“Stock Price Prediction using machine learning techniques”

B.Tech.project report,
submitted in fulfillment of the requirement for the award of
Integrated Post Graduation degree in Information Technology

by

AMIT KUMAR (2017IMT-009)

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR-474 015

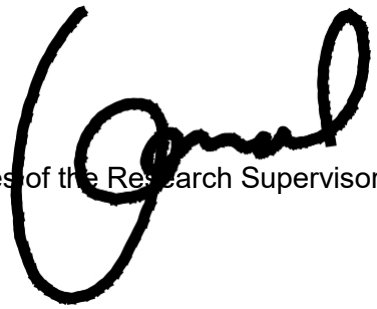
2020

CANDIDATES DECLARATION

I hereby certify that this work that has been presented in this report has been entitled Stock Price Prediction using machine learning techniques Content Degree of Integrated Post Graduation and submitted to the Indian Institute of information technology for an authentic record. Is is carried out by my own in duration October 2020 under the supervision of Dr. Vinal Patel. I also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date:30/10/2020 Signatures of the Research Supervisors



ABSTRACT

Computers for generations are used for detecting patterns.When we say patterns there are lots of areas in the real world that involve patterns.One of the areas is Trading and Stock prediction. The fluctuations and patterns in the stock market has always been a tedious task and involves lots of rigorous e orts to gain some conclusions about the market.As we know that Machine learning and Deep Learning techniques can be used to find patterns/high level features and predict some fruitful results from the given input [1].So we were looking for a problem statement which can have an impact on the people and which contains some patterns so we came up with the idea of stock prediction.In india less than 1% of the total mass of the country in the stock market with the majority of only two states:Gujarat and Maharashtra.In India majority of the people invest in the stock market without adequate knowledge and what they heard [2]. Most of the people start investing when the market is at its peak without any correct knowledge.To give Indian community adequate knowledge and information about the stock market and stock behavior of di erent companies.As stock market are highly inconsistent so we using top-down and bottom up approach to do the fundamental analysis of the market capitalisation to accumulate the stock market index.Then we will use stock split technique and reverse split technique to blue chip the stocks.We will further see the importance of dividend in the stock market.Stock trading is generally averaging the stop less and the do the technical analysis on the features and the options [3].The stock trading is basically long and short positions in the market.The day trading is circuit breakers and then this analysis is selling freeze and buying freeze.Target market can be taken into account while analysis company's profile for list research and creation.Direct promotion campaigns for the end-user survey.The trips to market is necessary for agent selection and training. The area manager development uses this information.Stock price prediction can be used in various sector like energy sector for various PSU's in the country so that they can have better BSE index.It can be used in the utilities and consumer staples. In this we have five phase early contraction for consumer staples then late contraction for financial services the early expansion for technology and transportation.Then we have middle expansion for capital goods and basic material after that we have late expansion for energy and healthcare [4].

Keywords: Machine Learning ,Deep Learning , High level features , Pattern Recogni-tion,Prediction

ACKNOWLEDGEMENTS

I am highly indebted to Dr. Vinal Patel for giving us such a opportunity to work under his guidance.I would like to show gratitude toward the helping nature of the our mentor in case of whenever we were stuck at some problem our mentor were there to help us and guide us what could be done for best results.

Finally, I am grateful to our Institution and batch-mates whose constant encouragement served to renew our spirit and helped us in carrying out this work.

(Amit kumar)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Prev Close	Last day close point
VWAP	volume-weighted average price
WMA	Weighted Moving Average
LR	Linear regression
NN	Neural Network
RNN	Recurrent neural network
ARIMA	Auto-regressive Integrated Moving Average
RMSE	Root Mean square Error

CHAPTER 1

Introduction

1.1 Background

This chapter talks about the introduction and literature Survey, of our thesis. In this section we will talk about the project to predict future prices. We will use simple linear regression and lstm with sliding window technique and fourier transfrom with arima model.

> In the Us 39 to 40 percent of adults invest in the stock market. In india less than 1% of the total mass of the country in the stock market with the majority of only two states:Gujarat and Maharashtra.In India majority of the people invest in the stock market without adequate knowledge and what they heard. Most of the people start investing when the market is at its peak without any correct knowledge [5].

1.2 Motivation

We are taking a step to contribute in filling this gap. To give Indian community ade-quate knowledge and information about the stock market and stock behavior of di erent companies. So that Indians realise that investing is not betting [6].

1.3 Literature survey

> In this paper, by Vatsal H. Shah discusses the Machine Learning techniques which have been applied for stock trading to predict the rise and fall of stock prices before the actual event of an increase or decrease in the stock price occurs. In particular the paper discusses the application of Linear Regression, Prediction using Decision Stumps, Expert Weighting and Online Learning in detail along with the benefits and pitfalls of each method [7].

1 S Abdulsalam Sulaiman Olaniyi, 2 Adewole, Kayode S. , 3 Jimoh, R. G has discussed about the data extracted from the daily activity summaries (equities). It was published by Nigerian Stock Exchange was used in building a database. The data was used in regression analysis using time series data that involved moving average method to predict future stock market prices [8].

Mohammad Almasarweh1 S. AL Wadi2 has mentioned Banking time series data and financial time series is very tedious tast to decompose and forecast be-cause the data are non-stationary and non-linear [9].

Hiransha Ma , Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.P* has talks about the datasets,from which we extract only the day-wise clos-ing price of each stock because stock closing price is preferred since investors make decision on buying which stock based on the closing price of the market [10].

