STOCK MARKET PREDICTION USING MACHINE LEARNING TECHNIQUES

Rutuja Rajebhau Mehetre[1], Arati Sharad Patil[2], Jagruti Kale[3], Pratiksha Barve[4], Prof. Ketaki Bhoyar[5]

[1-4]Department of Computer Engineering, Dr. D. Y. Patil Institute of Engineering , Management & Research , Pune 411044, India
[5]Asst. Professor, Department of Computer Engineering, Dr. D. Y. Patil Institute of Engineering, Management & Research, Pune 411044, India

---

ABSTRACT: The objective of the prediction of stock market behaviour is to determine the future value of a company asset or other financial instrument marketed on a financial exchange. The successful prediction of a stock's future value will maximize investor's profits. This paper proposes a machine learning model to predict stock market price. Forecasting accuracy is the most important factor in selecting any forecasting methods. The appropriate stock selections those are suitable for investment is a very difficult task. The key factor for each investor is to gain maximum profits on their investments. By using the dataset of stock market we are going to use preprocessing, processing and regression analysis. We will review the use of machine learning and deep learning algorithm on dataset and the result it generates. Prediction of how the stock market will perform is one of the most difficult things to do.

KEY WORD:

Machine Learning, Deep Learning, Time Series, Regression Analysis, Data Analysis.

INTRODUCTION:

The stock market refers to collection of markets and exchanges where regular activities of buying, selling, and yield of shares of public companies take place. While today it is possible to purchase almost everything online, there is usually a specified market for every commodity. A stock market is a similar specified market for exchanging various kinds of securities in a controlled, secure and managed the environment. The stock market brings together hundreds of thousands of market participants who wish to buy and sell shares. It ensures fair pricing practices and transparency in transactions. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument marketed on an exchange. The successful prediction of a stock's future price could yield remarkable profit. In Fundamental analysis, Stock Market price fluctuations are believed to derive from a security's relative data. Fundamentalists use numeric information such as earnings, ratios, and management effectiveness to determine future forecasts. In Technical analysis, it is believed that market timing is key. Technicians utilize charts and modeling techniques to identify trends in price and volume. These later individuals depend on historical data in order to predict future outcomes. Stock Market prediction has always had a certain need for researchers.

PROBLEM STATEMENT:

Investors 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 get more profits the input to our system will be historical data from Google. Appropriate data would be applied to find the stock prices of markets. Hence the prediction model will notify the up or down of the stock price movement for the next trading day and investors can act upon it so as to maximize their chances of gaining a profit. The entire system would be implemented in python. Hence it will effectively be a zero cost system.

LITERATURE SURVEY:

1. The model was comprised of four sub-models that were all based on different machine learning technique. Each sub-model used 6 input attributes including fuel price, commodity, foreign exchange, interest rate, general public sentiment and related NEWS. The historical data of the market was also used for predicting the market performance

using statistical techniques like Auto-Regressive Integrated Moving Average (ARIMA) and Simple Moving Average (SMA). Support Vector Machine, Radial Basis Function (RBF), Artificial Neural Network has two variants including Single Layer Perceptron and Multi-layer Perceptron were used to design four different sub-models.

2. The first is Efficient Market Hypothesis (EMH). In EMH, it is assumed that the price of a security considers all of the information available and that everyone has some level of access to the information. A different point of view on prediction comes from Random Walk Theory. In this theory, the forecast of Stock Market is believed to be impossible where prices are determined randomly and outperforming the market is infeasible. In one model that tested trading philosophies; LeBaron et. al. predicated that much can be learned from a simulated stock market with simulated traders. In their work, simulated traders mimicked human trading activity. Within this interval of time, Gidofalvi demonstrated that there exists a weak potential to predict the direction of a security before the market corrects itself.

EXISTING SYSTEM:

The existing system use data mining techniques. Data mining techniques are less accurate and time consuming to analyze big data. The system does not allow the import of raw data directly. The existing system cannot be used to analyze multi-variate time series. Lastly, the system does not have a user-interface which can be distributed as a GUI app to users for personal use Stock Market Prediction. The data frame features were date and the closing price for a particular day. We used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and the actual values.

PROPOSED SYSTEM:

In this proposed system we focus on predicting stock values using machine learning algorithms like SVM (Support Vector Machine) and LSTM (Long short-term memory). This system will use multiple linear regression for visualization of data on a particular dataset. We majorly use machine learning libraries like numpy which was used to clean and manipulate the data, and getting it into a form ready for analysis and pandas library for data processing which combined different datasets into a data frame.

