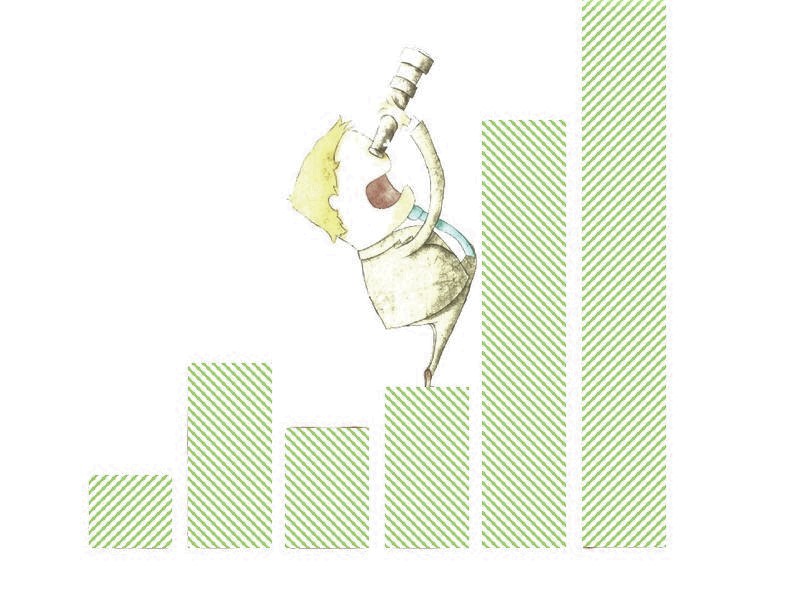
**Stock Price Prediction Using Historical Data**

***By: - Vaibhav Gaur, Shubham Sood and Lisha Uppal***

***Mentored By: - Ms. Manpreet Kaur***



**ABSTRACT:**

Sleek accuracy and precision are one of the most sought-after things when it comes to stock predictions, though stocks might not hold as much of an importance to normal folks as they do to stocks shelves, their ever-changing nature coupled with a risk factors which always seems much of a gamble, makes it worth-while for us to analyze them with a bunch of Machine Learning Algorithms. We thus seek to provide a machine-based approach replacing the tradition yet unreliable time-series forecasting to predict the stocks from the past historical data of a big firm and determine the underlying patterns to improve upon the same.

After a prolonged deliberation, we decide to use a couple of conventional machine learning algorithms including but not limited to Decision Tree, Logistic Regression, and K-Nearest Neighbors to bet on the algorithm that best predicts the stock prices and hence help in determining the underlying patterns to draw conclusion from to devise strategies in improving the stocks.

**INTRODUCTION:**

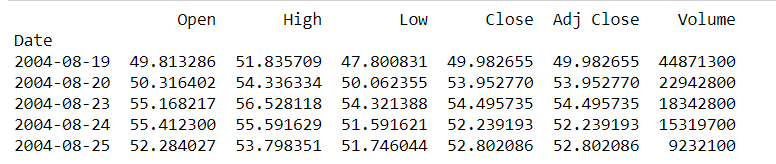
Over the last two decades, humans have grown a lot of dependence on data and information in society and with this advent growth, technologies have evolved for their storage, analysis and processing on a huge scale. The fields of Data Mining and Machine Learning have not only exploited them for knowledge and discovery but also to explore certain hidden patterns and concepts which led to the prediction of future events, not easy to obtain. And one of the difficult things to predict that caught our attention is stock or commonly called as shares.

Stock price prediction is one of the most important topics to be investigated in academic and financial researches. Various Data mining techniques are frequently involved in the studies. To solve this problem. But technique using machine learning/deep learning will give more accurate, precise and simple way to solve such issues related to stock and market prices.

There are ample amounts of data about stocks but the most difficult and intriguing thing is to predict the price of these stocks based on old data and this is what we have done in our project.