1.4 InteliCharts Predictive Stock Market Analytics

It is a powerful and reliable stock market analysis tool that delivers a rich and reliable stock or time series analysis, which are unique for each stock market analysis tool used for sophisticated investors and geometrical time series analysis. Intelicharts is a platform for investors and reliable stock data analysis [11].

CHAPTER 2

Objective, Methodology

2.1 Objective

This is an attempt to predict Stock prices based on Stock prices of previous days. The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place.

This is a time series analysis and we will see simple eight ways to predict the Stock prices. The various models to be used are: Average, Weighted Average, Moving Av-erage, Moving Weighted Average, Linear Regression, Weighted Linear Regression, Lasso Regression, Moving Window Neural Network, Classical model for time series predictions.

To develop a model which is as close as real world stock movement.

To develop a low error rate model for stock price movement.

To develop an approach that deal with anomalies and outliers.So that model per-forms well on the training data.

Applying techniques like cross validation score to make the model more robust. To develop an Api and interface for the user.

The data we use for prediction would be for closing price of Infosys in NSE for the business days in 2015. So we will import only the Date column and Closing price column.

2.2 Methodology

In this study, we are planning to use di erent machine learning techniques like RNN: Deep Neural Network with Gradual Input, Fine-Tuned IndRNN,Recurrent Neural Network ,Arima,Fourier Transform to make the study more comprehensive .

Python

This program is written in python, one of the widely used languages in Machine Learning.

Used Python

Libraries Numpy

Numpy is a python package used for computation.It is used to process multi-dimensional arrays. It is an high performance python library used for integrat-ing codes form c/c++. It contains multidimensional container which is used for broadcasting functions [12].

Pandas

Pandas is a python library which has Data frames its main data structure. It is used to manipulate data into rows and columns.We can use pandas in-build func-tions to find missing values , filtering,group-by function,filtering data around a condition ,reading from files with CSV, XLSX, TXT, among other formats ,Gen-erating data range, converting frequency, date shifting, lagging, and other time-series functionality,Manipulating high-dimensional data in a data structure with a lower dimension using hierarchical axis indexing,Manipulating data using in-tegrated indexing for Data-Frame objects [13].

Matplotlib

It is a python library used for visualization. It is used for plotting graphs, charts etc. We can use this library to visualize 1D , 2D,Multidimensional data. We can also preform exploratory data analysis and do scientific plotting for publication [14].

Scikit-learn

It is a free python library. It consists of vast variety of methods. It supports many machine learning algorithms like SVM, KNN, Decision Tree, K-Means Clustering, etc. It also supports numerical computations like calculating errors, accuracy, matrix, etc [15].

Keras

It is basically a high level API of Tensorflow that is used for high level and state of the art research [16].

CHAPTER 3

Design Details and Implementation

3.1 DATA FLOW OF THE PROJECT

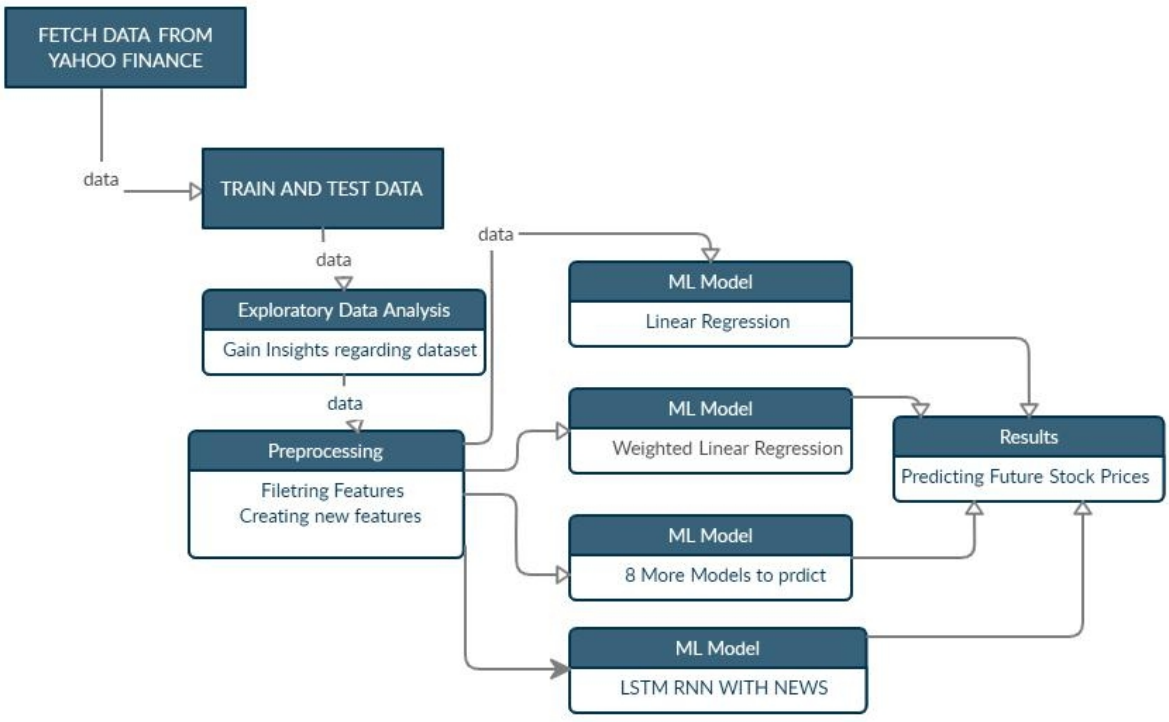


Figure 3.1: Model flow

First step towards the project was to collect the data and we have scraped the data from yahoo finance.

