Stock Price Prediction

Kritika Karamchandani, Rishabh Kasat, Siddharth Singh
Instructor : German Creamer

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®
Business Intelligence & Analytics

Introduction

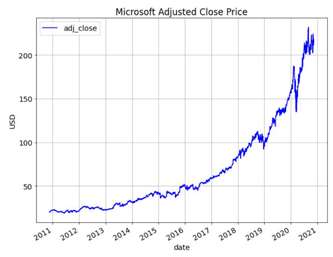
- Time series Analysis is one of the most challenging and interesting application of Machine Learning Algorithms
- Stock price prediction is an enticing prospect for data scientist to evaluate and predict future prices
- Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. psychological, rational and irrational behaviour, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.
- We see daily ups and downs in the market and assume a pattern which is highly dependent on sector as well as the company

Objective

- To predict tech industry stock price for next day and to help investors evaluate their portfolio
- Minimize the prediction error to build an accurate model

Data Understanding

- 10-year daily trading for Amazon, Google and Microsoft was collected from yahoo finance

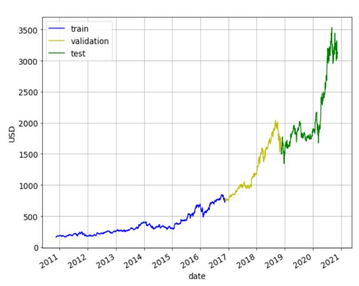


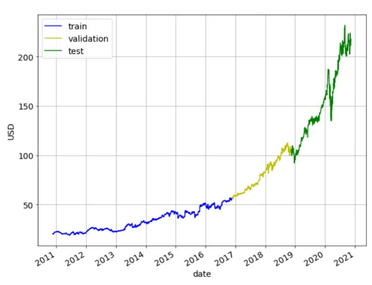
- The above graphs shows the adjusted closing price for Amazon,Microsoft and Google

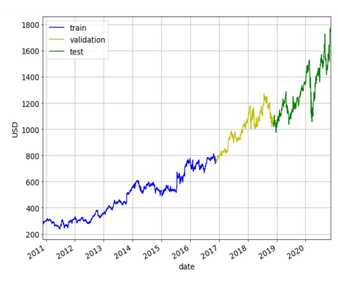
Data Preparation

- The data was divided into 3 sets:
- Training Set – 60% of the data
- Validation Set – 20% of the data
- Test Set – 20% of The data

- Since we know stock price are highly volatile and the range of data in training set , test set and validation set is different, We normalized the data using previous 5 days rolling average to have mean 0 and standard deviation 1 for all the data.



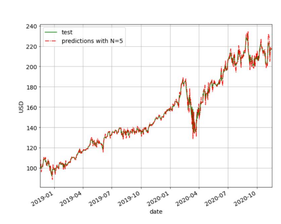
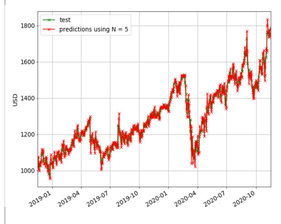
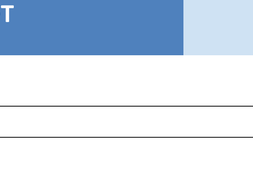
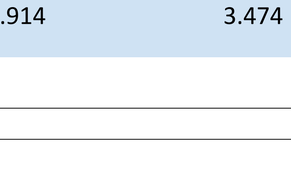
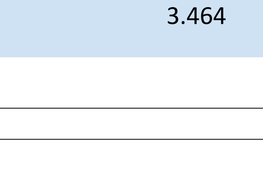
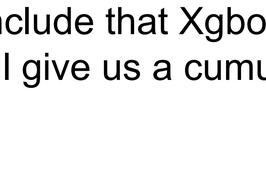
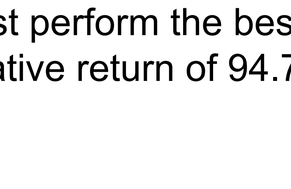
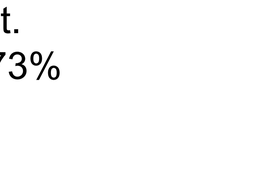




Modeling

- Linear Regression
- XgBoost
- Random Forest
- For the three algorithms we added additional features like lag for previous N days and OHLC difference for previous N days to know the importance of different features and relevance of previous lags.
- Features as input to our model:
  1. Difference between High and low of stock for past n days
     e.g:High-Low_lag_1 ,High-Low_lag_2,High-Low_lag_3
  2. Difference Between Open and close for previous n days
     e.g:Open-Close_lag_1,Open-Close_lag_2
  3. Difference Between Open and close for previous n days
     e.g:Open-Close_lag_1,Open-Close_lag_2
  4. Trading volume lag for previous n days
     e.g:Volume_lag_1,Volume_lag_2,Volume_lag_3
- We normalised the training, testing and validation set to have mean 0 and variance 1

Results & Evaluation

AMAZON	MICROSOFT	GOOGLE
Linear Regression	Linear Regression	Linear Regression
		
XGBOOST	XGBOOST	XGBOOST
		
Random Forest	Random Forest	Random Forest
		
The above graphs shows actual stock price and predicted stock price	The above graphs shows actual stock price and predicted stock price	The above graphs shows actual stock price and predicted stock price

RMSE VALUES	Linear Regression	RF Regressor	XGBoost
GOOG	32.621	26.249	26.227
AMZN	49.343	49.982	49.337
MSFT	3.914	3.474	3.464

Conclusion

- Using Feature Engineering and testing data on three different models we conclude that Xgboost perform the best.
- Our Investment of $100 in each stock through the test period (2018-2020) will give us a cumulative return of 94.73%
- S&P500 return is 31.65 % in the same period