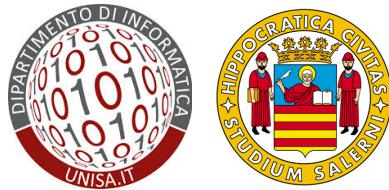


Easy Finances AI

Problem Statement

Francesco Castagnozzi, Vincenzo Mauro

gg/mm/aaaa



Contents

1	Introduzione al progetto	3
2	Definizione e approcci al problema	3
3	Objective & Success Criteria	5
3.1	Obiettivi del progetto	5
3.2	Criteri di successo	5
3.2.1	Valutazione quantitativa dei modelli	5
3.2.2	Valutazione metodologica dei modelli	5
4	Vincoli e limitazioni del progetto	6
5	Ambiti e confini del progetto	6

questa pagina è intenzionalmente lasciata vuota

1 Introduzione al progetto

Easy Finances AI è un progetto di finanza quantitativa che si propone di analizzare e prevedere il comportamento a breve termine di un titolo azionario mediante tecniche di Machine Learning.

Il progetto si affiderà ad alcune parti del processo CRISP-DM per assicurare un approccio strutturato: partendo da Business & Data Understanding, Data Preparation e Revisioning & Evaluation.

I dati considerati includeranno serie storiche di prezzo (open/high/low/close), volume e indicatori tecnici derivati; le analisi prenderanno in considerazione finestre temporali adeguate per il training dei modelli di ML che si andranno ad utilizzare, oltre che ad utilizzare tecniche per evitare bias temporali.

2 Definizione e approcci al problema

Il problema affrontato nel progetto consiste nella previsione del comportamento a breve termine dei titoli azionari di una o più specifiche organizzazioni prese in esame.

Questo progetto consente di analizzare e confrontare il comportamento di diversi approcci di Machine Learning in un contesto realistico e complesso come quello finanziario, ponendo particolare attenzione alla capacità di generalizzazione dei modelli e alla loro robustezza su dati non stazionari.

Sui dati finanziari storici degli ultimi 10 anni dell'organizzazione, verranno utilizzate tecniche di apprendimento supervisionato per addestrare un modello di ML in modo da renderlo capace di stimare la variazione percentuale del titolo nel giorno successivo in base a ciò che viene osservato precedentemente.

Formalmente, il dataset è costituito da una serie temporale finanziaria così definita:

$$\{X_t\}_{t=1}^T$$

dove ogni componente della serie X_t rappresenta lo stato del mercato al tempo t ed è descritta da feature derivate dai dati storici di prezzo (come open, high, low, close), volumi di scambio e indicatori valutati in finestre di tempo appropriate.

Come detto prima, l'orizzonte temporale di osservazione è di 10 anni, mentre l'orizzonte di previsione è fissato a un giorno ($h = 1$).

Inoltre, è stata definita una variabile custom che quantifica la variazione percentuale del titolo nel giorno $t+1$: **Target_Return**. Questa variabile viene utilizzata in due formulazioni del problema:

- Regressione: il modello apprende una funzione che stima la variazione percentuale definita come

$$f_{\theta} : X_t \rightarrow \mathbf{R}$$

Per questo tipo di modello verranno utilizzate metriche di errore quali RMSE e MAE per valutare i dati generati.

- Classificazione: il modello apprende una funzione, di seguito definita

$$g_{\theta} : X_t \rightarrow \mathbf{C}$$

che prevede la direzione del movimento in borsa (up/down), etichettando il valore in base ai dati originali e sui valori predetti nella regressione. Questo modello verrà valutato con le classiche metriche di valutazione Precision, Accuracy, Recall e F1-score.

3 Objective & Success Criteria

3.1 Obiettivi del progetto

Il progetto Easy finances AI non si propone di sviluppare un prodotto capace di prevedere i movimenti del mercato azionario, il nostro è un progetto accademico, mirato a studiare e sviluppare modelli di regressione e classificazione utilizzando approcci di machine learning differenti.