Then the scrapped data is prepossessed and fed to function that extracts technical indicators and features from the data.

After extracting the features Anomaly detection is preformed on the extracted features. patel2015predicting

Then the feature is fed to Fourier transform function for trend analysis.

Now we will give these features to our models and compare result and error.

3.2 Dataset Description

The dataset contains three csv files. Each resembling to INFOSYS, NIFTYITINDEX, and TCS, respectively. One can easily identify that by the name of CSV files.

Timeline of Data recording : 1-1-2015 to 31-12-2015.

Source of Data : Ocial NSE website.

Method : We have used the NSEpy api to fetch the data from NSE site. I have also mentioned my approach in this Kernel - "WebScraper to download data for NSE".

Shape of Dataset: INFOSYS - 248 x 15 || NIFTYI TI NDEX 248x7jj
T CS
248x15

Prev Close: Last day close point

Open: current day open point

High: current day highest point

Low: current day lowest point

Last: the final quoted trading price for a particular stock, or stock-market index, during the most recent day of trading.

Close: Closing point for the current day

VWAP: volume-weighted average price is the ratio of the value traded to total volume traded over a particular time horizon

Volume: the amount of a security that was traded during a given period of time. For every buyer, there is a seller, and each transaction contributes to the count of total volume.

Turnover: Total Turnover of the stock till that day

Trades: Number of buy or Sell of the stock.

Deliverable: Volume the quantity of shares which actually move from one set of people (who had those shares in their demat account before today and are selling today) to another set of people (who have purchased those shares and will get those shares by T+2 days in their demat account).

3.3 Visualize Closing price

Figure 3.2: Closing price

3.4 Classifiers

3.4.1 Average

This is the simplest model.We will use closing price of the dataset and the we will get as average of the previous values. By using this information we will predict it as the forecast. [17]

3.4.2 Weighted average

We shall give more weightage to the data which are close to the last day in training data, while calculating the mean. The last day in the training set will get a weightage of 1(=180/180) and the first day will get a weightage of 1/180 [18].

3.4.3 Moving Average

We have to predict the 68 values in data set and for each values we will get the average of previous 80 days. This will be a simple mean of each column in the y_test[19]:

3.4.4 Weighted Moving Average

We will obtain the stock price on the test date by calculating the weighted mean of past 80 days. The last of the 80 day will have a weightage of 1(=80/80) and the first will have a weightage of 1/80 [20].

3.4.5 Linear regression

A simple linear regression model attempts to explain the relationship between the two variables using a best fitting straight line called a regression line . X which is called the independent variable predicts the value of the dependent variable which is represented as Y linear regression has multiple possible strategies to calculate Direction line .The most popular strategy is least squares and it is as follows you can draw a line and then for each of the data points measure the vertical distance between the point and the line and add these up the regression line will be the one where this sum of distances is as small as possible .Let’s draw a comparative regression line what is your estimation does the red or the blue regression line fit these three data points better let’s draw the residuals and compare turns out that the blue line Stowell distance is smaller and is therefore the better regression line In this method, we will perform a linear regression on our dataset. The values will be predicted as a linear combination of the previous 80 days values [21].

3.4.6 Weighted Linear regression

In this model we will provide weight age to out input data rather than the feature. This technique is bit advance version of linear regression. After passing this feature as input the we will update the weights. After that we will use the output for forecasting.

3.4.7 Lasso regression

Lasso is least absolute shrinkage and selection operator.Lasso regression is Linear Re-gression with L1 regulations. The method to apply regularization(shrinkage) process where it penalizes the coe cient of regression variables shrinking some of then to zero.The goal of this process is to minimize the prediction error [22].

3.4.8 Moving average NN

In this model we will construct a simple Feed forward network. In which we will take 80 features as our input and pass it to the network and then the network will update the

random weights and give us the output which we will use for forecasting.

3.4.9 Auto regression

In this model we will construct a linear combination of the past values as a feature and the feed this information to our Auto regression model.That operates on non convex problems and produce results.Then we will use that result for forecasting [23].

3.4.10 ARMA

In this ARMA model we will use combination of Auto regression and Moving average which is a advance version of both these model. Feeding the input to this model and we will get our output which we will use for forecasting.

3.4.11 ARIMA

ARIMA in general termed as Auto-regressive Integrated Moving Average models. I have planned to use Ar to to develop the dependent relation between the observation and lagged observation.I stands for di erentiating the observation in order to make time series stationary.Ma stands for moving average of the data. Fourier transforms, take a function and create a series of sine waves (with di erent amplitudes and frames) [24].

3.4.12 Recurrent Neural Network

I have used the RNN model referred to as a Recurrent neural network used in the calculation of stock news articles and taking them as a feature to predict stock prices [25].

CHAPTER 4

Testing, Verification and Validation of results

4.1 Mean square error

MSE is a measure of values or the gap (y-y(hat)) between actual values and predicted values.Regression line of how concentrated the MSE represents the data [26].

$$MSE = \frac{1}{n} \sum_{i=1}^{n} e_{i2} \quad (4.1)$$

Where:
f = forecasts (expected values or unknown results), o = observed values (known results).

