestico basato sui dati dei WDI. Misura

l'espansione del credito concesso al settore privato dalle istituzioni finanziarie domestiche, un indicatore della disponibilità di finanziamenti per le attività economiche interne.

- **Crescita Reale del PIL (WDI):** Tasso di crescita reale del PIL secondo i dati WDI. Rappresenta la crescita economica annuale, depurata dagli effetti dell'inflazione, mostrando l'aumento del valore economico di beni e servizi prodotti in un paese.
- **Logaritmo del PIL Pro Capite (WDI):** Logaritmo del Prodotto Interno Lordo (PIL) pro capite secondo i dati dei WDI. Questa variabile cattura la ricchezza media per abitante, fornendo una misura del benessere economico su base logaritmica per stabilizzare la variabilità nei dati.
- **Rapporto tra Offerta di Moneta (M2) e Riserve Estere (WDI):** Rapporto tra l'offerta di moneta (M2) e le riserve valutarie estere secondo i dati dei WDI. Indica la capacità di un paese di sostenere l'offerta di moneta rispetto alle sue riserve estere, un fattore critico per la stabilità finanziaria e la resilienza agli shock esterni.
- **Tasso di Interesse Reale (IFS):** Tasso di interesse reale secondo l'IFS. Questo tasso rappresenta il rendimento effettivo degli investimenti, depurato dall'inflazione, ed è un indicatore chiave della politica monetaria e delle condizioni finanziarie.

- **Rapporto tra Attività Nette Estere e PIL (WDI):** Rapporto tra attività nette estere e Prodotto Interno Lordo PIL secondo i dati dei WDI. Questa variabile valuta la posizione finanziaria esterna di un paese, riflettendo la sua esposizione al rischio di valuta, la sostenibilità del debito estero e la dipendenza dagli investimenti stranieri.

2.2 Analisi Esplorativa

In questa sezione è riportata l'esplorazione delle variabili del dataset scelto. Il primo grafico mostra una mappa dove i singoli Paesi sono colorati sulla base del numero di anni di crisi che hanno registrato, i colori più scuri indicano un maggior numero di crisi, mentre i colori più chiari indicano un minor numero di crisi o nessuna crisi (i Paesi in Bianco sono quelli su cui non si hanno dati disponibili).

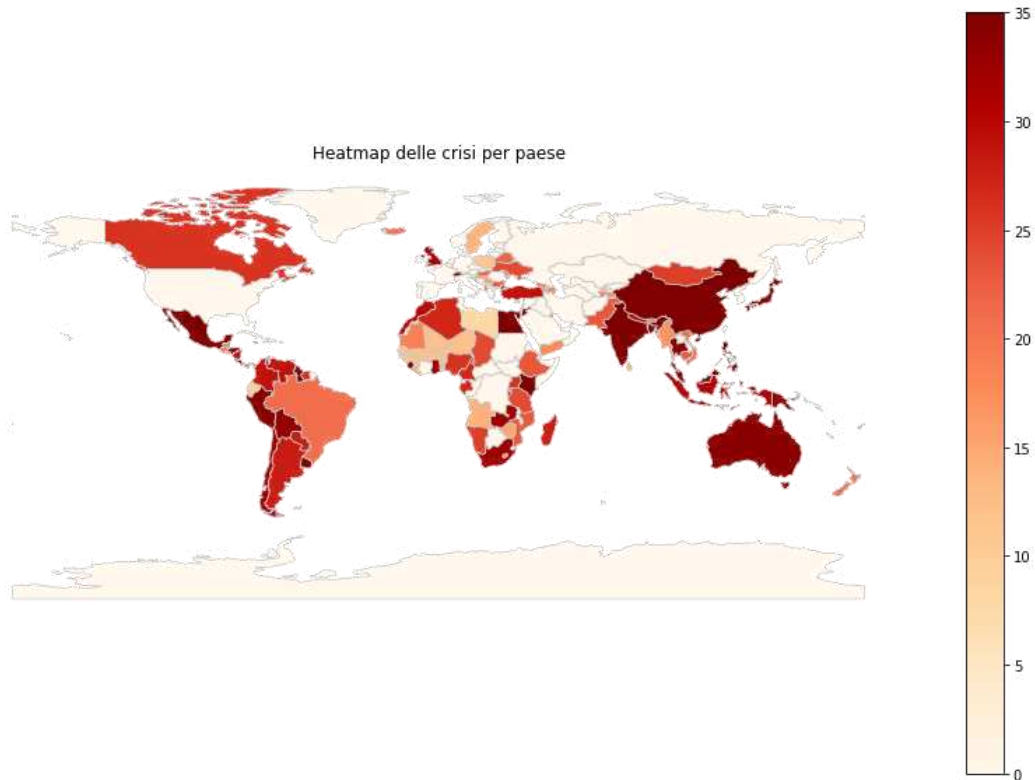


Fig 2.1 Heatmap Paesi per numero di crisi

Di seguito sono stati utilizzati dei box plot per rappresentare le variabili utilizzate. Questi grafici rappresentano graficamente le distribuzioni attraverso una serie di elementi chiave. La mediana è rappresentata come una linea orizzontale all'interno di un rettangolo, che costituisce l'intervallo interquartile. Questo intervallo contiene il 50% dei dati centrali, mentre le linee verticali esterne, chiamate "baffi", si estendono fino ai valori adiacenti più alti e più bassi. Gli outlier, ossia i punti dati che cadono al di fuori dei baffi, sono individuati e rappresentati con simboli distinti, fornendo informazioni preziose sulla distribuzione dei dati e sulla presenza di valori estremi.

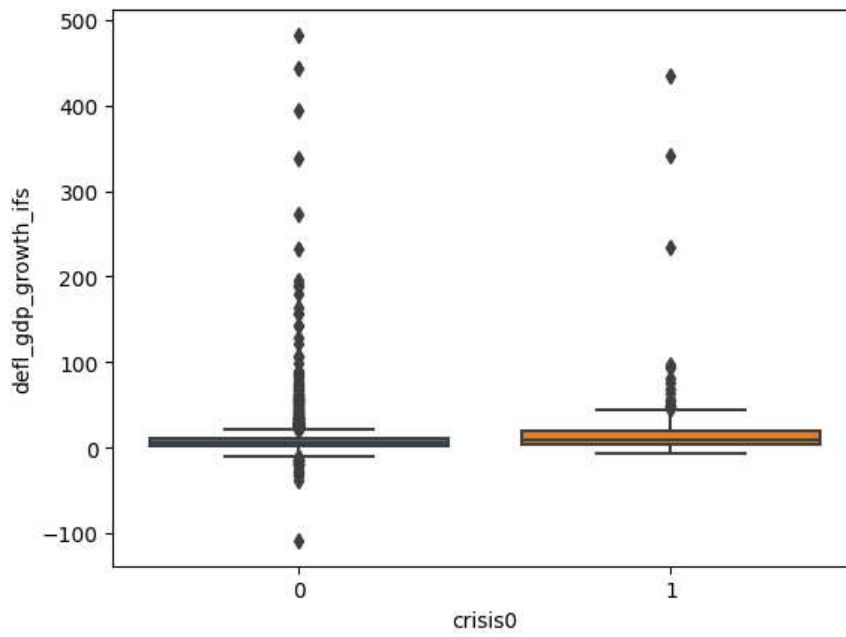


Fig 2.2 *Boxplot Deflector GDP growth (IFS) rispetto alla crisi*

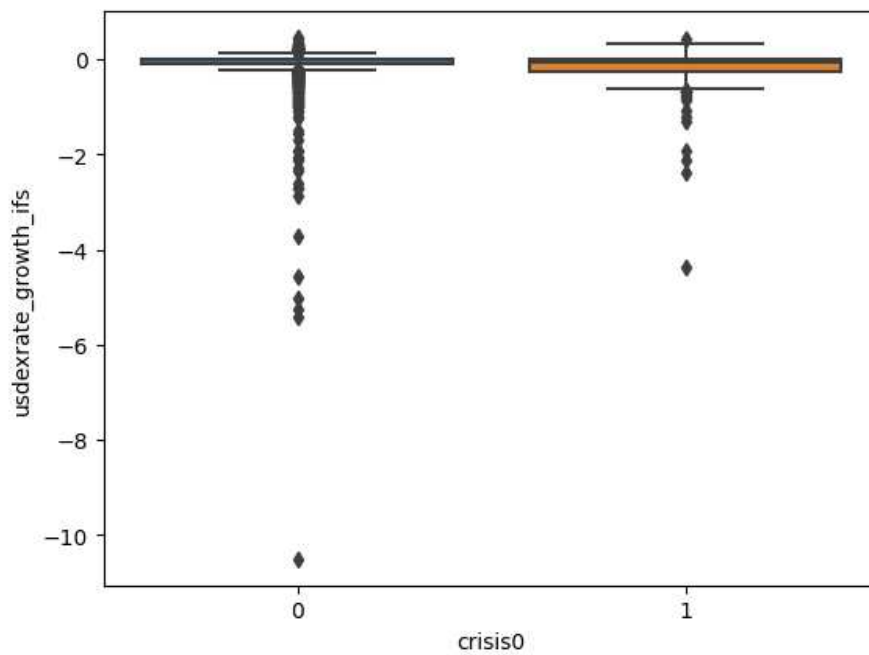


Fig 2.3 *Boxplot US Dollar Exchange Rate Growth (IFS) rispetto alla crisi*

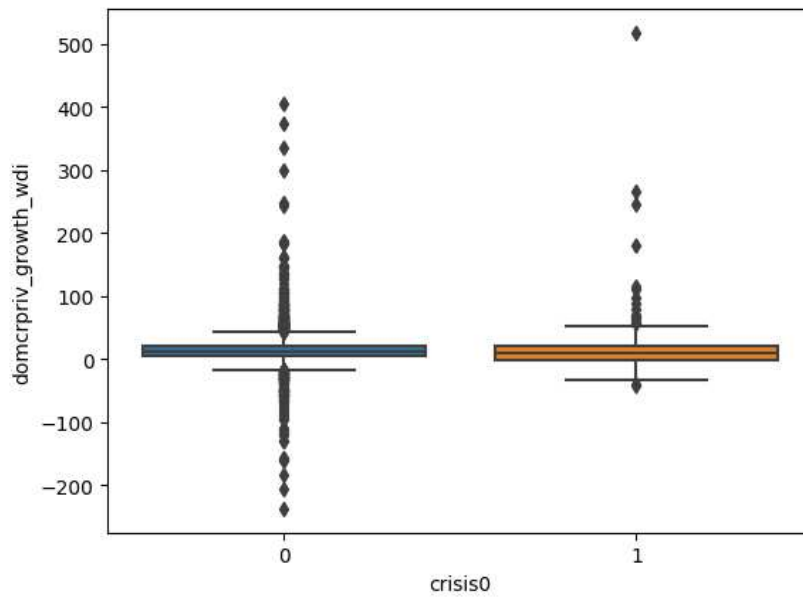


Fig 2.4 Boxplot Domestic Credit to the Private Sector Growth (WDI) rispetto alla crisi