Here, we are predicting future values by the use of past data of last five years. These data is used for training module and by applying this training module on month data for comparison to check accuracy for analyzing fluctuation of data between these time periods for forecasting in the format of seven days, fifteen days and one month.

Methodology:

Support Vector Machine

Support Vector Machine are a set of supervised learning methods which learn from the dataset and used for classification. SVM is a classifier derived from statistical learning. An SVM is a kind of large-margin classifier it is vector space based machine learning method where the goal is to find a decision boundary between two classes that is maximally far from any point in the training data. An SVM algorithm builds a model that predicts whether a new ex. falls into one class or the other.

Support Vector Regression

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. First of all, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem. But besides this fact, there is also a more complicated reason, the algorithm is more complicated therefore to be taken in consideration. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated. The method of Support Vector Machines can be extended to solve regression problems this is called as support vector regression. The model produced by Support vector regression also depends on a subset of training data as the cost function for building the model ignores any data close to the model prediction.

LSTM

LSTM stands for Long-Short Term Memory. LSTMs are a type of RNN with gates inside of each LSTM cell. LSTM cells as a cell with its own tiny neural network inside of each one. These gates inside LSTM cells help the LSTM decide what data is important to be remembered and what data can be forgotten even on long series of data. Types of gates are forget gate, input gate and the output gate. In LSTM the previous cell output is passed forward as input to the next cell. Gates contain sigmoid activation functions. It takes in numerical input and squishes the numbers into a range of 0 to 1. LSTM allows for accurate predictions on long series of sequential data.