4.2 Mean Square Error in Models

Table 4.1: Comparison of MSE of various models

Model	MSE
LinearRegression	1754.1645412925632
Weighted Linear Regression	2054.3614078787523
Lasso Regression	1467.333864613377
Moving window Neural Network	6866.732158445809
Auto Regression	1036.9010256538772
Moving Average	2264.330610755017
ARMA	874.2152436482215
ARIMA	9464.89640001397

CHAPTER 5

Results and Discussion

5.1 Mathematical models

5.1.1 Average

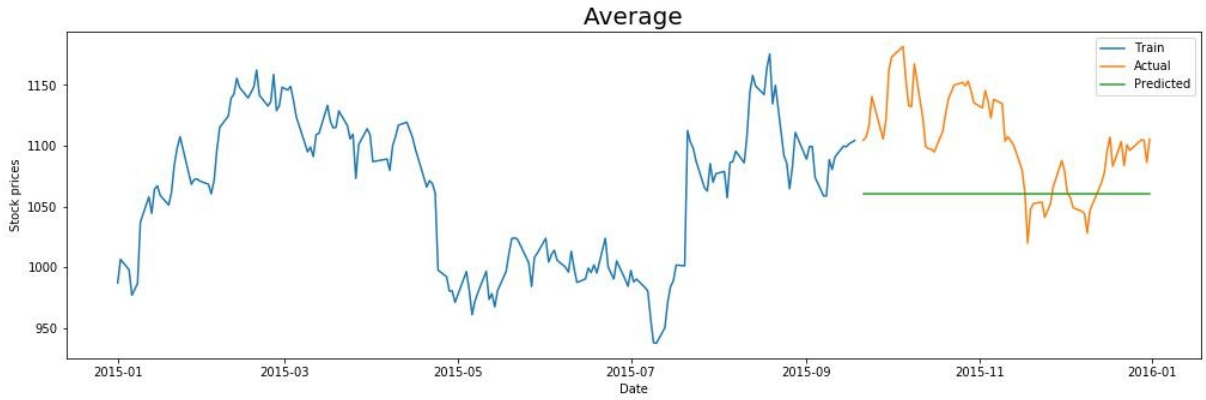


Figure 5.1: Test dataset prediction using average

Figure 5.1 shows stock price forecasting which uses closing price as its feature for pre-dicting the pattern in the closing price.In this mathematical model I have taken the av-erage of the closing price and trained the model.Then predicted on the testing data.But this model losses it's accuracy when the prices becomes slightly non linear.

5.1.2 Weighted average

Figure 5.2 shows stock price forecasting with MSE 2054.36 which uses closing price as its feature for predicting the pattern in the closing price.In this mathematical model I have taken weighted average of the closing price and trained the model.Then predicted on the testing data.

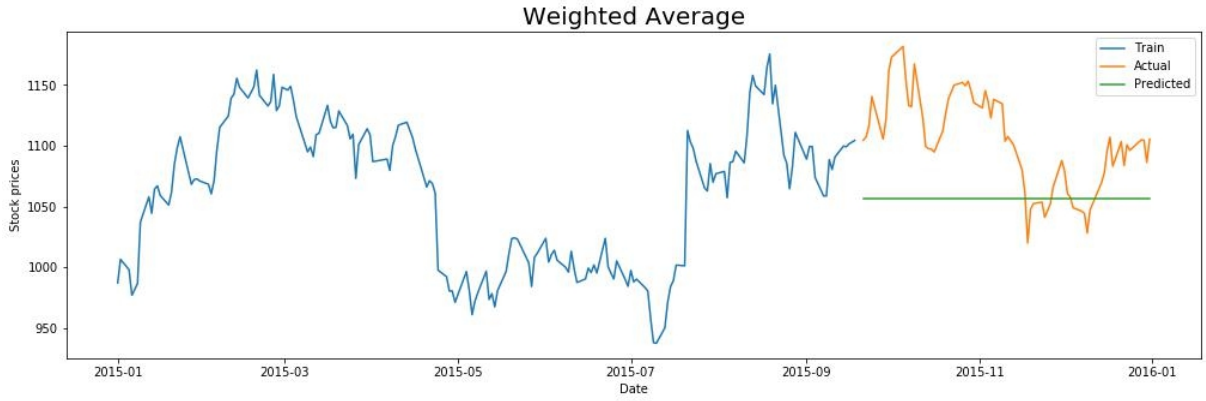


Figure 5.2: Test dataset prediction using weighted average

5.1.3 Moving average

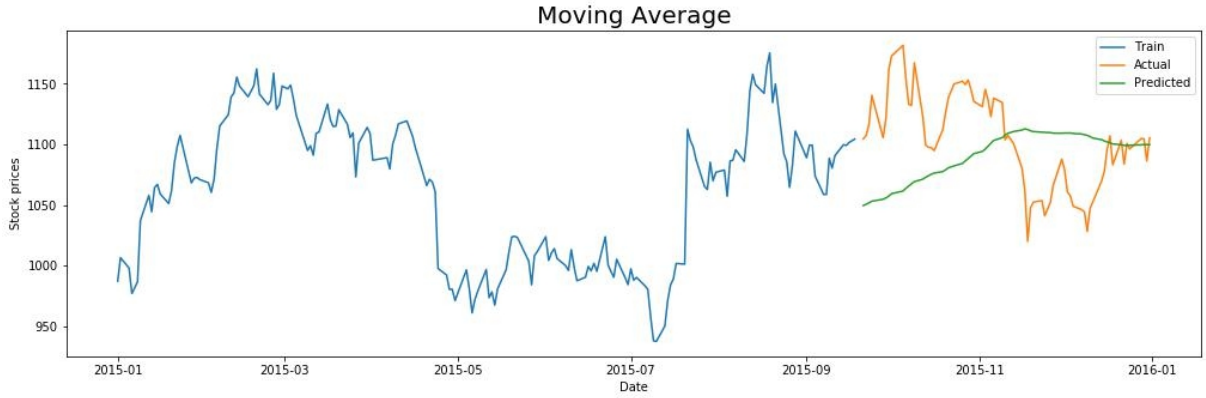


Figure 5.3: Test dataset prediction using moving average

Figure 5.3 shows stock price forecasting with MSE 6866.732 which uses closing price as its feature for predicting the pattern in the closing price.In this mathematical model I have taken Moving average of the closing price and trained the model.Then predicted on the testing data.

