

MINI PROJECT – II
(2019-2020)

Stock Price Prediction
Group No-

FINAL REPORT



Institute of Engineering & Technology

Team Members

Ishan Rathi (171500144)

Ayush Kumar Singh (171500075)

Ayush Sinha (161500163)

Supervised By

Mr. Pankaj Sharma

Asst. professor

Department of Computer Engineering & Applications



**Department of computer Engineering and
Applications**

GLA University, Mathura

17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha,

Declaration

I hereby declare that the work which is being presented in the Mini Project Titled: “Stock Price Prediction”, in fulfillment of the requirements for Mini-Project LAB, is an authentic record of our own work carried under the supervision of Mr. Pankaj Sharma, GLA University, Mathura

Signature of Candidate:

Name of Candidate: Ishan Rathi, Ayush Kumar Singh, Ayush Sinha

Roll. No. : 171500144, 171500075, 161500163

Course: B. Tech

Year: 3rd

Semester: VI

ACKNOWLEDGMENT

It gives us a great sense of pleasure to present the report of the B. Tech Mini Project undertaken during B. Tech. Third Year. This project in itself is an acknowledgement to the inspiration, drive and technical assistance contributed to it by many individuals. This project would never have seen the light of the day without the help and guidance that we have received.

Our heartiest thanks to Dr. (Prof). Anand Singh Jalal, Head of Dept., Department of CEA for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal.

We owe special debt of gratitude to Mr. Pankaj Sharma , Department of CEA, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. He has showered us with all his extensively experienced ideas and insightful comments at virtually all stages of the project & has also taught us about the latest industry-oriented technologies.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

CONTENT

1. Abstract
2. Introduction
3. Objective
4. Problem Statement
5. Motivation
6. Future Prospects
7. Requirements
8. Implementation
9. Conclusion
10. References

ABSTRACT

In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is Used to predict the stock market using machine learning is Python. In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a different machine learning techniques to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

INTRODUCTION

Basically, quantitative traders with a lot of money from stock markets buy stocks derivatives and equities at a cheap price and later on selling them at high price. The trend in a stock market prediction is not a new thing and yet this issue is kept being discussed by various organizations. There are two types to analyze stocks which investors perform before investing in a stock, first is the fundamental analysis, in this analysis investors look at the intrinsic value of stocks, and performance of the industry, economy, political climate etc. to decide that whether to invest or not. On the other hand, the technical analysis it is an evolution of stocks by the means of studying the statistics generated by market activity, such as past prices and volumes. In the recent years, increasing prominence of machine learning in various industries have enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results. This paper will develop a financial data predict or program in which there will be a dataset storing all historical stock prices and data will be treated as training sets for the program. The main purpose of the prediction is to reduce uncertainty associated to investment decision making. Stock Market follows the random walk, which implies that the best prediction you can have about tomorrow's value is today's value. Indisputably, the forecasting stock indices is very difficult because of the market volatility that needs accurate forecast model. The stock market indices are highly fluctuating and it effects the investor's belief. Stock prices are considered to be a very dynamic and susceptible to quick changes because of underlying nature of the financial domain and in part because of the mix of a known parameters (Previous day's closing price, P/E ratio etc.) and the unknown factors (like Election Results, Rumors etc.). There has been numerous attempts to predict stock price with Machine Learning. The focus of each research projects varies a lot in three ways. (1) The targeting price change can be near-term (less than a minute), short-term (tomorrow to a few days later), and a long-term (months later), (2) The set of stocks can be limited to less than 10 particular stock, to stocks in particular industry, to generally all stocks. (3) The predictors used can range from a global news and economy trend, to particular characteristics of the company, to purely time series data of the stock price. The probable stock market prediction target can be the future stock price or the volatility of the prices or market trend. In the prediction there are two types like dummy and a real time prediction which is used in stock market prediction system. In Dummy prediction they have define some set of rules and predict the future price of shares by calculating the average price. In the real time prediction compulsory used internet and saw current price of shares of the company. Computational advances have led to introduction of machine learning techniques for the predictive systems in financial markets. In this paper we are using a Machine Learning techniques i.e., Support Vector Machine (SVM), Linear Regression etc. in order to predict the stock market and we are using Python language for programming.