MATHEMATICAL MODEL

Let 'S' be the system

Where,

S= {I, O, P, Fs, Ss}

Where,

I = Set of input

O = Set of output

P = Set of technical processes

Fs = Set of Failure state

Ss = Set of Success state

Identify the input data I1, I2, In

I = {(Stock Data)}

Identify the output applications as

O = {(Stock market prediction)}

Identify the Process as P

P = {(Data pre-processing, Data Processing, Regression analysis, prediction)}

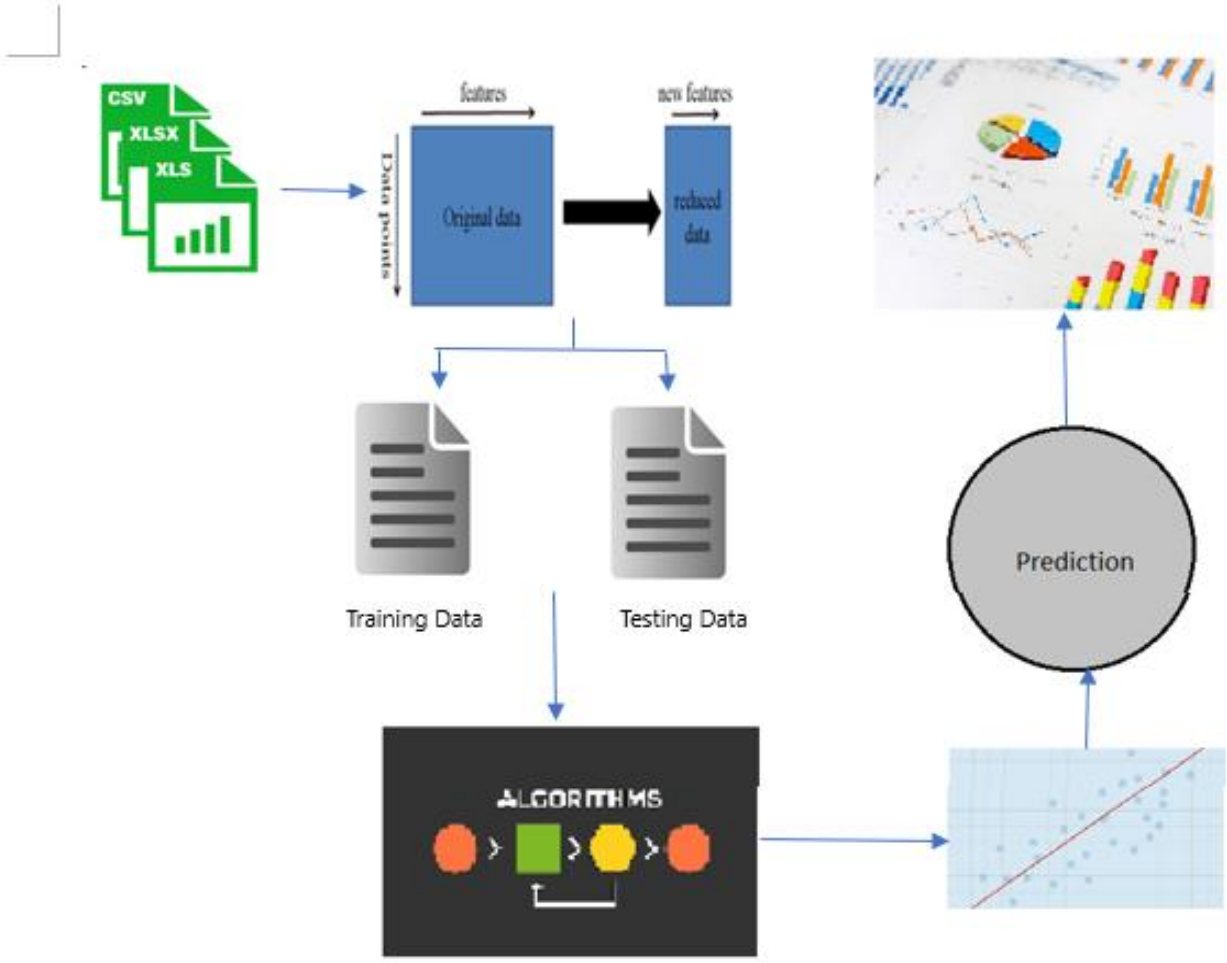
Identify the Failure state as Fs

Fs = {(If not predicted, if more time required to predict}

Identify the Success state as Ss

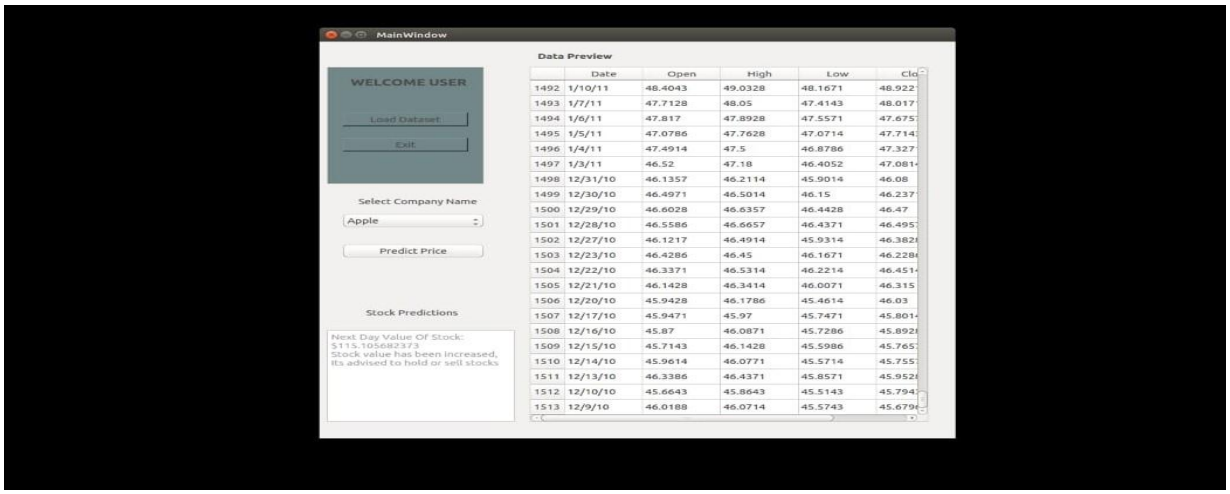
P = {(Correctprediction)}

SYSTEM ARCHITECTURE:

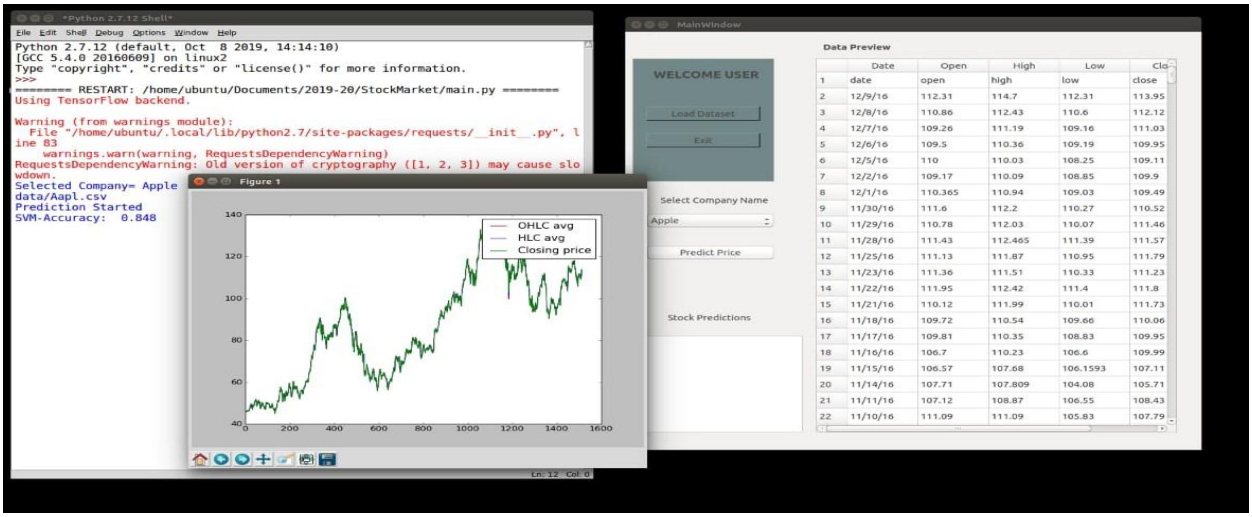


EXPERIMENTAL RESULT:

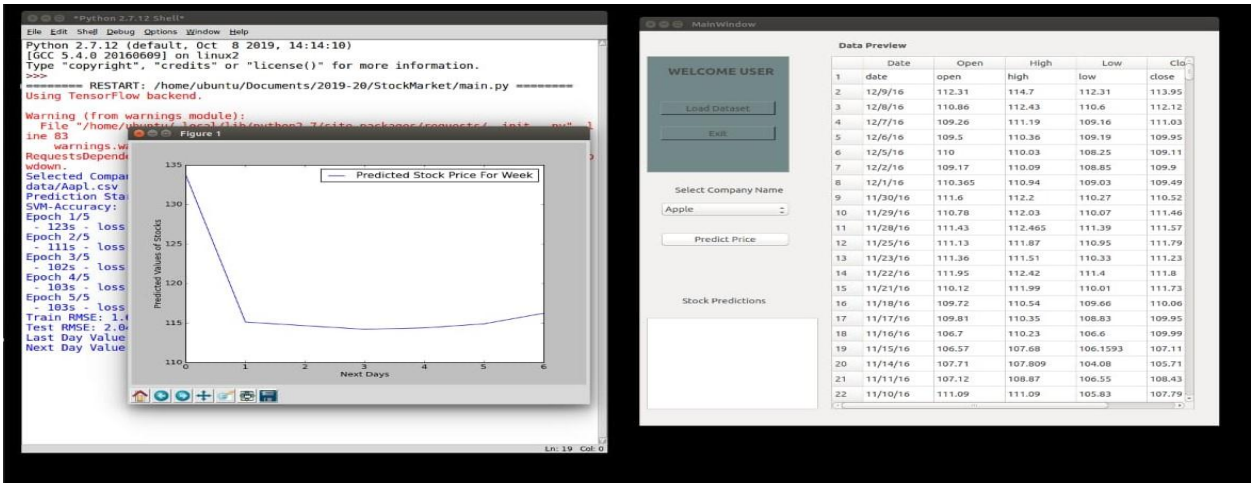
This GUI represents processed data which contains OPEN (the value of the stock at the very beginning of the trading day), HIGH(which describes the highest value the stock had in previous year), LOW (which describes the lowest value the stock had in previous year), CLOSE(the price at which the stock is valued before the trading day closes.) price values of stock market and user interface for loading datasets of various companies like apple, yahoo etc. for predicting it's future values in different visualization format.