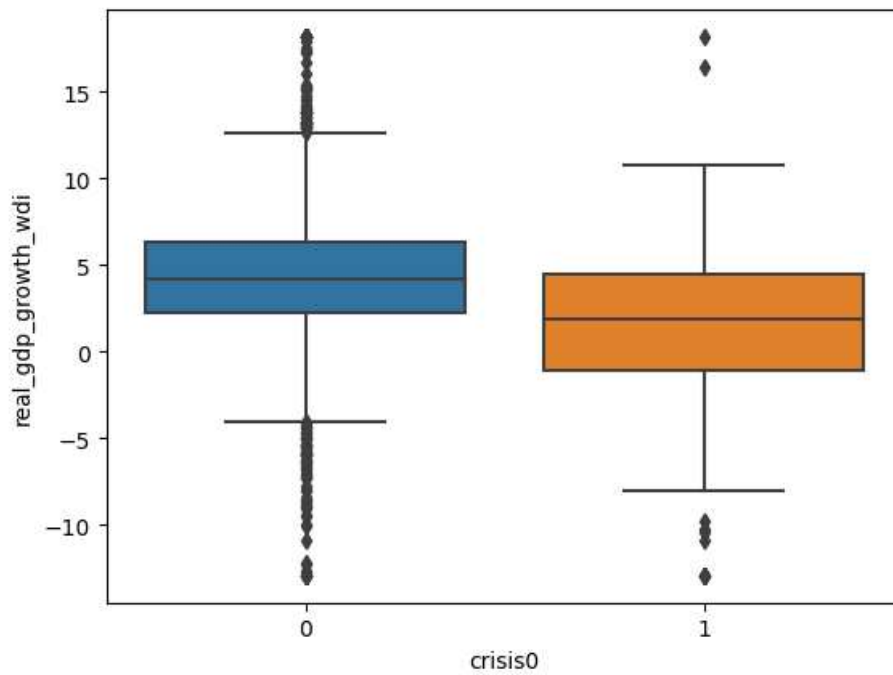


Fig 2.5 Boxplot Real GDP growth (WDI) rispetto alla crisi

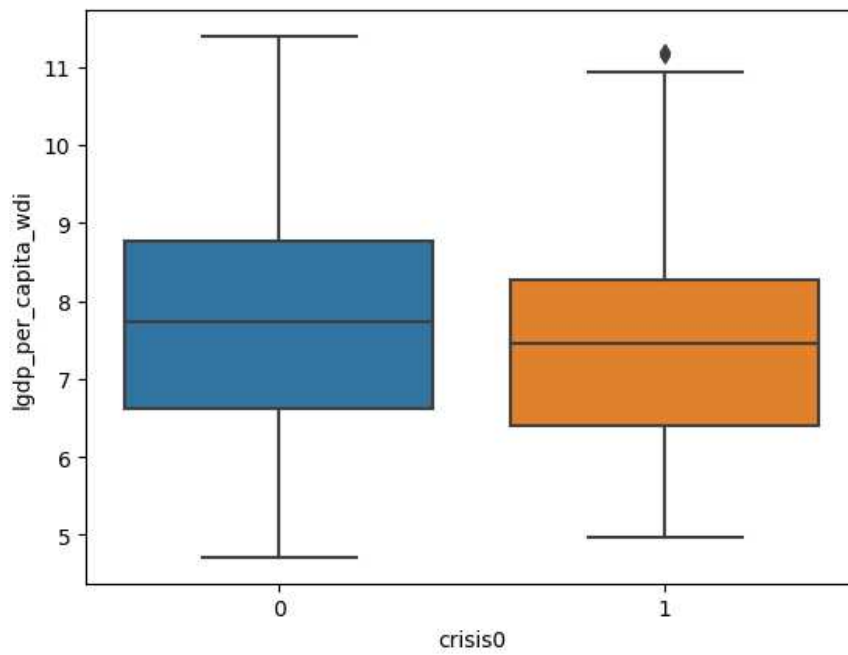


Fig 2.6 Boxplot Log GDP per capita (WDI) rispetto alla crisi

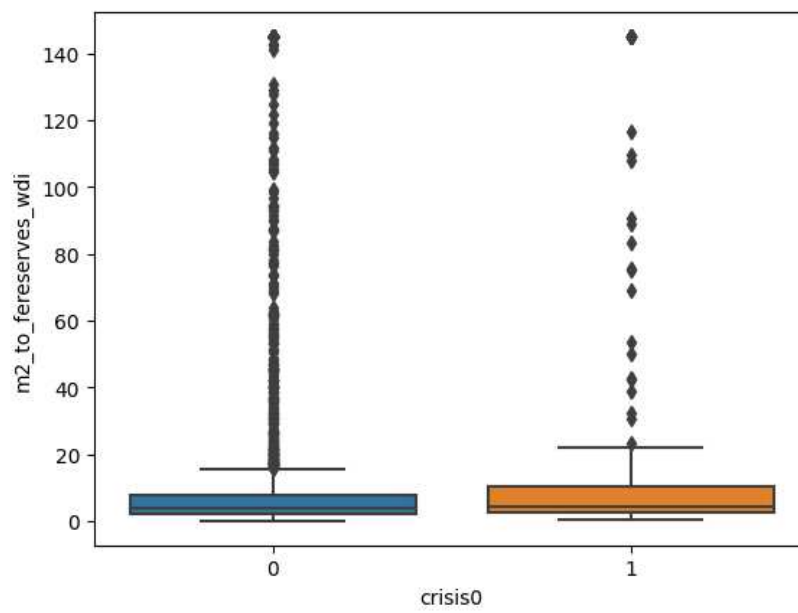


Fig 2.7 Boxplot M2 to Foreign Exchange Reserves Ratio (WDI) rispetto alla crisi

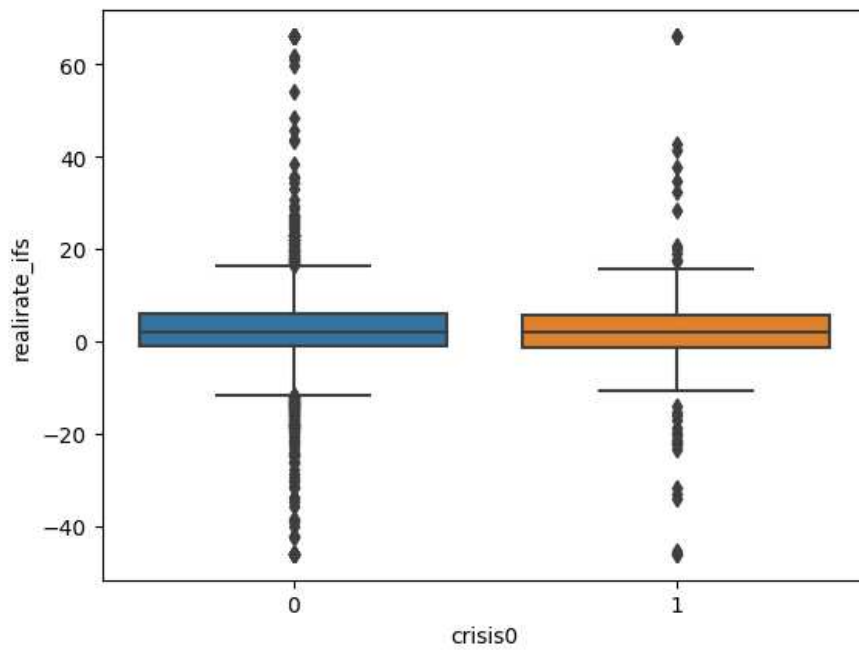


Fig 2.8 Boxplot Real interest rate (IFS) rispetto alla crisi

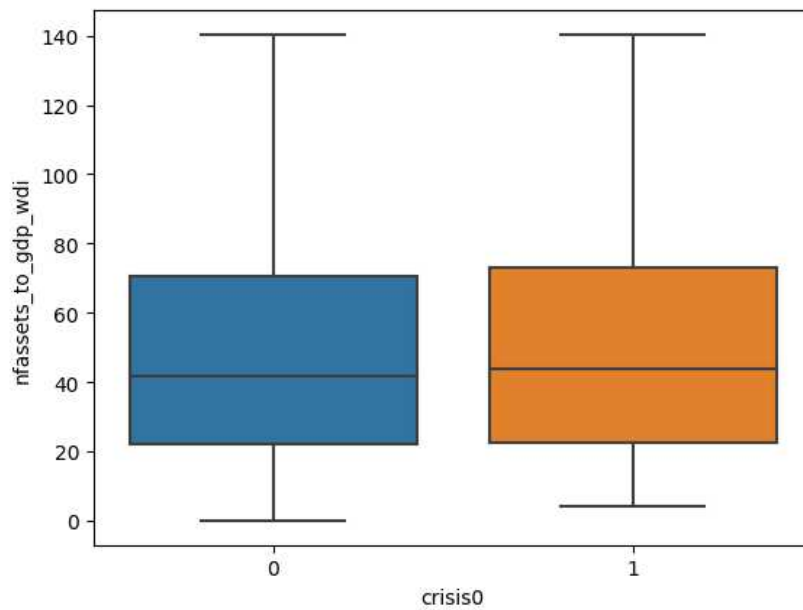


Fig 2.9 Boxplot Net Foreign Assets to GDP Ratio (WDI) rispetto alla crisi

Di seguito è riportata una tabella contenente le principali statistiche descrittive delle variabili utilizzate (media, standard deviation, valore minimo, 25° percentile, 50° percentile, 75° percentile e valore massimo).

	mean	std	min	25%	50%	75%	max
crisis0	0.068331	0.252352	0.000000	0.000000	0.000000	0.000000	1.000000
defl_gdp_growth_ifs	9.727042	24.093835	-108.564003	2.197161	5.538104	10.471513	482.339447
usdextrate_growth_ifs	-0.078981	0.356372	-10.515289	-0.090591	-0.014895	0.004896	0.462925
domcrpriv_growth_wdi	15.227779	29.735175	-237.334274	5.082256	12.327372	20.762953	518.081116
real_gdp_growth_wdi	4.062976	4.134500	-12.912101	2.029986	4.161050	6.300000	18.180000
lgdp_per_capita_wdi	7.753656	1.457921	4.712799	6.590851	7.704715	8.737250	11.389804
m2_to_fereserves_wdi	14.856432	33.259088	0.029423	2.269971	3.915581	7.637676	144.997437
realirate_ifs	2.281910	10.884504	-46.111092	-1.085380	2.042040	5.939442	66.151863
nfassets_to_gdp_wdi	50.902692	36.902385	0.004139	22.304121	41.715454	70.548950	140.290222

Tab 2.1 Descrizione delle variabili

Dai boxplot e dalle statistiche descrittive precedenti è possibile notare come le distribuzioni delle variabili non si differenziano molto nel caso della presenza o meno della crisi. Inoltre è possibile osservare come il problema che affronteremo nei seguenti paragrafi sia estremamente sbilanciato.

