

UE19CS312: Data Analytics – Assignment

Problem Statement

Stock market prediction is defined as “the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange”. Predicting future values of stock prices can yield great profits for the company. However, the task is not particularly an easy one as it can involve numerous factors (both physical and psychological), making share prices particularly volatile. Investment banks are institutions that act in an advisory role to help companies, individuals or governments make financial transactions.

Major shareholders of the media conglomerate, Waystar Royco (WAYA US) have hired the investment bank you work at, as consultants to analyse their performance in the stock market and make future predictions. Your team, as business analysts in the investment bank, has to predict the closing stock price of the company from 30th July, 2021 to 10th September, 2021 as accurately as possible, given data from 14th August, 2015 to 29th July, 2021. The data you have been provided with, consists of the opening, high, low and closing price of Waystar Royco’s stock as well as the volume of stocks traded in a day, for 1500 days. Your team has been instructed to use ONLY this information and disregard any other factors while making predictions. In addition to this, the bank also wants to know what models are best suited for stock price predictions given similar data and wants your team to perform a comparison between regression and time series models so that they know what to adopt for future scenarios.

DISCLAIMER: Waystar Royco is a fictional company, so any information apart from what is given in the description above, is irrelevant to the problem statement at hand.

This is what you need to do...

- Analyse the dataset provided
- Model using regression and time series techniques
- Provide comparative analysis based on performance metrics and technical domain knowledge (this needs to be done in your kernels itself, with the use of suitable documentation and/or graphs)
- Predict closing stock prices from 30/07/2021 to 10/09/2021

Deliverables

1. A submission file (.csv) with predictions
2. Python or R file containing the executable code with detailed comments

Evaluation: The metric that will be used to evaluate your submission is RMSE

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

Dataset: [train.csv](#), [test.csv](#)

Submission format: [sample file](#)

Deadline: 7:00pm IST, November 6, 2021