5.1.4 Weighted Moving average

Figure 5.4 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price of last n days and try to predict next day stock price.In this mathematical model I have taken weighted moving average of the closing price and trained the model.Then predicted on the testing data.

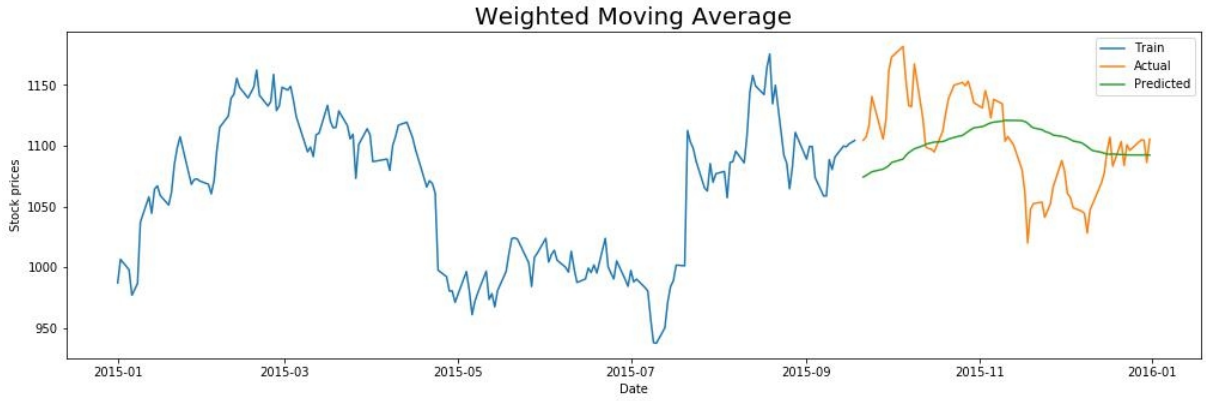


Figure 5.4: Test dataset prediction using weighted moving average

5.2 Linear models

5.2.1 Linear regression

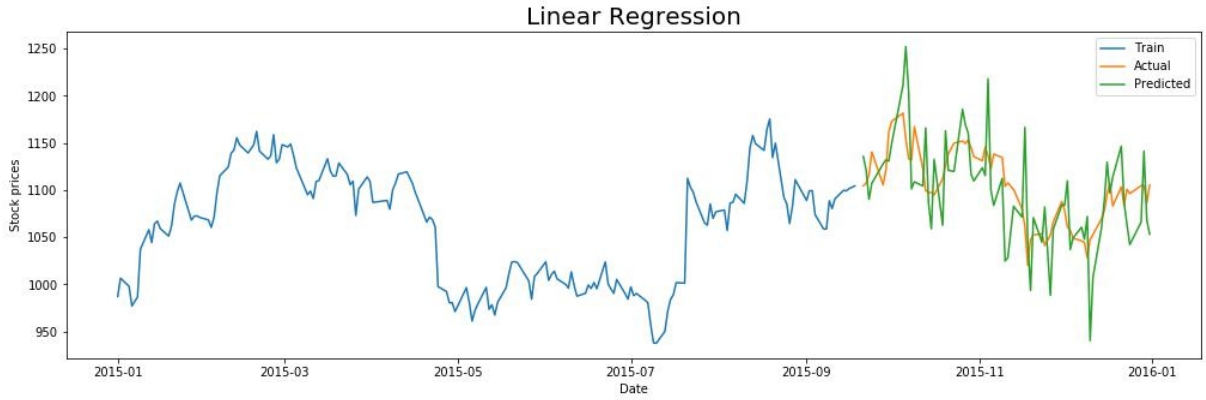


Figure 5.5: Test dataset prediction using linear regression

Figure 5.5 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this linear model I have taken the closing price and trained the model.Then predicted on the testing data with MSE 1754.16.

5.2.2 Lasso regression

Figure 5.6 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this lasso model I have taken the closing price and trained the model.Then predicted on the testing data with MSE 1467.33.

– Lasso regression preformed better than Linear regression in terms of MSE.