2.3 Preprocessing

La fase di preprocessing dei dati è cruciale per garantire l'efficacia e l'accuratezza dell'analisi successiva. Innanzitutto, è stata effettuata l'eliminazione delle variabili con valori mancanti (NA). Questa operazione

è fondamentale per mantenere l'integrità dei dati e per ridurre il rumore che potrebbe interferire con i risultati del modello. Le variabili con troppi valori mancanti possono introdurre bias e compromettere la qualità delle previsioni, mentre le variabili non pertinenti possono distorcere i risultati e aumentare inutilmente la complessità computazionale.

Dopo aver ripulito il dataset dalle variabili incomplete, il passo successivo è stato la trasformazione del dataset da una struttura cross-section a una struttura time series (Fig. 2.13). La trasformazione in serie temporale è essenziale per analizzare come le variabili cambiano nel tempo, permettendo di catturare le dinamiche temporali e le tendenze che possono influenzare l'andamento delle crisi finanziarie. Questa struttura temporale consente di utilizzare modelli che tengono conto della sequenzialità dei dati, migliorando la capacità predittiva.

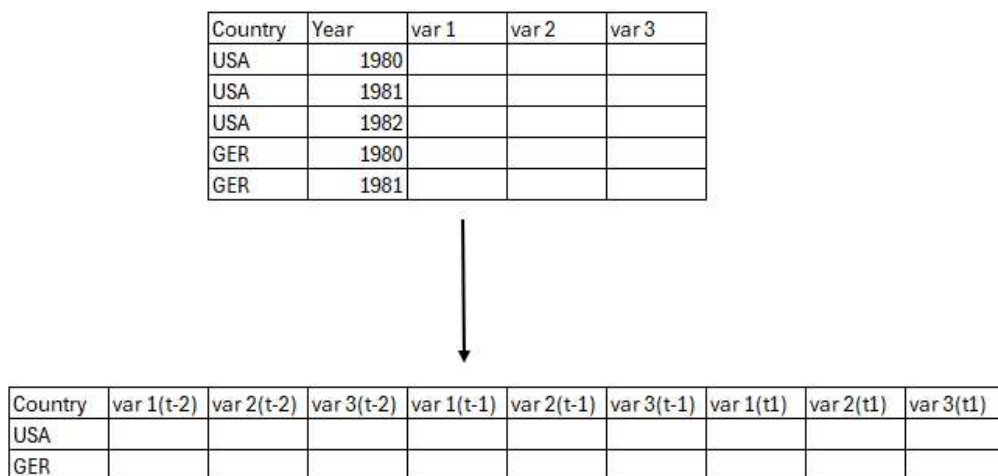


Fig 2.10 Trasformazione applicata al Dataset

Infine, tutte le variabili sono state scalate utilizzando il metodo Min-Max. Questo metodo normalizza i dati mappandoli su un intervallo tra 0 e 1, il che è particolarmente importante quando si lavora con algoritmi di deep learning. La scalatura Min-Max assicura che tutte le variabili abbiano lo stesso peso nel modello, evitando che quelle con range di valori più ampi dominino il processo di apprendimento. Inoltre, la normalizzazione facilita la convergenza dei metodi di ottimizzazione, rendendo l'addestramento del modello più stabile ed efficiente.

2.4 Reti Neurali

Le reti neurali artificiali rappresentano una delle innovazioni più rivoluzionarie nel campo dell'apprendimento automatico e dell'intelligenza artificiale. Ispirate alla struttura e al funzionamento del cervello umano, queste reti sono costituite da un insieme di nodi interconnessi, chiamati neuroni, che lavorano in sinergia per riconoscere pattern complessi nei dati. Organizzate in strati, le reti neurali possono essere utilizzate per una vasta gamma di applicazioni, tra cui la visione artificiale, il riconoscimento vocale, la traduzione automatica, la diagnosi medica e la previsione di eventi futuri come le crisi finanziarie.

Un principio teorico fondamentale che sottolinea la potenza delle reti neurali è la teoria dell'approssimazione universale. Secondo questa teoria,

una rete neurale con almeno un livello nascosto, contenente un numero sufficiente di neuroni, è in grado di approssimare qualsiasi funzione continua con un livello di precisione arbitrariamente elevato. Questo significa che, indipendentemente dalla complessità e dalla natura non lineare delle relazioni nei dati, una rete neurale adeguatamente configurata può modellarle con successo. La teoria dell'approssimazione universale conferisce alle reti neurali una versatilità eccezionale, rendendole capaci di affrontare una vasta gamma di problemi di apprendimento supervisionato e non supervisionato.

Tra le possibili reti neurali una delle più famose è rappresentata dal Multi Layer Perceptron (MLP), il cui funzionamento si basa su una struttura stratificata di neuroni, suddivisi in tre tipi principali di strati:

- **Strato di Input (input layer):** Questo è il livello iniziale che riceve i dati grezzi o pre-processati. Ogni neurone in questo strato rappresenta una caratteristica del dataset.
- **Strati Nascosti (hidden layers):** Questi strati, che possono essere uno o più, e si trovano tra l'input e l'output. Ogni neurone in uno strato nascosto riceve input dai neuroni dello strato precedente, li elabora attraverso una funzione di attivazione e trasmette il risultato ai neuroni dello strato successivo. Gli strati nascosti sono

responsabili dell'estrazione delle caratteristiche e della modellazione delle complessità nei dati. Una funzione di attivazione è una componente fondamentale nelle reti neurali artificiali che introduce non linearità nel modello, permettendo alla rete di apprendere e rappresentare relazioni complesse nei dati. Dopo che un neurone ha ricevuto i suoi input ponderati e ha calcolato la loro somma, la funzione di attivazione viene applicata a questo valore aggregato per determinare l'output del neurone. Le funzioni di attivazione più usate in questi strati sono

- ReLU (Rectified Linear Unit):

$$f(x) = \max(0, x)$$

- Sigmoid:

$$\text{sigma}(x) = \frac{1}{1 + e^{-x}}$$

- Tanh (Tangente Iperbolica):

$$\text{tanh}(x) = \frac{e^x + e^{-x}}{e^x - e^{-x}}$$

- **Strato di Output (output layer):** Questo è il livello finale che produce la previsione o la classificazione. Il numero di neuroni in questo strato dipende dal tipo di problema; ad esempio, per una

classificazione binaria ci sarà un solo neurone con una funzione di attivazione sigmoid, mentre per una classificazione multiclasse ci sarà un neurone per ciascuna classe con una funzione di attivazione softmax.

Il funzionamento di un singolo neurone è descritto come segue:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

dove y è l'output del neurone, f è la funzione di attivazione, w_i sono i pesi, x_i gli input e b il bias.

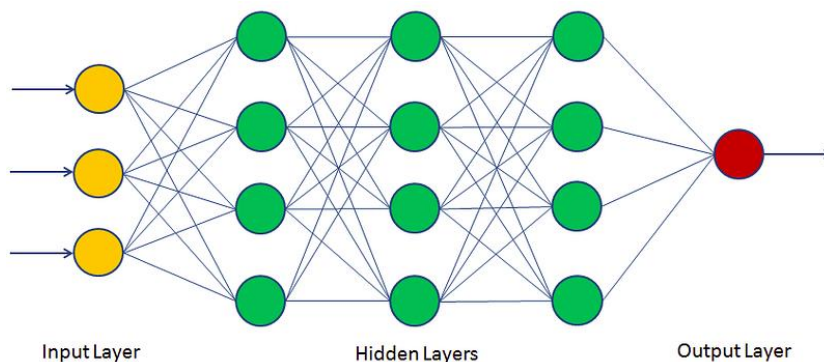


Fig 2.11 Rappresentazione di una rete neurale MLP

Il processo di addestramento delle reti neurali può seguire diversi metodi, uno dei più utilizzati è la backpropagation, ovvero un metodo efficiente per calcolare i gradienti necessari per aggiornare i pesi dei neuroni. Nel caso del MLP, l'algoritmo di backpropagation si svolge in due fasi principali:

- **Propagazione in Avanti (Forward Pass):** Durante questa fase, gli input vengono trasmessi attraverso la rete strato per strato fino a raggiungere l'output finale. Ogni neurone calcola un output utilizzando una combinazione lineare dei suoi input ponderati e una funzione di attivazione. La formula per il forward pass di un singolo neurone è:

$$a_i^{(l)} = f \left(\sum_{j=1}^{n^{(l-1)}} w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)} \right)$$

Dove a è l'attivazione del neurone i nel layer l , $w_{ij}^{(l)}$ è il peso che collega il neurone j del layer $l-1$ al neurone i del layer l e $b_i^{(l)}$ è il bias del neurone i nel layer l .

- **Propagazione all'Indietro (Backward Pass):** Nella fase successiva, l'errore calcolato viene propagato all'indietro attraverso la rete. Questo processo inizia dal livello di output e procede verso il livello di input. L'obiettivo è calcolare i gradienti dei pesi rispetto all'errore utilizzando la regola della catena del calcolo differenziale (chain rule of differentiation). Il primo passo è calcolare l'errore all'output layer:

$$\delta^{(L)} = \nabla_a \mathcal{L} \odot f'(z^{(L)})$$

dove $\delta^{(l)}$ è l'errore del layer di output, $\nabla_a \mathcal{L}$ è il gradiente della funzione di perdita rispetto all'attivazione del layer di output, f' è la derivata della funzione di attivazione e $z^{(l)}$ è la somma ponderata degli input al layer di output.

Successivamente, l'errore viene propagato all'indietro attraverso i layer nascosti:

$$\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)}) \odot f'(z^{(l)})$$

dove $\delta^{(l)}$ è l'errore al layer l , $W^{(l+1)}$ sono i pesi del layer $l+1$, e $f'(z^{(l)})$ è la derivata della funzione di attivazione rispetto alla somma ponderata degli input al layer l .