OBJECTIVE

Build a model using machine learning algorithms for predicting the stock price.

Problem Statement:

Financial analysts investing in stock market usually are not aware of the stock market behavior. They are facing the problem of trading as they do not properly understand which stocks to buy or which stocks to sell in order to more profits. In today's world, all the information pertaining to stock market is available. Analyzing all this information individually or manually is tremendously difficult. As such, automation of the process is required. This is where data mining techniques help.

Motivation:

Nowadays, as the connections between worldwide economies are tightened by globalization, external perturbations to the financial markets are no longer domestic. With evolving capital markets, more and more data is being created daily.

The intrinsic value of a company's stock is the value determined by estimating the expected future cash flows of a stock and discounting them to the present, which is known as the book value. This is distinct from the market value of the stock that is determined by the company's stock price. This market value of a stock can deviate from the intrinsic value due to reasons unrelated to the company's fundamental operations, such as market sentiment.

The fluctuation of stock market is violent and there are many complicated financial indicators. Only few people with extensive experience and knowledge can understand the meaning of the indicators and use them to make good prediction to get fortune. Most people have to rely solely on luck to earn money from stock trading. However, the advancement in technology, provides an opportunity to gain steady fortune from stock market and also can help experts to find out the most informative indicators to make better prediction. The prediction of the market value is of paramount importance to help in maximizing the profit of stock option purchase while keeping the risk low.

Future Prospects:

- It is used to predict the future value of company stocks.
- Can be used to predict the price hike
- Can be beneficial for stock buyer as they get to know the future hike of stock
- Explaining stocks value to others

Requirements:

a) Hardware:

- Computer/Laptop
- 8GB RAM
- Processor i3 or more

b) Software:

- Jupyter
- Anaconda Navigator

c) Technology:

- Python
- Numpy
- Sklearn
- Pandas

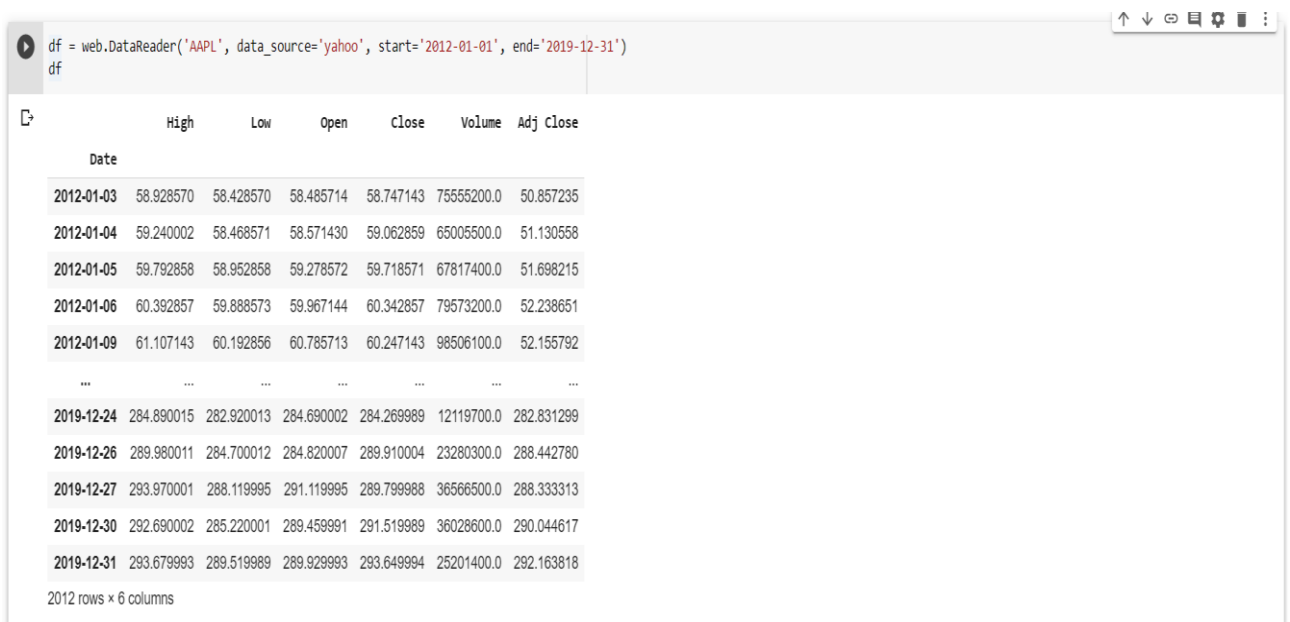
IMPLEMENTATION

1. Dependencies Needed

- python 3.6+
- numpy
- pandas
- matplotlib
- scikitplot
- sklearn

2. Loading data

UCI repository provide the stock price data and start preparing a machine learning model for predicting future stock price. First we have to label the data according to the algorithm.



```
df = web.DataReader('AAPL', data_source='yahoo', start='2012-01-01', end='2019-12-31')
df
```

Date	High	Low	Open	Close	Volume	Adj Close
2012-01-03	58.928570	58.428570	58.485714	58.747143	75555200.0	50.857235
2012-01-04	59.240002	58.468571	58.571430	59.062859	65005500.0	51.130558
2012-01-05	59.792858	58.952858	59.278572	59.718571	67817400.0	51.698215
2012-01-06	60.392857	59.888573	59.967144	60.342857	79573200.0	52.238651
2012-01-09	61.107143	60.192856	60.785713	60.247143	98506100.0	52.155792
...
2019-12-24	284.890015	282.920013	284.690002	284.269989	12119700.0	282.831299
2019-12-26	289.980011	284.700012	284.820007	289.910004	23280300.0	288.442780
2019-12-27	293.970001	288.119995	291.119995	289.799988	36566500.0	288.333313
2019-12-30	292.690002	285.220001	289.459991	291.519989	36028600.0	290.044617
2019-12-31	293.679993	289.519989	289.929993	293.649994	25201400.0	292.163818

2012 rows x 6 columns

3. Graph of Original Data



4. Algorithm Used

1. Support Vector Machines

SVM is the newest extremely fast machine learning (data mining) algorithm for solving multiclass classification problems from ultra large data sets that implements an original proprietary version of a cutting plane algorithm for designing a linear support vector machine. SVM is a linearly scalable routine meaning that it creates an SVM model in a CPU time which scales linearly with the size of the training data set. Our comparisons with other known SVM models clearly show its superior performance when high accuracy is required. We would highly appreciate if you may share SVM performance on your data sets with us.

Features:

- Efficiency in dealing with extra large data sets (say, several millions training data pairs)
- Solution of multiclass classification problems with any number of classes
- Working with high dimensional data (thousands of features, attributes) in both sparse and dense format
- No need for expensive computing resources (personal computer is a standard platform)

- Ideal for contemporary applications in digital advertisement, e-commerce, web page categorization, text classification, bioinformatics, proteomics, banking services and many other areas.

2. Linear Regression

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog).

5. Prediction values by Svm and Linear Regression

```
[28] #Show the SVM model Prediction
      svm_prediction = svr_rbf.predict(x_future)
      print('SVM MODEL PREDICTION',svm_prediction)
      print()
      #Show the Linear Regression Prediction
      lr_prediction = lr.predict(x_future)
      print('LINEAR REGRESSION MODEL PREDICTION',lr_prediction)
```

```
☞ SVM MODEL PREDICTION [227.99819542 223.44694864 227.94138466 234.32396111 249.41342736
248.54864701 247.14711585 244.72815923 247.0451143 249.92106048
259.86592534 258.60762282 265.47989423 266.2415923 271.23149431
274.40421009 265.69155332 265.63399511 274.07180635 280.30353931
281.64414064 281.33878782 281.42881975 283.3827637 284.10059164
286.48478114 286.18194112 289.70036795 287.05806951 291.81247163]
```