**DATASET:**

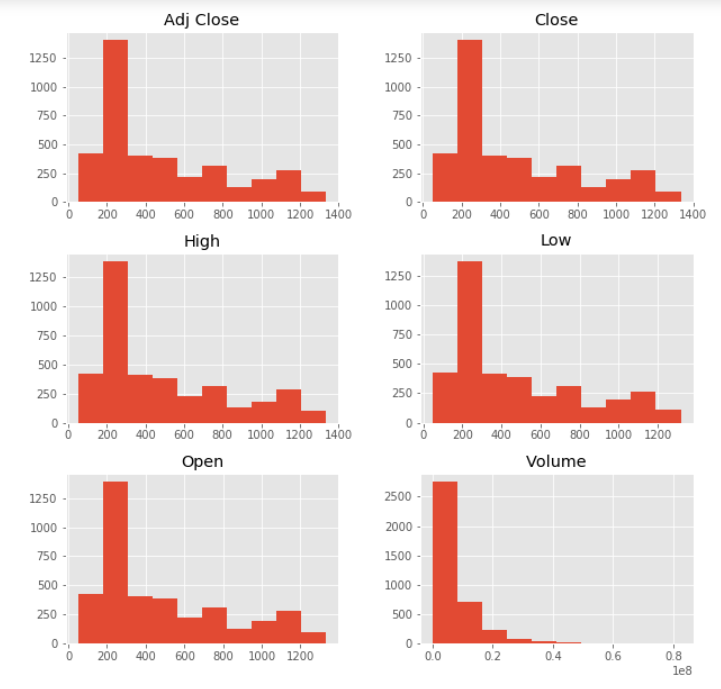
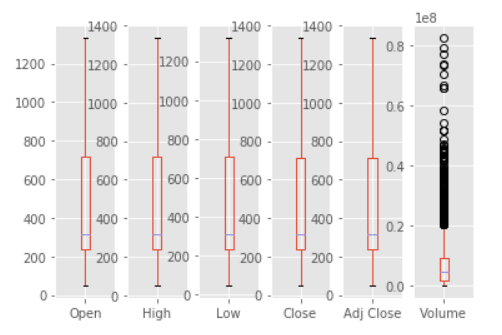
In this project we have mainly used data consisting of stock prices for the well-known company Google from Yahoo! Finance from the year 2004 to November 2019. This includes Date, Open, High, Low, Close, Adj Close and Volume for a given day.

****

**METHODS:**

**Step 1: Data Visualization**

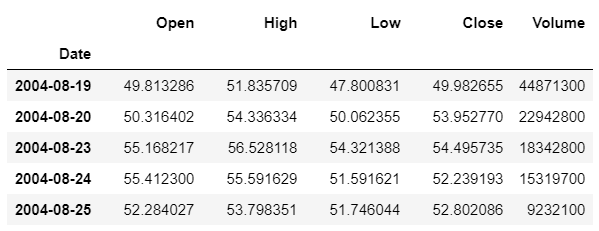
We have plotted a box plot as shown in a Figure 1. That shows the mean of each attribute and the highest and lowest value they take. We also plotted a histogram as shown in Figure 2 for every attribute of the data to observe the dependency of stocks on the given attributes.



*Figure 1: Box plot of all the attributes Figure 2: Histogram of all the attributes*

**Step 2: Data Preprocessing**

In data preprocessing we dropped the unnecessary attributes from the given set of attributes and the final data is shown in Figure 3.



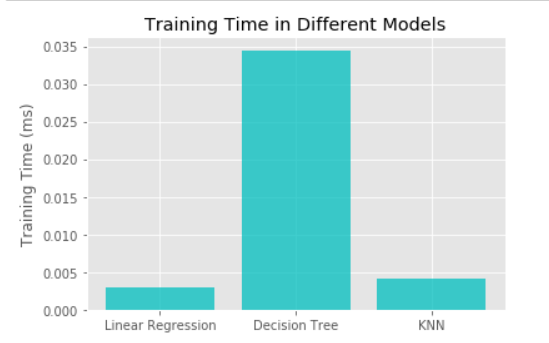
*Figure 3: Adjusted Close attribute dropped from the given attribute*

**Step 3: Model Training**

We are able to use different algorithms like Decision Tree, KNN and Linear regression to train the model and then obtain the accuracy of each model. The accuracies of the said algorithms are computed in the table below.

|  |  |
| --- | --- |
| **Algorithm** | **Accuracy (%)** |
| Linear Regression | 98.10 |
| Decision Tree | 97.52 |
| KNN | 70.18 |

We also tried to compute the time an algorithm take to train the dataset and the computation time can be compared from the Figure 4. We observed that Linear Regression takes the least time to train the dataset whereas Decision Tree takes the maximum time.