Inoltre, verranno valutate le capacità generative dei modelli su dati temporali non utilizzati durante il training, la robustezza delle prestazioni rispetto a lievi cambiamenti negli intervalli temporali di test. Infine, verrà applicato il processo CRISP-DM, garantendo tracciabilità e una corretta metodologia.

3.2 Criteri di successo

3.2.1 Valutazione quantitativa dei modelli

- **Regressione:** il modello di regressione sarà considerato efficace nel momento in cui i valori del RMSE e del MAE risultano inferiori alle baseline di riferimento.
- **Classificazione:** per il modello di classificazione la valutazione avverrà riferendosi alle metriche Precision, Recall, Accuracy e F1-Score, oltre che al bilanciamento delle classi e delle feature.

Infine, le prestazioni saranno stimate con dei test basati su dati non presenti nel dataset di training, simulando una previsione effettuata con dati "reali".

3.2.2 Valutazione metodologica dei modelli

Il progetto avrà successo nel momento in cui verranno evitati fenomeni di leakage informativo e bias temporali, entrambi raggiungibili utilizzando procedure di validazione adeguate alle serie temporali.

I modelli scelti e i risultati sperimentali da loro generati saranno analizzati in modo critico, evidenziando eventuali instabilità, errori e limiti.

La valutazione del successo del progetto quindi non si attenderà solo al raggiungimento di specifici valori metrici, ma riguarderà anche la coerenza dell'analisi e la consapevolezza dei limiti della previsione finanziaria.

4 Vincoli e limitazioni del progetto

Il progetto è soggetto ad una serie di vincoli intrinseci del mercato azionario e al dominio finanziario che sono fuori dal controllo dei progettisti. Di seguito questi vincoli definiscono chiaramente l'ambito del progetto e ne circoscrivono i risultati al fronte di evita misinterpretazioni degli stessi:

- **Caoticità dell'ambiente:** i dati utilizzati riflettono un contesto di mercato caratterizzato da elevato rumore, non stazionarietà e possibili cambi di regime, questo ambiente caotico limita la capacità dei modelli di apprendere relazioni stabili e generalizzabili nel tempo, favorendo la comparsa di bias ed errori nelle previsioni.
- **Fattori esterni non considerati:** l'analisi dei dati si basa esclusivamente su informazioni storiche di prezzo, volume e indicatori tecnici derivati, dati facilmente ottenibili e non influenzati da fattori esogeni quali notizie macroeconomiche, eventi societari o dinamiche geopolitiche, che possono influenzare significativamente l'andamento dei titoli azionari e il comportamento del mercato.
- **Limiti metodologici:** l'orizzonte di previsione è limitato ad un solo giorno e le prestazioni dei modelli sono valutate con dei test out-of-sample che, pur garantendo una stima più realistica della capacità di generalizzazione, non eliminano l'incertezza intrinseca del problema.

5 Ambiti e confini del progetto

L'ambito del progetto è definito sull'analisi metodologica e sperimentale di tecniche di Intelligenza Artificiale supervisionata per la previsione a breve termine di un titolo azionario, escludendo volutamente tutti gli aspetti che non rientrano negli obiettivi del lavoro.

Il progetto non affronterà aspetti come l'ottimizzazione automatica degli iperparametri su larga scala, la selezione automatica dei modelli o l'uso di approcci di apprendimento online, ma si limiterà ad un contesto accademico di addestramento e valutazione.

Altri confini del progetto riguardano l'assenza di analisi approfondite sulla sensibilità delle feature, la mancata esplorazione di tecniche avanzate di riduzione della dimensionalità o di interpretabilità dei modelli, e la non considerazione di aspetti etici, normativi o socio-economici legati all'uso di sistemi predittivi in finanza. Queste scelte permettono di concentrarsi sugli aspetti fondamentali dell'apprendimento supervisionato e sul confronto controllato tra modelli, chiarendo che i risultati devono essere interpretati solo all'interno del contesto metodologico e sperimentale definito dal progetto.