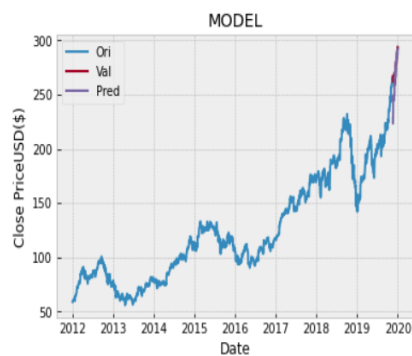
```
LINEAR REGRESSION MODEL PREDICTION [232.31217745 229.59449903 232.28152823 235.4078733 241.66057903
241.31319523 240.75128248 239.78067191 240.71040647 241.86491235
246.05380006 245.49188731 248.78168969 249.19037193 252.25541856
254.77897486 248.89407535 248.86342614 254.48267829 261.69576779
263.41218642 263.03416899 263.14653907 265.3840256 266.10944242
268.21410528 267.96888034 270.53331267 268.66364794 271.85129145]
```

6. Graph made by SVM Model

```
[46] valid = df[X.shape[0]:]
      valid['Prediction'] = svm_prediction
      plt.title('MODEL')
      plt.xlabel('Date')
      plt.ylabel('Close PriceUSD($)')
      plt.plot(df['Close'])
      plt.plot(valid[['Close', 'Prediction']])
      plt.legend(['Ori', 'Val', 'Pred'])
      plt.show()
```

⚠ /usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

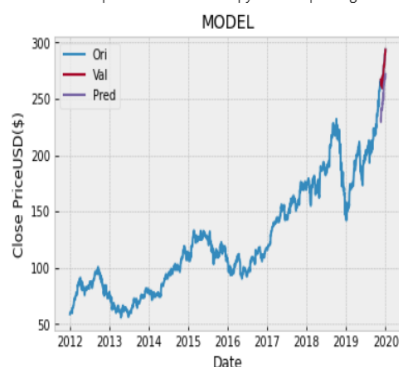


7. Graph made by Linear Regression

```
[47] prediction = lr_prediction
      valid = df[X.shape[0]:]
      valid['Prediction'] = prediction
      plt.title('MODEL')
      plt.xlabel('Date')
      plt.ylabel('Close PriceUSD($)')
      plt.plot(df['Close'])
      plt.plot(valid[['Close', 'Prediction']])
      plt.legend(['Ori', 'Val', 'Pred'])
      plt.show()
```

⚠ /usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
This is separate from the ipykernel package so we can avoid doing imports until



8. Confidence Score of SVM Model and Linear Regression Model

```
lr_confidence = lr.score(x_test,y_test)
print('LINEAR REGRESSION MODEL CONFIDENCE',lr_confidence)
```

```
↳ LINEAR REGRESSION MODEL CONFIDENCE 0.9351180747893716
```

```
[40] svm_confidence = svr_rbf.score(x_test,y_test)
print('SVM MODEL CONFIDENCE',svm_confidence)
```

```
↳ SVM MODEL CONFIDENCE 0.9369403664809587
```

CONCLUSION

In the project, we proposed the use of the data collected from different global financial markets with machine learning algorithms in order to predict the stock index movements. SVM algorithm works on the large dataset value which is collected from different global financial markets. Also, SVM does not give a problem of over fitting. Various machine learning based models are proposed for predicting the daily trend of Market stocks. Numerical results suggest the high efficiency. The practical trading models built upon our well-trained predictor. The model generates higher profit compared to the selected benchmarks.

REFERENCE

1. <http://www.nasdaq.com/>
2. <https://in.finance.yahoo.com>
3. <https://www.google.com/finance>
4. <https://www.princeton.edu/~erp/ERParchives/archivepdfs/R7.pdf>
5. <http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningInStockPriceTrendForecasting.pdf>
6. <http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf>
7. http://cs229.stanford.edu/proj2015/009_report.pdf