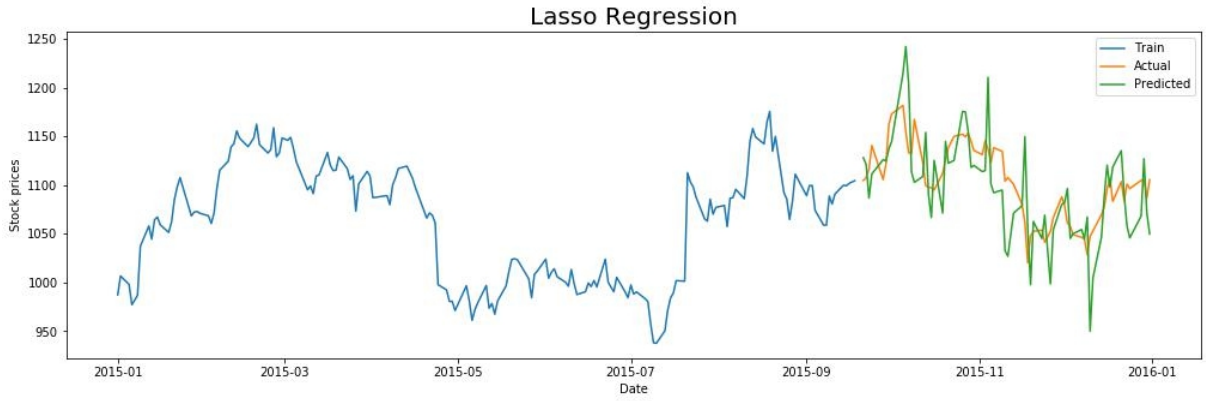


Figure 5.6: Test dataset prediction using Lasso regression

5.2.3 Moving average Neural Network

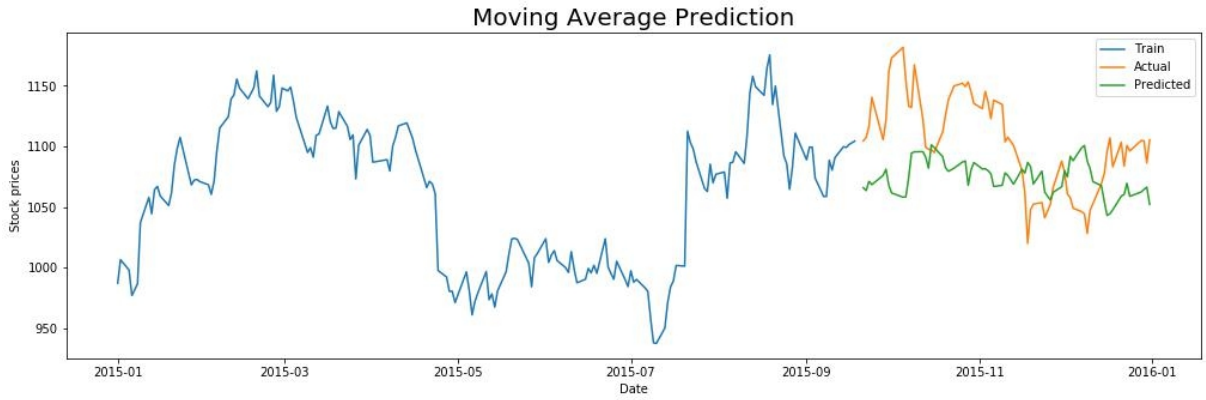


Figure 5.7: Test dataset prediction using moving average NN

Figure 5.7 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this Moving average Neural Network I have taken the closing price and trained the model.Then predicted on the testing data with MSE 6866.73.

5.2.4 Auto regression

Figure 5.8 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this Auto regression model I have taken the closing price and trained the model.Then predicted on the testing data with MSE 1036.90.