Infine, i gradienti dei pesi e dei bias vengono calcolati come segue:

$$\frac{\partial \mathcal{L}}{\partial W_{ij}^{(l)}} = a_j^{(l-1)} \delta_i^{(l)} \quad \frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = \delta_i^{(l)}$$

Questi gradienti sono poi utilizzati per aggiornare i pesi e i bias della rete tramite un algoritmo di ottimizzazione come la discesa del gradiente:

$$W_{ij}^{(l)} := W_{ij}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial W_{ij}^{(l)}} \quad b_i^{(l)} := b_i^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial b_i^{(l)}}$$

dove η è il tasso di apprendimento.

Di seguito lo pseudo codice che riassume l'apprendimento di una rete neurale MLP:

```
1: Let  $D.train = \{(\mathbf{x}_k, y_k) \mid k = 1, 2, \dots, n\}$  be the set of training instances.
2: Set counter  $c \leftarrow 0$ .
3: Initialize the weight and bias terms  $(\mathbf{w}^{(0)}, \mathbf{b}^{(0)})$  with random values.
4: repeat
5:   for each training instance  $(\mathbf{x}_k, y_k) \in D.train$  do
6:     Compute the set of activations  $(a_i^l)_k$  by making a forward pass using  $\mathbf{x}_k$ .
7:     Compute the set  $(\delta_i^l)_k$  by backpropagation using Equations 6.58 and 6.59.
8:     Compute  $(\partial \text{Loss} / \partial w_{ij}^l, \partial \text{Loss} / \partial b_i^l)_k$  using Equations 6.56 and 6.57.
9:   end for
10:  Compute  $\partial E / \partial w_{ij}^l \leftarrow \sum_{k=1}^n (\partial \text{Loss} / \partial w_{ij}^l)_k$ .
11:  Compute  $\partial E / \partial b_i^l \leftarrow \sum_{k=1}^n (\partial \text{Loss} / \partial b_i^l)_k$ .
12:  Update  $(\mathbf{w}^{(c+1)}, \mathbf{b}^{(c+1)})$  by gradient descent using Equations 6.53 and 6.54.
13:  Update  $c \leftarrow c + 1$ .
14: until  $(\mathbf{w}^{(c+1)}, \mathbf{b}^{(c+1)})$  and  $(\mathbf{w}^{(c)}, \mathbf{b}^{(c)})$  converge to the same value
```

Fig 2.12 Pseudocodice apprendimento rete neurale MLP

2.5 Reti Neurali Ricorrenti

Le reti neurali ricorrenti rappresentano una classe di reti neurali artificiali progettate per lavorare con dati sequenziali, come il linguaggio naturale o serie temporali. A differenza delle reti neurali tradizionali, che trattano ciascun input come un'entità indipendente, le RNN sono in grado di considerare la struttura sequenziale dei dati, mantenendo una memoria interna che consente loro di catturare le dipendenze a lungo termine nei dati.

Il processo di addestramento delle reti neurali ricorrenti coinvolge un'estensione dell'algoritmo di backpropagation chiamata "backpropagation

through time" (BPTT). A differenza delle reti neurali feedforward, in cui l'errore è propagato solo in avanti attraverso la rete, nelle RNN l'errore viene propagato sia in avanti che all'indietro attraverso il tempo, considerando ogni istante di tempo come uno strato separato della rete.

Durante la fase di forward pass, l'input corrente e lo stato precedente vengono utilizzati per calcolare l'output corrente e lo stato successivo. Durante la fase di backward pass, l'errore viene propagato all'indietro attraverso la rete, partendo dall'ultimo istante di tempo fino al primo, e i pesi vengono aggiornati utilizzando la regola della catena del calcolo differenziale.

La principale sfida nell'addestramento delle reti neurali ricorrenti riguarda il problema del "vanishing gradient" o "exploding gradient", in cui i gradienti diventano estremamente piccoli o estremamente grandi durante la fase di backpropagation attraverso il tempo. Questo può portare a problemi di apprendimento instabili o lenti. Per mitigare questo problema, sono state proposte diverse varianti di RNN, come le LSTM (Long Short-Term Memory) e le GRU (Gated Recurrent Unit), che utilizzano meccanismi di gating per controllare il flusso di informazioni attraverso la rete e mitigare il problema del vanishing gradient.

2.6 Reti Neurali LSTM (Long-Short Term Memory)

Le Long Short-Term Memory (LSTM) sono una variante avanzata delle reti neurali ricorrenti (RNN) progettate per superare le limitazioni di queste, in particolare il problema del "vanishing gradient" che impedisce l'apprendimento efficace di dipendenze a lungo termine nei dati sequenziali. Le LSTM sono state introdotte da Hochreiter e Schmidhuber nel 1997 e sono state sviluppate per consentire alle reti neurali di ricordare informazioni per periodi di tempo prolungati, mantenendo al contempo la capacità di dimenticare informazioni non rilevanti. Questa capacità rende le LSTM estremamente utili in una vasta gamma di applicazioni, come il riconoscimento vocale, la traduzione automatica, l'analisi delle serie temporali e molte altre.

Le LSTM sono composte da una serie di celle, ciascuna delle quali contiene quattro componenti principali: la cella di memoria, le porte di input, output e forget (dimenticanza). Queste porte regolano il flusso di informazioni all'interno della cella e tra le celle successive, consentendo alla rete di gestire dipendenze a lungo termine e breve termine:

- **Cella di Memoria (Cell State):** La cella di memoria è il cuore della LSTM, progettata per conservare le informazioni lungo la sequenza. Questo permette alla cella di mantenere informazioni intatte per lunghi periodi. La cella di memoria viene aggiornata continuamente

attraverso la combinazione di nuove informazioni in ingresso e quelle esistenti, regolate dalle porte di input e forget.

- **Porta di Input (Input Gate):** La porta di input decide quali nuove informazioni devono essere aggiunte alla cella di memoria. Utilizza una funzione sigmoide che limita i valori tra 0 e 1, determinando l'importanza delle informazioni in ingresso. La funzione sigmoide è definita come:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

dove i_t è la porta di input, σ è la funzione sigmoide, W_i sono i pesi associati, h_{t-1} è lo stato nascosto precedente, x_t è l'input attuale e b_i è il bias.

- **Porta di Output (Output Gate):** La porta di output controlla quali informazioni dalla cella di memoria devono essere utilizzate per calcolare l'output della LSTM. Anch'essa viene calcolata utilizzando una funzione sigmoide:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

- **Porta di Dimenticanza (Forget Gate):** La porta di dimenticanza decide quali informazioni devono essere scartate dalla cella di memoria. Questa funzione è cruciale per evitare che informazioni

irrilevanti persistano nella memoria. La porta di dimenticanza è definita come:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- **Aggiornamento della Cella di Memoria:** La cella di memoria viene aggiornata combinando le informazioni esistenti e le nuove informazioni selezionate dalla porta di input e filtrate dalla porta di dimenticanza:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Questo aggiornamento permette di mantenere le informazioni rilevanti e scartare quelle irrilevanti, bilanciando l'importanza delle vecchie e nuove informazioni.

- **Calcolo dell'Output:** L'output della LSTM viene calcolato combinando la cella di memoria aggiornata e la porta di output:

$$h_t = o_t \cdot \tanh(C_t)$$

In questo modo, l'output h_t è una versione filtrata della cella di memoria, che incorpora sia le informazioni attuali sia quelle passate.

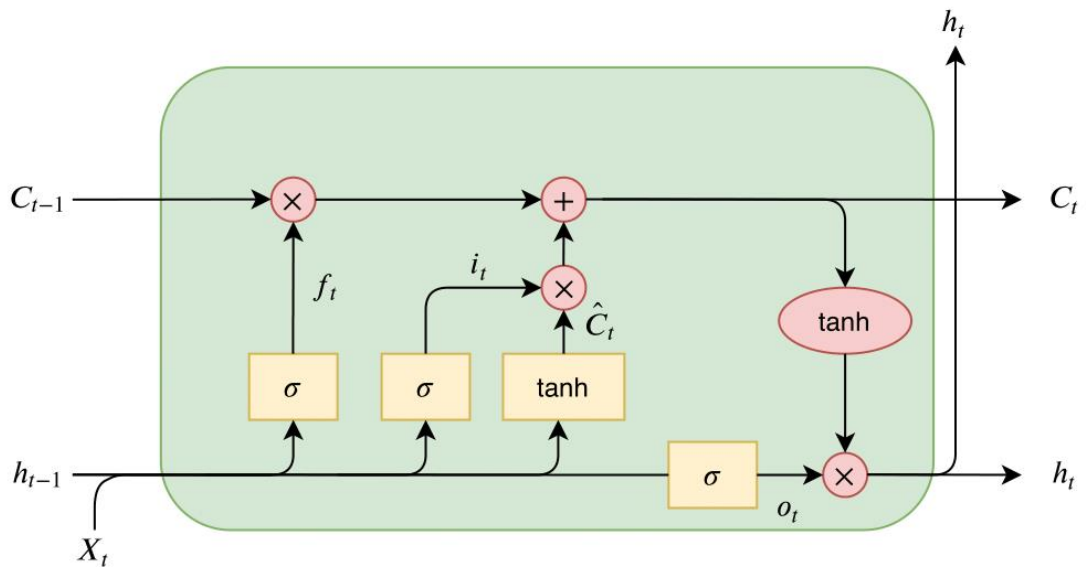


Fig 2.13 Rappresentazione LSTM

Il processo di backpropagation in una LSTM, noto come "backpropagation through time" (BPTT), è una versione estesa della backpropagation standard, adattata per le reti neurali ricorrenti. Il BPTT è più complesso nel contesto delle LSTM a causa delle molteplici porte e della cella di memoria. Durante il forward pass, vengono calcolati gli stati delle porte e della cella di memoria. Durante il backward pass, l'errore viene propagato all'indietro attraverso il tempo, aggiornando i pesi delle porte di input, output e forget, oltre ai pesi associati alla cella di memoria.

- **Forward Pass:** Durante il forward pass, ogni input sequenziale viene processato attraverso le porte della LSTM. Le porte di input, output e forget regolano il flusso di informazioni in ogni passo temporale.

