|  |
| --- |
| Progetto  machine learning  Analisi e predizione |
| Manuel Tanzi - 307720 E-mail: manuel.tanzi1@studenti.unipr.it  Progetto GitLab:  <https://gitlab.com/tanzone2000/machine-learning.git>  Dataset Riferimento:  <https://www.kaggle.com/mattiuzc/stock-exchange-data> |





## Introduzione

### **Argomento**

Si utilizza un Dataset creato tramite i dati generati dalle API di yahoo-finance che fornisce notizie finanziarie, dati e commenti, tra cui le quotazioni in borsa.

I dati forniti interessano il rendimento di diversi anni di andamento di alcuni titoli e azioni, in particolare del valore della moneta di tutto il mondo divisa per simbolo di appartenenza.

Si vuole quindi avere una rappresentazione chiara dell’andamento accompagnata da alcune informazioni necessarie e utili per l’analisi delle stesse.

Inoltre, si cerca di eseguire un’analisi predittiva tramite alcuni degli algoritmi di machine learning visti a lezione implementate attraverso diverse tecniche e strategie.

### **Procedimento**

Si utilizza una scaletta prefissata per la realizzazione del progetto:

1. **Raccolta dei dati**
2. **Esplorazioni dati**
3. **Data cleaning**
4. **Feature selection**
5. **Feature engineering**
6. **Feature scaling**
7. **Costruzione modelli**
8. **Confronto modelli**
9. **Analisi prestazioni**

Si è deciso di non utilizzare algoritmi come T-sne e PCA poiché per questo tipo di problema risulta utile semplicemente una visualizzazione a grafico XY senza preoccuparsi di Curse of Dimensionality.

### **Sviluppo**

Durante lo sviluppo vengono eseguiti molteplici casi, dai più irreali ad una possibile condizione realistica in cui si prova a predire un certo numero di giorni futuri stabili da una costante presente nel file apposito (constants.py).

La maggior parte delle funzioni sono generalizzate e parametrizzabili in maniera tale da avere una maggior libertà per simulare i vari modelli attraverso più combinazioni possibili e sperimentare.

### **Considerazioni**

Il sistema da analizzare è un processo aleatorio non markoviano ovvero che l’istante al tempo t non dipende dalla storia passata. Questo fa si che i modelli di Machine Learning non riconoscano un pattern ricorrente e quindi costruirne una funziona adeguata solamente utilizzando i dati forniti.

### **Ipotesi**

### La regressione è il processo per trovare un modello o una funzione per distinguere i dati in valori reali continui.

### Matematicamente, con un problema di regressione, si sta cercando di trovare l'approssimazione della funzione con la deviazione minima dell'errore. Nella regressione, si prevede che la dipendenza numerica dei dati la distingua.

### L’analisi di regressione è il modello statistico che viene utilizzato per prevedere i dati numerici anziché le etichette ma può anche identificare il movimento di distribuzione in base ai dati disponibili o ai dati storici.

### Si ipotizzi invece di aggiungere due features (BUY, SELL) caratterizzate da un valore identificativo [0, 1], calcolati secondo alcuni algoritmi che studiano l’andamento medio come SMA, EMA, STD, MEAN e CHANGE.

### A questo punto si potrebbe eseguire un task di classificazione per prevedere se nell’immediato futuro sia meglio eseguire un’azione di BUY o SELL.