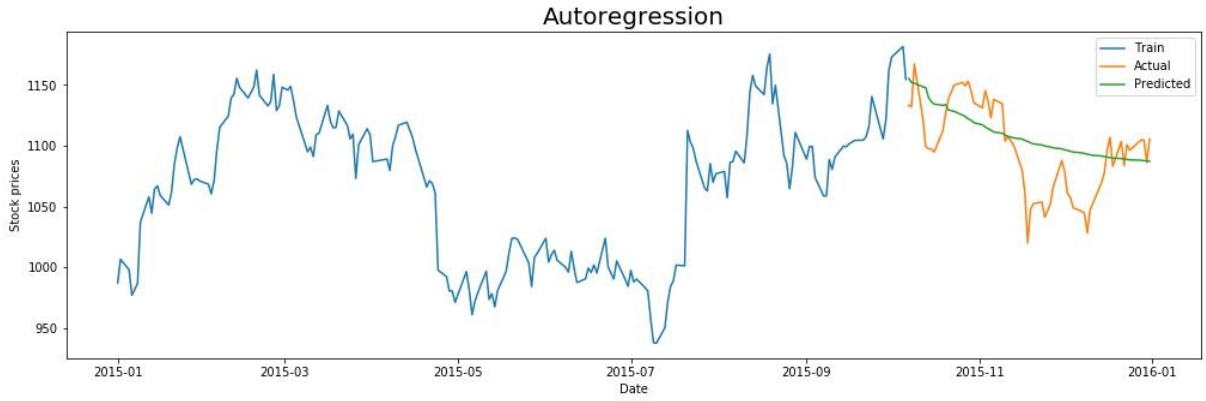


Figure 5.8: Test dataset prediction using Auto regression

5.3 ARMA model,ARIMA model

5.3.1 ARMA

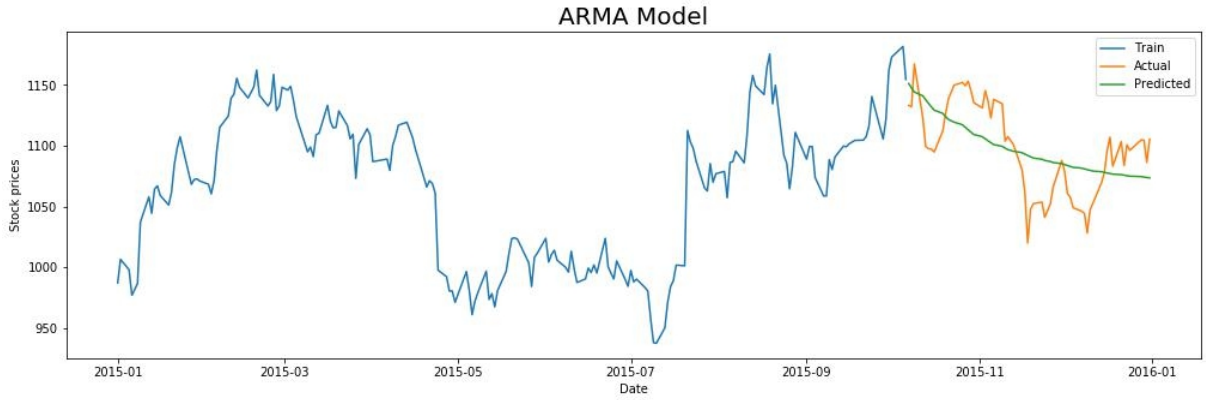


Figure 5.9: Test dataset prediction using ARMA

Figure 5.9 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this ARMA model I have taken the closing price and trained the model.Then predicted on the testing data with MSE 874.21.ARMA model performed slightly better than than the ARIMA model in terms of MSE

5.3.2 ARIMA

Figure 5.10 shows stock price forecasting which uses closing price as its feature for predicting the pattern in the closing price .In this ARIMA model I have taken the closing price and trained the model.Then predicted on the testing data with MSE 9464.89.ARMA model performed slightly better than than the ARIMA model in terms of MSE

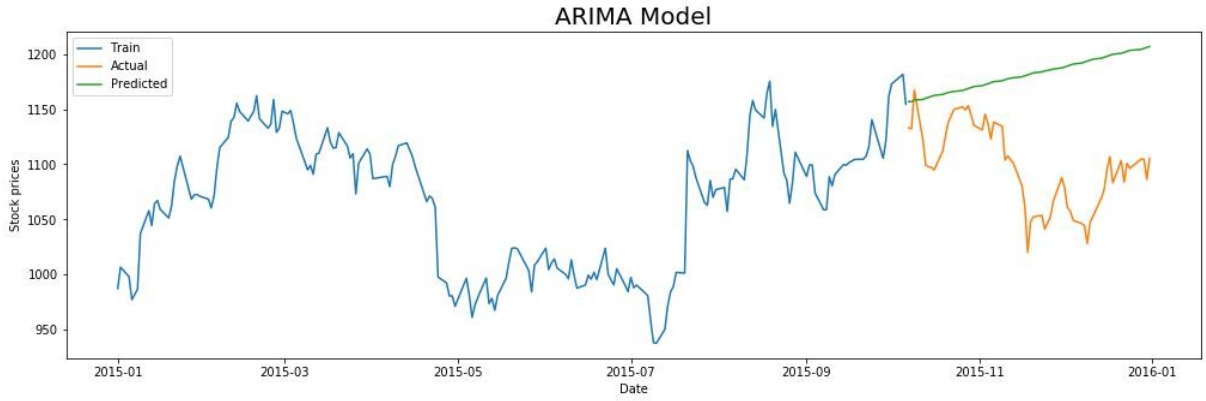


Figure 5.10: Test dataset prediction using ARIMA

5.4 RNN model taking news as feature

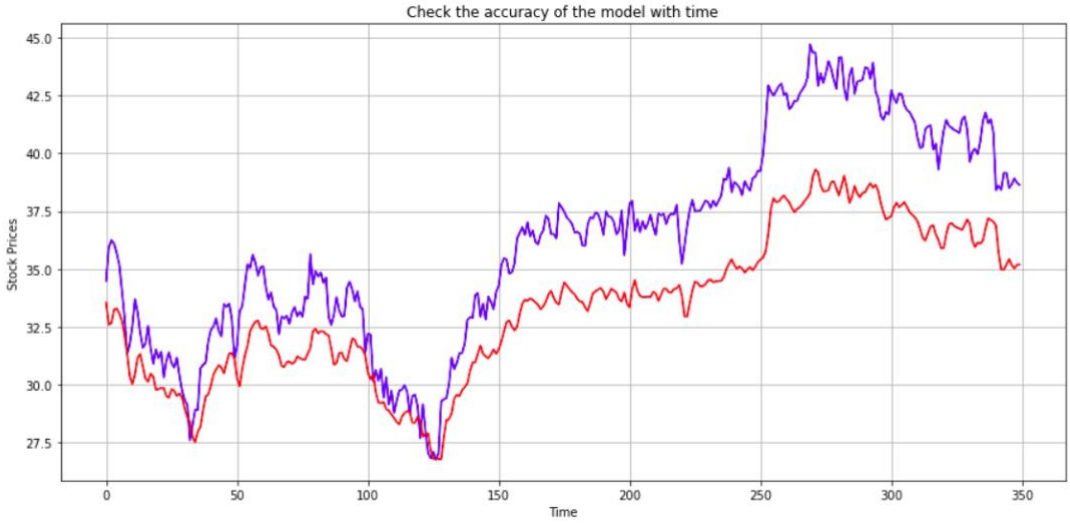


Figure 5.11: Test dataset prediction using Arima(moving average)

Figure 5.11 shows stock price forecasting which uses closing price and news as its feature for predicting the pattern in the closing price . Sentimental analysis is done on the news dataset.Then it is converted into o/1 class.RNN model taking news as a feature performed better than 9 models listed above with a mse of 608.21.