- **Backward Pass:** Nel backward pass, l'errore viene propagato indietro attraverso la sequenza temporale. I gradienti dell'errore rispetto ai pesi della rete vengono calcolati e aggiornati. La presenza delle porte di input, output e forget nelle LSTM aiuta a mantenere i gradienti stabili, superando il problema del vanishing gradient.
- **Aggiornamento dei Pesi:** I pesi delle varie porte vengono aggiornati utilizzando un algoritmo di ottimizzazione, come l'algoritmo di discesa del gradiente, che minimizza l'errore totale della rete. Questo aggiornamento tiene conto delle informazioni provenienti da tutti i passi temporali della sequenza.

2.7 K-means

L'algoritmo K-means è una tecnica di clustering non supervisionato ampiamente utilizzata per suddividere un dataset in k gruppi distinti, dove k rappresenta il numero di cluster predefiniti. Questo metodo è particolarmente apprezzato per la sua semplicità ed efficienza computazionale, rendendolo adatto per una vasta gamma di applicazioni, dalla segmentazione del mercato all'analisi delle immagini. Il funzionamento dell'algoritmo può essere descritto attraverso i seguenti passaggi:

1. **Inizializzazione:** Vengono selezionati casualmente k punti del dataset come centroidi iniziali. Questo passaggio può influenzare significativamente i risultati finali, motivo per cui spesso si utilizzano tecniche come K-means++ per migliorare la scelta iniziale dei centroidi.
2. **Assegnazione:** Ogni punto dati viene assegnato al centroide più vicino in base ad una metrica di distanza. Questo crea k -cluster provvisori.
3. **Calcolo dei nuovi centroidi:** Per ciascun cluster, viene calcolato un nuovo centroide come la media dei punti dati appartenenti a quel cluster. Questo passaggio sposta i centroidi verso il centro dei dati assegnati.
4. **Iterazione:** I passaggi di assegnazione e calcolo dei centroidi vengono ripetuti fino a quando i centroidi non cambiano più in modo significativo tra un'iterazione e l'altra. Questo processo di iterazione continua fino al raggiungimento della convergenza, che è determinata da un criterio di arresto predefinito, come un numero massimo di iterazioni o una soglia di tolleranza per il movimento dei centroidi.

L'obiettivo finale dell'algoritmo K-means è minimizzare la somma delle distanze tra i punti dati e i rispettivi centroidi, ottenendo cluster compatti e ben separati. Nonostante la sua efficacia, K-means presenta alcune limitazioni. È sensibile ai valori iniziali dei centroidi e può convergere a soluzioni locali, piuttosto che a quella globale ottimale. Per mitigare questo problema, l'algoritmo viene spesso eseguito più volte con diverse inizializzazioni, selezionando infine la soluzione con il miglior valore della funzione obiettivo. Inoltre, K-means assume che i cluster siano di forma sferica e di dimensioni simili, il che può non essere sempre appropriato per tutti i tipi di dati.

2.8 Problema dei dati sbilanciati

Il problema dei dati sbilanciati è una sfida comune negli EWS, particolarmente rilevante quando si lavora con dataset in cui le classi non sono rappresentate in modo equo. Nel contesto del nostro dataset, questa sproporzione si manifesta con una forte discrepanza tra il numero di record di crisi (219, pari al 7%) e quelli di non crisi (2986, pari al 93%). Questa situazione può influenzare negativamente la performance dei modelli predittivi, in quanto i modelli tendono ad essere più accurati nel predire la classe maggioritaria, ignorando o sottostimando la classe minoritaria. Le implicazioni dei dati sbilanciati possono essere:

- I modelli di machine learning possono sviluppare un bias verso la classe maggioritaria, portando a una predizione sproporzionata della classe di non crisi. Questo significa che, nonostante un'alta accuratezza complessiva, il modello può fallire nel rilevare eventi di crisi, che sono di fondamentale importanza per l'analisi.
- Le metriche tradizionali come l'accuratezza non sono adeguate per valutare modelli su dataset sbilanciati, poiché un modello che predice sempre la classe maggioritaria può ancora apparire performante. Metriche alternative come la precisione, il recall, il punteggio F1 e l'Area sotto la Curva ROC diventano cruciali per una valutazione più equilibrata delle performance.

Nei seguenti paragrafi verranno analizzate delle tecniche applicate per ridurre questi problemi.

2.9 Oversampling

L'oversampling è una strategia comune per affrontare il problema dei dati sbilanciati, che prevede l'aumento della rappresentazione della classe minoritaria nel dataset. Questo può essere fatto replicando gli esempi esistenti della classe minoritaria o, in modo più sofisticato, generando nuovi esempi sintetici. Una delle tecniche più efficaci e ampiamente utilizzate per l'oversampling è SMOTE (Synthetic Minority Over-sampling

Technique). Questa tecnica è stata sviluppata per migliorare la qualità dell'oversampling generando esempi sintetici della classe minoritaria invece di semplicemente duplicare gli esempi esistenti. Questo metodo aiuta a ridurre il problema dell'overfitting, che può verificarsi quando gli stessi esempi vengono replicati più volte. Il suo funzionamento:

1. **Selezione degli Esempi:** Per ogni esempio della classe minoritaria nel dataset, SMOTE seleziona uno o più dei suoi k-nearest neighbors (vicini più prossimi). Il numero di vicini k è un parametro che può essere regolato in base alle necessità del modello.
2. **Generazione degli Esempi Sintetici:** Per ogni coppia formata dall'esempio originale e uno dei suoi vicini più prossimi, SMOTE genera un nuovo esempio sintetico. Questo avviene creando una combinazione lineare tra l'esempio originale e il vicino scelto. La formula per generare un esempio sintetico è la seguente:

$$\text{Esempio Sintetico} = x_i + \lambda \cdot (x_i - x_{nn})$$

dove x_i è l'esempio originale, x_m è il vicino più prossimo, e λ è un numero casuale tra 0 e 1. Questa interpolazione produce un nuovo punto che si trova sulla linea che congiunge l'esempio originale e il suo vicino

3. Ripetizione del Processo: Questo processo viene ripetuto per ogni esempio della classe minoritaria, fino a quando la classe minoritaria raggiunge una dimensione desiderata o un rapporto predefinito rispetto alla classe maggioritaria.

2.10 Undersampling

L'undersampling è una strategia per affrontare il problema dei dati sbilanciati riducendo la rappresentazione della classe maggioritaria nel dataset. Una tecnica comune di undersampling è il metodo Cluster Centroids, che seleziona un sottoinsieme rappresentativo degli esempi della classe maggioritaria.

Il metodo Cluster Centroids è una tecnica di undersampling che riduce la dimensione della classe maggioritaria identificando i suoi centroidi e rimuovendo gli esempi che si trovano in prossimità di questi centroidi. Questo approccio aiuta a mantenere la struttura e le caratteristiche salienti della classe maggioritaria, riducendone la dimensione senza eliminare informazioni importanti. Il suo funzionamento:

- 1. Clusterizzazione:** Innanzitutto, il dataset viene clusterizzato utilizzando un algoritmo di clustering, tipicamente K-Means, per identificare i centroidi dei cluster della classe maggioritaria.

- 2. Selezione dei Centroidi:** Dopo aver individuato i centroidi, vengono selezionati come rappresentanti della classe maggioritaria. Questi centroidi sono punti che rappresentano in modo significativo la distribuzione e le caratteristiche della classe maggioritaria.
- 3. Rimozione degli Esempi Vicini ai Centroidi:** Gli esempi della classe maggioritaria che si trovano in prossimità dei centroidi selezionati vengono rimossi dal dataset. La distanza tra gli esempi e i centroidi può essere calcolata utilizzando diverse metriche, come la distanza euclidea o la distanza di Mahalanobis.
- 4. Ridimensionamento del Dataset:** Dopo la rimozione degli esempi vicini ai centroidi, il dataset viene ridimensionato in modo che la classe maggioritaria raggiunga una dimensione desiderata o un rapporto predefinito rispetto alla classe minoritaria.

2.11 Metacost

Il meta-algoritmo MetaCost è progettato per affrontare efficacemente il problema dei dataset sbilanciati nell'apprendimento automatico supervisionato. La sua caratteristica distintiva è quella di personalizzare i costi associati agli errori di classificazione, consentendo al modello di concentrarsi maggiormente sulle classi meno rappresentate e di migliorare

l'efficacia nella predizione di eventi rari, come le crisi finanziarie. Fasi del MetaCost:

- 1. Allenamento di Diversi Modelli:** Nella prima fase, vengono allenati diversi modelli di classificazione utilizzando il metodo del bagging, che consiste nel creare m campioni di training ottenuti dal dataset originale tramite campionamento con sostituzione. Questo consente di generare una varietà di modelli che verranno successivamente combinati.
- 2. Rietichettamento del Dataset:** Nella fase successiva, ogni record del dataset viene rietichettato. Per ogni record, vengono individuati i modelli che non lo hanno utilizzato nel training set e vengono utilizzati per effettuare una previsione. Le previsioni di questi modelli vengono poi combinate mediante media e moltiplicate per una matrice di costo personalizzata. Il record viene quindi etichettato con la classe associata al costo inferiore.
- 3. Allenamento di un Nuovo Modello:** Nella fase finale, viene allenato un nuovo modello utilizzando il dataset rietichettato. Questo nuovo modello tiene conto dei costi differenziati associati agli errori di classificazione e mira a minimizzare la funzione di costo complessiva.

Qui di seguito lo pseudocodice associato a questo algoritmo:

Table 1: The MetaCost algorithm.

Inputs:
 S is the training set,
 L is a classification learning algorithm,
 C is a cost matrix,
 m is the number of resamples to generate,
 n is the number of examples in each resample,
 p is *True* iff L produces class probabilities,
 q is *True* iff all resamples are to be used for each example.

Procedure MetaCost (S, L, C, m, n, p, q)

For $i = 1$ to m
 Let S_i be a resample of S with n examples.
 Let $M_i =$ Model produced by applying L to S_i .

For each example x in S
 For each class j
 Let $P(j|x) = \frac{1}{\sum_i 1} \sum_i P(j|x, M_i)$
 Where
 If p then $P(j|x, M_i)$ is produced by M_i
 Else $P(j|x, M_i) = 1$ for the class predicted by M_i for x , and 0 for all others.
 If q then i ranges over all M_i
 Else i ranges over all M_i such that $x \notin S_i$.
 Let x 's class = $\operatorname{argmin}_i \sum_j P(j|x)C(i, j)$.

Let $M =$ Model produced by applying L to S .

Return M .

