

Assignment 4

Data Preparation

‘Mine Dime’, a stock trend forecasting company has just employed you as a Data Scientist. As a first task in your new job, your manager has provided you with a company’s stock data and asked you to check the quality of the data for the next step of analysis. Following are the additional description and information about the data which your manager has shared with you.

- a) The data set contains six variables namely-
 - i. Date
 - ii. Open
 - iii. High
 - iv. Low
 - v. Close
 - vi. Volume
- b) Typically, the stock market opens at 9:15 hours and closes at 15:30 hours. Each stock is defined by an opening price and a closing price which are the prices it opens and closes with. Within the operating hours, the stock price touches a maximum and minimum which are the highest and lowest prices achieved by the stock in the working hours of the stock market. You have access to ten years of monthly stock price data with the Open, High, Low and Close price and the number of stocks traded for each day given by the feature Volume. On some days when there is no trading, the parameters Open, High, Low and Close remain constant and Volume is zero.

Furthermore, your manager also claims that the model prediction is too bad since the data is polluted. Try to impress your new boss by preprocessing the data and by giving a proper rationale behind the steps you would follow. The two datasets should be merged before preprocessing.