CHAPTER 6

Conclusion

– First Phase:This thesis work proposed a stock price forecasting model which uses closing price as its feature for predicting the pattern in the closing price.I have first started with mathematical models like average, moving average in which they took the closing price of last n days and try to predict next day stock price. But this model losses it’s accuracy when the prices becomes slightly non linear.

– Second Phase:After using the Mathematical models then I moved on to the linear models like Linear regression, Weighted linear regression and Lasso regression. After feeding the input to these models I observed that lasso regression preforms better in terms MSE which was 1467.33 better than all the three models. But the linear model start’s preforming ine ciently when the data was inconsistent.

– Third Phase: After using Linear models then I moved on to the ARMA model,ARIMA model.After training and testing these models and visualis-ing the forecasting of these models.I observed that ARMA model performed slightly better than than the ARIMA model in terms of MSE which was 874.2152.But these models started giving high MSE in the period of infla-tion.So we do not have a good prediction as there is no significant trend in the data.

– Forth Phase: In this phase I took economic news as a feature.In which I have performed sentimental analysis on the news and classified into 3 classes (Positive ,Negative,Neutral).After doing thing then feed thing input to the RNN model.RNN model was able find pattern in the data-set and gave minimum error among all the models used.But still the error rate can be decreased the future.

6.1 Future work

– Future research on the more advanced method or analyzing various indexes of the stock market can improve the model's accuracy, robustness, and per-formance.

– As economic news has been taken as a feature. A more e cient sentimental analysis model can be build to classify the news with less error rate.

– Research on making a more explainable and better understanding of each feature's influence for the prediction can be worked upon.

REFERENCES

[1] K. R. French, G. W. Schwert, and R. F. Stambaugh, “Expected stock returns and volatility,” Journal of financial Economics, vol. 19, no. 1, p. 3, 1987.

[2] D. M. Cutler, J. M. Poterba, and L. H. Summers, “What moves stock prices?” National Bureau of Economic Research, Tech. Rep., 1988.

[3] M. J. Gordon, “Dividends, earnings, and stock prices,” The review of eco-nomics and statistics, pp. 99–105, 1959.

[4] E. F. Fama, “The behavior of stock-market prices,” The journal of Business, vol. 38, no. 1, pp. 34–105, 1965.

[5] ——, “Random walks in stock market prices,” Financial analysts journal, vol. 51, no. 1, pp. 75–80, 1995.

[6] R. T. Baillie and R. P. DeGennaro, “Stock returns and volatility,” Journal of financial and Quantitative Analysis, pp. 203–214, 1990.

[7] V. H. Shah, “Machine learning techniques for stock prediction,” Foundations of Machine Learning| Spring, vol. 1, no. 1, pp. 6–12, 2007.

[8] S. Abdulsalam, K. S. Adewole, and R. Jimoh, “Stock trend prediction using regression analysis–a data mining approach,” 2011.

[9] M. Almasarweh and S. Alwadi, “Arima model in predicting banking stock market data,” Modern Applied Science, vol. 12, no. 11, 2018.

[10] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, and K. Soman, “Nse stock market prediction using deep-learning models,” Procedia computer science, vol. 132, pp. 1351–1362, 2018.

[11] R. J. Barro, “The stock market and investment,” The review of financial studies, vol. 3, no. 1, pp. 115–131, 1990.

[12] I. Idris, NumPy 1. 5 Beginner’s Guide. Packt Publishing Ltd, 2011.

[13] ——, NumPy Beginner's Guide. Packt Publishing Ltd, 2013.

[14] W. L. Megginson and J. M. Netter, "From state to market: A survey of empirical studies on privatization," Journal of economic literature, vol. 39, no. 2, pp. 321–389, 2001.

[15] J. Wurgler, "Financial markets and the allocation of capital," Journal of financial economics, vol. 58, no. 1-2, pp. 187–214, 2000.

[16] M. P. Naeini, H. Taremian, and H. B. Hashemi, "Stock market value prediction using neural networks," in 2010 international conference on computer information systems and industrial management applications (CISIM). IEEE, 2010, pp. 132–136.

[17] K.-j. Kim and W. B. Lee, "Stock market prediction using artificial neural networks with optimal feature transformation," Neural computing & appli-cations, vol. 13, no. 3, pp. 255–260, 2004.

[18] C. M. Jones, "A century of stock market liquidity and trading costs," Avail-able at SSRN 313681, 2002.

[19] R. Gencay, "Non-linear prediction of security returns with moving average rules," Journal of Forecasting, vol. 15, no. 3, pp. 165–174, 1996.

[20] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in 1990 IJCNN international joint conference on neural networks. IEEE, 1990, pp. 1–6.

[21] Y. E. Cakra and B. D. Trisedya, "Stock price prediction using linear re-gression based on sentiment analysis," in 2015 international conference on advanced computer science and information systems (ICACSIS). IEEE, 2015, pp. 147–154.

[22] S. S. Roy, D. Mittal, A. Basu, and A. Abraham, "Stock market forecasting using lasso linear regression model," in Afro-European Conference for Industrial Advancement. Springer, 2015, pp. 371–381.

[23] R. J. Shiller and A. E. Beltratti, "Stock prices and bond yields: Can their comovements be explained in terms of present value models?" Journal of monetary economics, vol. 30, no. 1, pp. 25–46, 1992.

[24] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the arima model," in 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014, pp. 106–112.

[25] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, “Stock price prediction using lstm, rnn and cnn-sliding window model,” in 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 2017, pp. 1643–1647.

[26] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance,” Climate research, vol. 30, no. 1, pp. 79–82, 2005.