Fig 2.14 Pseudocodice Metacost

Capitolo 3

Esperimenti e Risultati

3.1 Grid Search

Per individuare il modello ottimale è stato implementato un approccio di Grid Search. Questo metodo, essenziale per l'ottimizzazione dei modelli di machine learning, permette di esplorare sistematicamente una vasta gamma di iperparametri per identificare la configurazione più performante. L'idea alla base della Grid Search è di eseguire un'esplorazione esaustiva di tutte le possibili combinazioni di parametri predeterminati, al fine di determinare quella che offre le migliori prestazioni secondo metriche specifiche.

I parametri critici che influenzano significativamente la performance del modello e che sono stati considerati nella ricerca includono il numero di lag input, il numero di epoche, il numero di neuroni, il batch size, il learning rate, la percentuale di oversampling, la percentuale di undersampling e la

matrice di costo del MetaCost. Per ogni modello l'ottimizzatore è stato 'Adam' mentre i valori esaminati per ciascun parametro erano i seguenti:

- Numero di lag input: [1, 2, 3, 4, 5]. Il numero di lag input si riferisce al numero di ritardi temporali considerati per l'input del modello, un elemento cruciale per la previsione basata su serie temporali.
- Numero di epoche: [50, 100, 150, 200, 250, 500, 1000]. Il numero di epoche rappresenta il numero di volte in cui l'intero set di dati o batch di addestramento viene passato attraverso la rete, influenzando il grado di apprendimento del modello.
- Numero di neuroni: [5, 10, 15, 20, 25, 30]. Il numero di neuroni per strato è un parametro fondamentale che determina la capacità della rete di apprendere rappresentazioni complesse dai dati.
- Batch size: [32, 63, 128]. Il batch size è il numero di campioni elaborati prima di aggiornare i pesi del modello, influenzando l'efficienza e la stabilità del processo di addestramento.
- Learning rate: [0.001, 0.005, 0.0001, 0.0005]. Il learning rate è la dimensione del passo con cui il modello aggiorna i pesi durante l'addestramento, un parametro critico per la convergenza del modello.

- Percentuale di oversampling: [0.00, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 1.00]. L'oversampling è una tecnica utilizzata per bilanciare i dati quando ci sono classi minoritarie, aumentando il numero di esempi di queste classi.
- Percentuale di undersampling: [0.00, 0.15, 0.20, 0.25]. L'undersampling, al contrario, riduce il numero di esempi delle classi maggioritarie per bilanciare il dataset.
- Metacost: [[0.0, 4.1], [1.0, 0.0]], [[0, 10], [1, 0]], [[0, 2], [1, 0]], [[0, 3], [1, 0]], [[0, 4], [1, 0]], [[0, 5], [1, 0]], [[0, 6], [1, 0]], [[0, 7], [1, 0]], [[0, 8], [1, 0]], [[0, 9], [1, 0]], [[0.0, 0.00041425], [0.00684932, 0.0]], [[0.0, 0.00684932], [0.00041425, 0.0]], [[0.0, 3.5], [1.0, 0.0]], [[0.0, 4.2], [1.0, 0.0]], [[0.0, 4.3], [1.0, 0.0]], [[0.0, 4.4], [1.0, 0.0]], [[0.0, 4.5], [1.0, 0.0]], [[0.0, 4.6], [1.0, 0.0]], [[0.0, 4.7], [1.0, 0.0]], [[0.0, 4.8], [1.0, 0.0]], [[0.0, 4.9], [1.0, 0.0]]. Il metacost viene utilizzato per gestire le situazioni di squilibrio tra le classi mediante l'assegnazione di costi differenti a errori di classificazione diversi, consentendo al modello di penalizzare maggiormente gli errori più critici.

La Grid Search è stata condotta in maniera iterativa, partendo da una ricerca ampia e generale delle combinazioni di parametri per poi restringere

gradualmente il campo verso le combinazioni più promettenti. Nella fase iniziale, l'obiettivo era di coprire un ampio spettro di configurazioni per assicurarsi che nessuna potenziale combinazione fosse trascurata. Questa fase ha implicato l'esecuzione di numerosi esperimenti su un insieme di dati di validazione, valutando le performance del modello con ciascuna combinazione di parametri.

Durante questa prima fase, sono stati raccolti dati dettagliati sulle performance del modello, incluse metriche come accuratezza, precisione, recall e F1-score. Queste metriche hanno permesso di identificare quali combinazioni di parametri portavano a modelli che erano non solo accurati, ma anche equilibrati nelle loro capacità di rilevare sia falsi positivi che falsi negativi.

Una volta identificate le combinazioni più promettenti, la Grid Search è stata affinata restringendo la griglia attorno a queste combinazioni. Questo processo ha permesso di eseguire un'ottimizzazione più fine, valutando parametri con incrementi più piccoli e combinazioni più specifiche. Questo ha garantito che il modello finale non solo fosse accurato, ma anche robusto e ben generalizzabile per l'applicazione pratica nei sistemi di allerta precoce per le crisi finanziarie.

3.2 Valutazione dei modelli

Ogni combinazione della Grid Search è stata valutata utilizzando una tecnica di cross validation X-fold con $X = 10$. Questo metodo divide il dataset in 10 sottogruppi (folds), utilizzando a turno ciascun sottogruppo come set di test e gli altri 9 come set di addestramento, garantendo una valutazione robusta e affidabile delle performance del modello. Le combinazioni di iperparametri sono state valutate tramite diverse metriche di performance: Accuracy, F1, Recall 0, Recall 1, Precision 0, Precision 1 e AUC (Area Under the Curve).

Per comprendere il calcolo di queste metriche, è essenziale partire dalla matrice di confusione. La matrice di confusione è una tabella che permette di visualizzare le performance di un algoritmo di classificazione, presentando il numero di veri positivi (TP), falsi positivi (FP), veri negativi (TN) e falsi negativi (FN) (Fig. 3.1).

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negatives (TN)	False Positives (FP) Type I error
	Positive +	False Negatives (FN) Type II error	True Positives (TP)

Fig 3.1 Matrice di Confusione

Le metriche:

- **Accuracy:** misura la proporzione di previsioni corrette (sia vere positive che vere negative) sul totale delle previsioni. È calcolata come

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

- **F1 Score:** media armonica tra precisione e recall. Fornisce un bilanciamento tra i due, particolarmente utile quando si ha un disequilibrio tra le classi. Si calcola come

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Recall 0:** anche noto come specificity o TNR (True Negative Rate) per la classe negativa, misura la capacità del modello di identificare correttamente i veri negativi ed è calcolato come

$$R_0 = \frac{TN}{TN + FP}$$

- **Recall 1:** anche noto come sensitivity o TPR (True Positive Rate) per la classe positiva, misura la capacità del modello di identificare correttamente i veri positivi ed è calcolato come

$$R_1 = \frac{TP}{TP + FN}$$

- **Precision 0:** misura la proporzione di veri negativi tra tutti i negativi predetti, calcolata come

$$P_0 = \frac{TN}{TN + FN}$$

- **Precision 1:** misura la proporzione di veri positivi tra tutti i positivi predetti, calcolata come

$$P_1 = \frac{TP}{TP + FP}$$

- **AUC (Area Under the Curve)** rappresenta l'area sotto la curva ROC (Receiver Operating Characteristic), che traccia il rapporto tra TPR e FPR (False Positive Rate) a vari livelli di soglia (Fig. 3.2). L'AUC fornisce una misura aggregata delle performance del modello su tutte le possibili soglie di classificazione, dove un valore di 1 indica un modello perfetto e un valore di 0.5 indica una performance casuale.

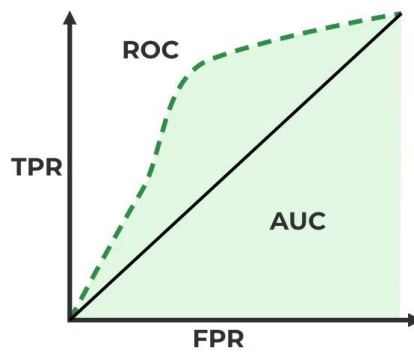


Fig 3.2 Grafico ROC-AUC

3.3 Risultati

Dopo l'esecuzione della Grid Search, sono stati identificati i migliori 10 modelli in base all'F1 score:

Input (anni)	Layers	N_neuroni	Epoche	Batch Size	Learning Rate	Over	Under	Matrice di Costo	MetaCost	Accuracy	F1	AUC	Recall 0	Recall 1	Precision 0	Precision 1
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.3][1. 0.]]		0,970	0,731	0,846	0,986	0,706	0,982	0,758
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.2][1. 0.]]		0,970	0,729	0,846	0,986	0,706	0,982	0,753
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.7][1. 0.]]		0,970	0,727	0,846	0,986	0,706	0,982	0,750
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.4][1. 0.]]		0,970	0,726	0,846	0,986	0,706	0,982	0,749
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.5][1. 0.]]		0,970	0,726	0,846	0,986	0,706	0,982	0,749
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4.9][1. 0.]]		0,969	0,722	0,842	0,986	0,699	0,982	0,747
5	1	15	100	128	0,0010	0,40	0,00	[[0. 4][1. 0]]		0,969	0,720	0,842	0,985	0,699	0,982	0,745
5	1	15	100	128	0,0010	0,40	0,00	[[0. 3.5][1. 0.]]		0,969	0,718	0,839	0,986	0,692	0,981	0,747
5	1	15	100	128	0,0010	0,4	0,00		/	0,969	0,718	0,836	0,986	0,686	0,981	0,759
5	1	15	100	128	0,0010	0,40	0,00	[[0. 5][1. 0]]		0,969	0,718	0,836	0,986	0,685	0,981	0,755

Tab 3.1 Top 10 modelli

Osservando i risultati, emerge che i modelli presentano un'accuracy elevata, superiore al 96%, tuttavia, a causa dell'alto sbilanciamento delle classi, questa metrica potrebbe non riflettere necessariamente una performance positiva. L'F1-score, che considera sia la precision che il recall, mostra valori compresi tra il 71% e il 73%, suggerendo

un buon bilanciamento tra la capacità del modello di fare previsioni precise e la sua sensibilità nel rilevare le istanze positive. L'AUC risulta essere circa 0.84 per ogni modello, indicando una buona capacità di discriminazione del modello tra le classi. Inoltre, sia la recall che la precision per entrambe le classi indicano che il modello è in grado di identificare correttamente sia le istanze positive che negative con un alto grado di accuratezza. Infine, è importante notare che nella top 10 sono presenti numerosi modelli che hanno fatto uso del MetaCost, suggerendo che questa tecnica può essere estremamente utile per migliorare le performance dei modelli nella previsione delle crisi finanziarie.

Dopo aver selezionato il miglior modello, questo è stato utilizzato per effettuare predizioni sull'intero dataset impiegato come training set. Questa operazione ha reso possibile l'analisi dei singoli record.