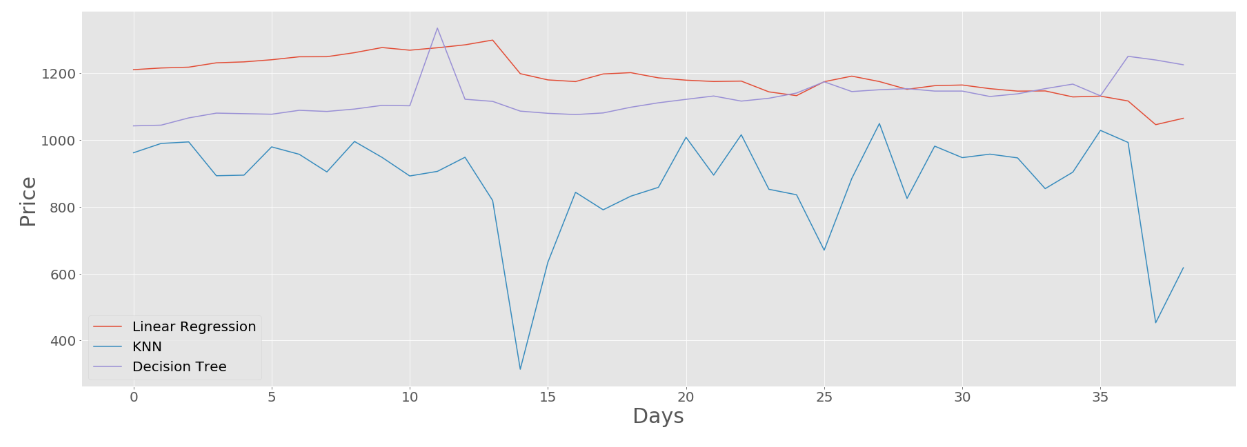
*Figure 4: Comparison in Training time for different models*

**EXPERIMENTS/RESULTS/DISCUSSION:**

For our project, we used several common mathematical and computational libraries like matplotlib, NumPy, pandas and sklearn. For our code we looked through various tutorials and documentations for references.

We tried to use different value of k for k-nn algorithm and it gives the highest accuracy in case of k = 13.

It can be clearly seen that Linear Regression had the maximum accuracy out of the three algorithms applied of 98%. Thus, the stock prices as predicted by the model are 98% correct.



*Figure 4: Depiction of stock prices as predicted by different algorithms.*

**CONCLUSION AND FUTURE WORK:**

By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the linear regression algorithm. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are considered more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation’s performance structure as a whole.

**CONTRIBUTION:**

Vaibhav Gaur gathered the dataset of Google’s Stock prices from Yahoo! Finance and preprocessed it so that unnecessary data can be cleaned and then implemented the Linear Regression on the data and concluded that it predicted the stock prices more accurately.

Shubham Sood did the data visualization and comparison graphs to see the pattern and trends in the data; after that, computation of training time needed for each algorithm is carried out, and the k-nn algorithm is applied to the data.

Lisha Uppal did the literature survey about the project so that we can get the knowledge about our project and we can explore about it then she applied Decision Tree algorithm and observed that it had the fastest execution time but takes the maximum time in training of the dataset moreover she made the proposal and report of the project.

**REFERENCES:**

[1] S. A. R. Nai-Fu Chen and Richard Roll, *Economic Forces and the Stock Market, The Journal of Business, vol. 59, no. 3, pp. 383–403, (1986)*.[Online]. Available: http://www.jstor.org/stable/2352710.

[2] E. F. Fama, *Random Walks in Stock Market Prices, Financial Analysts Journal, vol. 51, no. 1, pp. 75–80, (1995)*.[Online]. Available: <http://www.jstor.org/stable/4479810>.

[3] S. J. Grossman and R. J. Shiller, *The Determinants of the Variability of Stock Market Prices, National Bureau of Economic Research, Working Paper 564, October (1980)* [Online]. Available: <http://www.nber.org/papers/w0564>.

[4] A. W. Lo and A. C. MacKinlay, *Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, Review of Financial Studies, vol. 1, no. 1, pp. 41–66, (1988)*.

[5] *P. P¨a¨akk ¨onen and D. Pakkala, Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems, Big Data Research, vol. 2, no. 4, pp.166–186,(2015)*.[Online].Available: <http://www.sciencedirect.com/science/article/pii/S2214579615000027>.

[6] P. A. G. Xue Zhang and Hauke Fuehres, *Predicting Stock Market Indicators through Twitter I Hope it is not as Bad as I Fear, Procedia – Social and behavioral Sciences, vol. 26, pp. 55–62, (2011)*.