Successivamente, è stata eseguita un'analisi di clustering utilizzando l'algoritmo K-means su due distinti sottoinsiemi del dataset predetto: il primo contenente tutti i record che non sono stati correttamente predetti, e il secondo contenente solo i record correttamente predetti di classe 1.

Per determinare il numero ottimale di cluster k da utilizzare, è stato impiegato l'indice di silhouette, eseguendo K-means per vari valori di k . Questo metodo permette di identificare il valore di k che massimizza la qualità del clustering, valutando quanto i punti dati sono ben raggruppati all'interno dei loro cluster e separati dagli altri cluster. Per entrambi i sottoinsiemi analizzati, il valore migliore di k è risultato essere 3, com'è possibile osservare nei seguenti grafici.

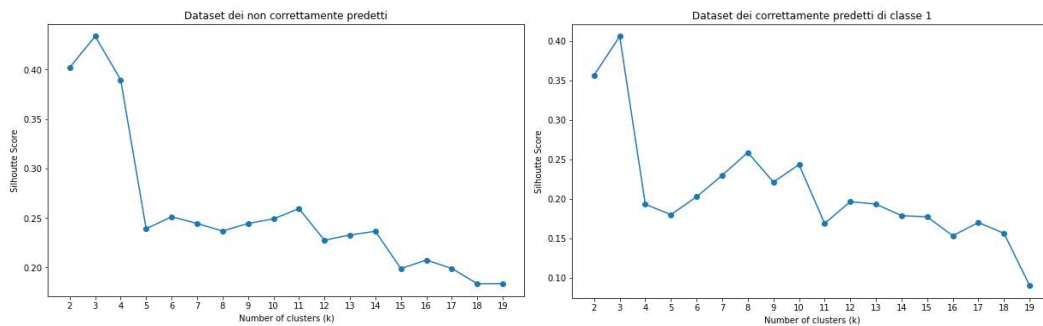


Fig 3.3 e Fig 3.4 Silhouette dei vari k

Dopo aver condotto questa analisi cluster preliminare, si è deciso di abbandonare questo approccio a favore di uno studio più analitico delle differenze nei predittori utilizzando il t-test. Sebbene l'analisi cluster abbia permesso di identificare gruppi di record con caratteristiche simili, non ha fornito una chiara comprensione delle differenze statisticamente rilevanti tra i record di classe 1 correttamente predetti e quelli non predetti correttamente. Pertanto, per approfondire la presenza di differenze

significative nei predittori, sono stati creati due dataset distinti per analizzare le variazioni statisticamente rilevanti tra i record di classe 1 correttamente predetti e quelli non predetti correttamente sempre di classe 1 a differenza dell'altro approccio.

Il primo dataset (Dataset 1) contiene tutti i record di classe 1 che sono stati correttamente predetti dal modello, indicando quindi le previsioni corrette.

Al contrario, il secondo dataset (Dataset 2) include i record di classe 1 che non sono stati predetti correttamente dal modello, ovvero i falsi negativi.

Questi rappresentano i casi in cui il modello ha erroneamente classificato un record di classe 1 come appartenente ad un'altra classe. In questa applicazione specifica, questi errori di previsione sono particolarmente rilevanti poiché comportano i costi economici più elevati.

Per investigare se vi siano differenze significative nei predittori tra questi due gruppi, si è proceduto come segue:

1. **Aggiunta di una Dummy Variable:** È stata aggiunta una variabile dummy per indicare il dataset di appartenenza di ciascun record (0: Dataset 1, 1: Dataset 2).
2. **Regressioni Lineari con Errori Robusti:** Sono state eseguite tante regressioni lineari quanti sono i predittori, utilizzando errori robusti. In ciascuna regressione, la variabile target era rappresentata da uno

dei predittori, mentre le variabili esogene erano la dummy e l'intercetta.

3. **Registrazione dei Risultati:** I risultati delle regressioni, comprensivi dei valori del t-test, p-value e intervalli di confidenza della variabile dummy, sono stati registrati in una tabella. Questo approccio ha permesso di valutare, attraverso i valori del test, se vi fosse una differenza significativa tra i predittori dei due dataset. Di seguito è proposta la tabella risultante:

t-test	P> z	Significance Stars	[0.025	0.975]	Target
0,15777805	0,8746317		-9,67E-64	1,14E-63	crisis0(t-5)
1,42684352	0,153625		-0,85468016	5,42961186	defl_gdp_growth_ifs(t-5)
-0,42496065	0,67086536		-0,04762674	0,03065388	usdextrate_growth_ifs(t-5)
1,72855681	0,08388845	.	-0,65138774	10,3828123	domcrpriv_growth_wdi(t-5)
0,74504359	0,45624537		-0,8033677	1,78868979	real_gdp_growth_wdi(t-5)
-1,81235636	0,06993115	.	-1,02682771	0,04017888	lgdp_per_capita_wdi(t-5)
-0,73979095	0,45942685		-4,31056311	1,948189	m2_to_fereserves_wdi(t-5)
-0,67375571	0,50046664		-3,03824074	1,48376093	realirate_ifs(t-5)
-1,83295019	0,06680998	.	-26,7369678	0,89534423	nfassets_to_gdp_wdi(t-5)
3,48760707	0,00048736	***	1,53E-21	5,46E-21	crisis0(t-4)
1,6908085	0,09087338	.	-0,41936027	5,68812084	defl_gdp_growth_ifs(t-4)
-0,31416732	0,75339396		-0,04651449	0,0336627	usdextrate_growth_ifs(t-4)
2,24983202	0,02445961	*	0,80622901	11,7089824	domcrpriv_growth_wdi(t-4)
2,80213342	0,00507659	**	0,53612613	3,03155723	real_gdp_growth_wdi(t-4)
-1,80094269	0,07171191	.	-1,0003774	0,04229866	lgdp_per_capita_wdi(t-4)
-1,12374778	0,26112004		-5,16938476	1,40179226	m2_to_fereserves_wdi(t-4)
-1,5267011	0,12683535		-3,67171497	0,45625771	realirate_ifs(t-4)
-1,63403842	0,10225078		-25,839464	2,34327667	nfassets_to_gdp_wdi(t-4)
0,59530035	0,55164272		-9,94E-09	1,86E-08	crisis0(t-3)
1,50487354	0,13235659		-0,68742057	5,23368626	defl_gdp_growth_ifs(t-3)
0,55094243	0,58167314		-0,03804714	0,06780081	usdextrate_growth_ifs(t-3)
1,330196	0,1834537		-1,66263881	8,68629046	domcrpriv_growth_wdi(t-3)
3,66609365	0,00024628	***	1,17166971	3,86364616	real_gdp_growth_wdi(t-3)
-1,51247713	0,13041253		-0,969056	0,12488039	lgdp_per_capita_wdi(t-3)
-1,14200976	0,25344996		-5,57291543	1,46951268	m2_to_fereserves_wdi(t-3)
-1,51502197	0,12976684		-3,87173086	0,49574178	realirate_ifs(t-3)
-2,05856712	0,03953573	*	-30,0558881	-0,73748457	nfassets_to_gdp_wdi(t-3)
-3,46165231	0,00053687	***	-0,4315262	-0,11952485	crisis0(t-2)
-0,04510953	0,96402001		-2,9028621	2,7722467	defl_gdp_growth_ifs(t-2)
1,59692332	0,1102828		-0,01194586	0,11703944	usdextrate_growth_ifs(t-2)
1,40238055	0,16080164		-1,48509331	8,9554228	domcrpriv_growth_wdi(t-2)
4,27163609	1,94E-05	***	1,71715498	4,62895369	real_gdp_growth_wdi(t-2)
-1,15337788	0,24875532		-0,89312263	0,2313849	lgdp_per_capita_wdi(t-2)
-1,36793777	0,17133156		-5,706351	1,01514697	m2_to_fereserves_wdi(t-2)
0,30571355	0,75982275		-1,90232714	2,60544582	realirate_ifs(t-2)
-2,08527186	0,03704463	*	-30,4053518	-0,94185609	nfassets_to_gdp_wdi(t-2)
-1,4338E+15	0	***	-1	-1	crisis0(t-1)
-0,60204525	0,54714402		-3,77094262	1,99867884	defl_gdp_growth_ifs(t-1)
1,15713719	0,24721631		-0,02444683	0,09491867	usdextrate_growth_ifs(t-1)
2,3102052	0,0208768	*	0,98425981	12,0001751	domcrpriv_growth_wdi(t-1)
5,51117753	3,56E-08	***	2,56149163	5,38893711	real_gdp_growth_wdi(t-1)
-0,63597982	0,52478958		-0,74464251	0,37978283	lgdp_per_capita_wdi(t-1)
-0,93938199	0,34753466		-4,34489749	1,52942222	m2_to_fereserves_wdi(t-1)
0,55752563	0,57716835		-1,44264737	2,58966802	realirate_ifs(t-1)
-1,9695884	0,04888556	*	-29,8638263	-0,0731437	nfassets_to_gdp_wdi(t-1)

Tab 3.2 Tabelle t-test dei predittori

Uno degli aspetti più rilevanti emersi dall'analisi è che le differenze significative tra i predittori dei due dataset si manifestano principalmente nei lag temporali più recenti. Predittori come Crisis0, Real GDP Growth WDI e NFAssets to GDP WDI mostrano differenze significative soprattutto nei lag t-1, t-2 e t-3. Ad esempio, Crisis0(t-1) e Crisis0(t-2) presentano differenze molto significative ($P < 0.001$), così come Real GDP Growth WDI nei lag t-1, t-2 e t-3. Questo potrebbe suggerire che gli eventi di crisi e la crescita del PIL reale abbiano un impatto più immediato e diretto sulla corretta classificazione della classe 1, influenzando le previsioni in modo più marcato nei periodi più vicini al momento della previsione.

In contrasto, alcuni predittori come Defl GDP Growth IFS e USDExrate Growth IFS non mostrano differenze significative, indipendentemente dal lag temporale. Ciò potrebbe indicare che queste variabili non hanno un impatto rilevante sulla classificazione della classe 1 o che il loro effetto è più distribuito nel tempo, rendendo difficile rilevare differenze significative in singoli lag. Per esempio, Defl GDP Growth IFS(t-5) e USDExrate Growth IFS(t-5) presentano t-test e P-value che non indicano significatività, suggerendo che queste variabili potrebbero avere un'influenza limitata o complessa che non si manifesta chiaramente in un'analisi basata su singoli lag.

I risultati mostrano anche che variabili economiche come Domcrpriv Growth WDI e Real GDP Growth WDI hanno differenze significative, soprattutto in lag come t-1 e t-3. Questo potrebbe essere dovuto al fatto che la crescita del credito privato domestico e del PIL reale sono indicatori economici chiave che riflettono la salute economica generale, influenzando in modo più diretto la capacità di previsione del modello. In particolare, Real GDP Growth WDI(t-1) presenta un t-test di 5.511 con un P-value estremamente basso (3.56E-08), indicando una differenza molto marcata e immediata tra i due dataset. Questo può essere dovuto al fatto che il PIL reale è un indicatore consolidato delle condizioni economiche correnti, influenzando la probabilità di corretta classificazione della classe 1.

Le differenze significative riscontrate nei lag più recenti (t-1, t-2, t-3) possono essere spiegate dalla natura temporale delle variabili economiche e finanziarie. Gli eventi e le condizioni economiche recenti tendono ad avere un impatto più diretto e immediato sulla capacità predittiva di un modello. Pertanto, è più probabile che le variabili nei lag più recenti mostrino differenze significative tra i due dataset. Al contrario, i lag più lontani (t-4, t-5) potrebbero non mostrare differenze significative perché l'influenza di queste variabili può attenuarsi nel tempo o essere confusa da altri fattori intervenienti che agiscono nel periodo intermedio.

Inoltre, alcune variabili potrebbero non mostrare differenze significative perché la loro influenza è meno diretta o più complessa. Ad esempio, i tassi di cambio (USDExrate Growth IFS) possono essere influenzati da molti fattori esterni che rendono difficile rilevare una differenza chiara e diretta tra i due gruppi. Variabili come M2 to Fereserves WDI non mostrano significatività marcata, il che potrebbe indicare che la liquidità relativa alle riserve estere non è un fattore determinante nella differenziazione della capacità predittiva per la classe 1.

L'analisi condotta sembra confermare che le crisi passate influenzano in qualche modo l'errore di previsione, come evidenziato dalle differenze significative tra i dataset nei predittori legati agli eventi di crisi. Tuttavia, dato che il modello include molti ritardi temporali e utilizza una rete neurale molto flessibile, è difficile capire esattamente quale aspetto di questi eventi non venga catturato.

Nonostante alcune variabili economiche e finanziarie mostrino differenze significative, non emergono regolarità chiare che suggeriscano strategie di miglioramento specifiche per i modelli predittivi. La correlazione tra la storia passata delle crisi e i falsi positivi sembra indicare che questi predittori potrebbero rilevare informazioni omesse dal modello corrente. Ciò suggerisce che, piuttosto che concentrarsi esclusivamente sul

miglioramento dei modelli esistenti, gli sviluppi futuri dovrebbero indirizzarsi verso l'arricchimento del dataset.

Arricchire il dataset potrebbe includere l'aggiunta di nuovi predittori o la raccolta di dati più dettagliati e aggiornati, che potrebbero fornire al modello informazioni aggiuntive e migliorare la sua capacità predittiva. Questo approccio potrebbe aiutare a catturare meglio le dinamiche complesse che influenzano gli errori di previsione, riducendo così il tasso di falsi positivi e migliorando la precisione complessiva del modello.

Conclusione

Questo lavoro ha esplorato l'efficacia di modelli di Deep Learning LSTM nell'identificazione e previsione delle crisi finanziarie, impiegando un dataset ampio e complesso. L'analisi è stata strutturata attraverso diverse fasi fondamentali, che hanno incluso la pre-elaborazione dei dati, l'ottimizzazione degli iperparametri e la valutazione delle performance tramite l'uso di metriche specifiche.

I risultati ottenuti mostrano chiaramente che i modelli di Deep Learning utilizzati hanno raggiunto un'alta accuratezza complessiva. Tuttavia, è importante sottolineare che l'accuratezza, da sola, potrebbe non essere un indicatore sufficiente della performance complessiva a causa del possibile sbilanciamento delle classi nel dataset. Per questo motivo, sono state utilizzate metriche aggiuntive per valutare meglio la performance dei modelli, come l'F1-score e l'AUC, che hanno fornito un quadro più completo e dettagliato. Queste metriche hanno indicato una buona capacità dei modelli di bilanciare precisione e sensibilità, e di discriminare efficacemente tra le diverse classi, dimostrando così il loro potenziale nell'ambito delle previsioni finanziarie.

Durante l'intero processo, sono emerse diverse sfide. Un aspetto rilevante è stato rappresentato dall'alta flessibilità dei modelli di rete neurale, che, se da un lato permettono una maggiore capacità di adattamento, dall'altro rendono più complesso comprendere quali specifiche caratteristiche degli eventi di crisi non vengono catturate. Questa complessità intrinseca richiede un'analisi approfondita e un continuo miglioramento delle strategie utilizzate. Inoltre, la natura variabile e spesso imprevedibile delle variabili economiche e finanziarie ha reso difficile sviluppare strategie di miglioramento specifiche per i modelli predittivi, evidenziando la necessità di ulteriori ricerche e raffinamenti metodologici.

Guardando al futuro, diventa cruciale l'arricchimento del dataset utilizzato. Questo potrebbe essere realizzato tramite l'inclusione di nuovi predittori o la raccolta di dati più dettagliati e aggiornati. Un dataset più ricco e diversificato potrebbe fornire al modello informazioni supplementari, migliorando così la sua capacità predittiva e riducendo il tasso di falsi positivi. Inoltre, sarebbe utile considerare l'integrazione di dati provenienti da fonti eterogenee, come dati macroeconomici globali, e dati non strutturati da fonti di notizie finanziarie, che potrebbero fornire ulteriori indizi predittivi e aumentare la robustezza dei modelli.

Oltre all'ampliamento del dataset, un'altra area di potenziale sviluppo è l'adozione di tecniche di machine learning più avanzate. L'esplorazione di modelli di Gated Recurrent Units (GRU) oppure modelli ensemble, che combinano le previsioni di diversi modelli per migliorare la precisione complessiva, potrebbe rappresentare una direzione promettente.

In conclusione, la ricerca svolta ha sottolineato l'importanza di adottare un approccio multifattoriale e dettagliato nella previsione delle crisi finanziarie. I risultati ottenuti indicano chiaramente che, nonostante le sfide e le complessità inerenti ai dati economici, i modelli di Deep Learning rappresentano strumenti potenti e promettenti. Per raggiungere previsioni ancora più accurate e affidabili, è fondamentale continuare a migliorare sia la qualità dei dati che le tecniche analitiche impiegate.

Bibliografia

Antunes, F., Erustes, A. G., Costa, A. J., Nascimento, A. C., Bincoletto, C., Ureshino, R. P., Pereira, G. J. S., & Smaili, S. S. (2018). Autophagy and intermittent fasting: the connection for cancer therapy? *Clinics*, 73, e814s. <https://doi.org/10.6061/clinics/2018/e814s>

Barthélémy, S., Gautier, V., & Rondeau, F. (2024). Early warning system for currency crises using long short-term memory and gated recurrent unit neural networks. *Journal of Forecasting*. <https://doi.org/10.1002/for.3069>

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 145. <https://doi.org/10.1016/j.jinteco.2023.103773>

Brownlee, J. (2020, August 27). The 5 Step Life-Cycle for Long Short-Term Memory Models in Keras. *Machine Learning Mastery*. <https://machinelearningmastery.com/5-step-life-cycle-long-short-term-memory-models-keras/>

Brownlee, J. (2020, October 21). Multivariate Time Series Forecasting with LSTMs in Keras. Machine Learning Mastery. <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>

Brownlee, J. (2021). Imbalanced Classification with Python: Choose Better Metrics, Balance Skewed Classes, and Apply Cost-Sensitive Learning. Independently published.

Casabianca, E. J., Catalano, M., Forni, L., Giarda, E., & Passeri, S. (2022). A machine learning approach to rank the determinants of banking crises over time and across countries. *Journal of International Money and Finance*, 129. <https://doi.org/10.1016/j.jimonfin.2022.102739>

Caggiano, G., Calice, P., Leonida, L., & Kapetanios, G. (2016). Comparing logit-based early warning systems: Does the duration of systemic banking crises matter? *Journal of Empirical Finance*, 37, 104–116. <https://doi.org/10.1016/j.jempfin.2016.01.005>

Demirguc-Kunt, A., & Detragiache, E. (1998). The Determinants of Banking Crises in Developing and Developed Countries. *IMF Staff Papers*, 45(1).

Domingos, P. (1999). MetaCost: a general method for making classifiers cost-sensitive. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 155–164. <https://doi.org/10.1145/312129.312220>

Duttagupta, R., & Cashin, P. (2008). The Anatomy of Banking Crises; Rupa Duttagupta and Paul Cashin; IMF Working Paper 08/93; April 1, 2008.

Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: A comparative approach. Journal of Financial Stability, 4(2), 149–164. <https://doi.org/10.1016/j.jfs.2008.01.001>

Holopainen, M., & Sarlin, P. (2015). Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2584343>

Kaminsky, G., Lizondo, S., & Reinhart, C. (1998). Leading Indicators of Currency Crisis. IMF Staff Papers, 45(1)

Kamalrathne, T., Amaratunga, D., & Haigh, R. P. (2023). Epidemic and pandemic preparedness and response in a multi-hazard context: COVID-19 pandemic as a point of reference.

Ristolainen, K. (2018). Predicting Banking Crises with Artificial Neural Networks: The Role of Nonlinearity and Heterogeneity. *Scandinavian Journal of Economics*, 120(1), 31–62. <https://doi.org/10.1111/sjoe.12216>

Saxena, S. (2024, January 4). What is LSTM? Introduction to Long Short-Term Memory. *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/>

Schularick, M., & Taylor, A. M. (2012). Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008. *American Economic Review*, 102(2), 1029–1061. <https://doi.org/10.1257/aer.102.2.1029>

Shapley, L. S. (1953). Stochastic Games. *Proceedings of the National Academy of Sciences*, 39(10), 1095–1100. <https://doi.org/10.1073/pnas.39.10.1095>

Tölö, E. (2020). Predicting systemic financial crises with recurrent neural networks. *Journal of Financial Stability*, 49, 100746. <https://doi.org/10.1016/j.jfs.2020.100746>

Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). *Python Deep Learning* (2nd ed.). Packt Publishing.

Waidyanatha, N. (2010). Towards a typology of integrated functional early warning systems. *International Journal of Critical Infrastructures*, 6(1), 31.

<https://doi.org/10.1504/IJCIS.2010.029575>

Zommers, Z., & Singh, A. (2014). *Reducing Disaster: Early Warning Systems for Climate Change*. Springer.