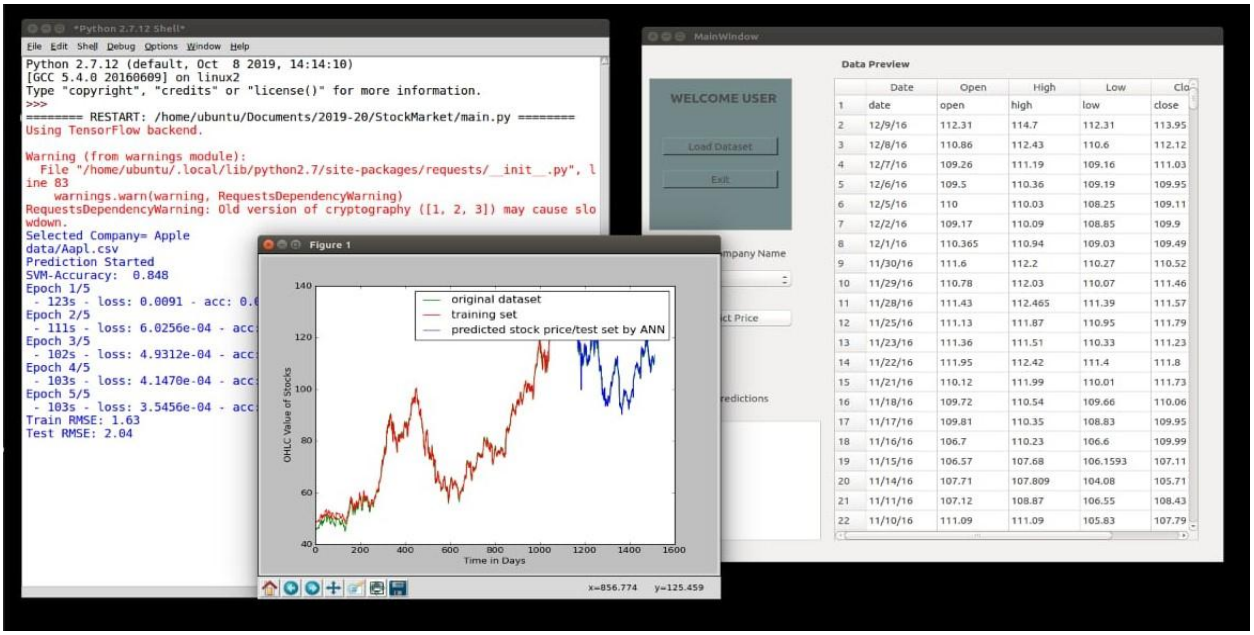
This is a time series plot generated from using the “matplotlib.pyplot” library



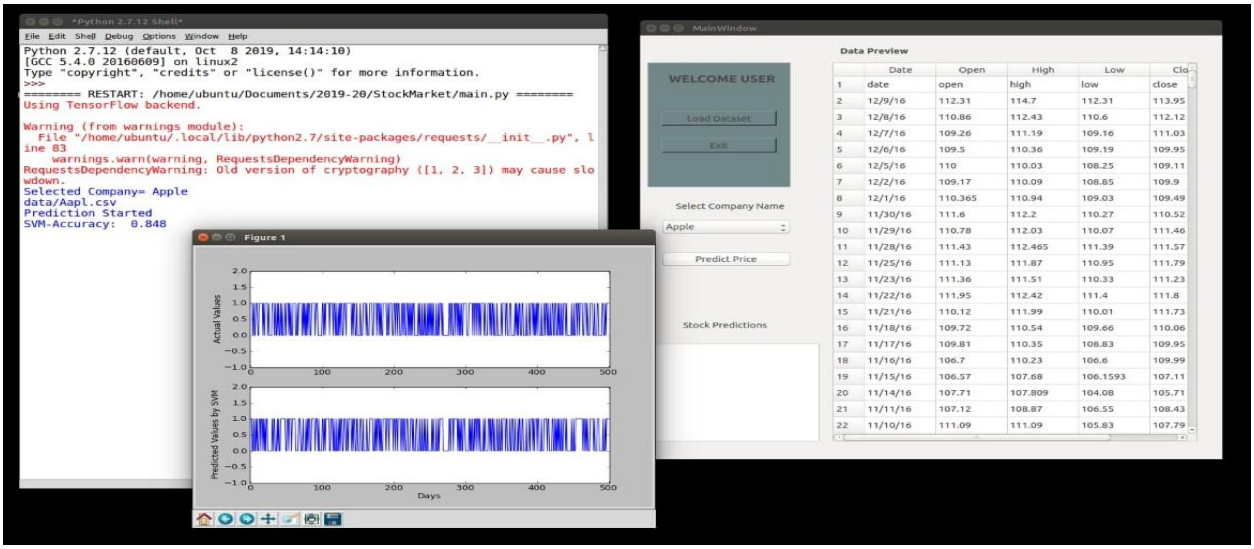
Predicting stock values for “APPLE” company for first week.



Train data for last five years used for testing purpose by using ANN(Artificial Neural Network) with the help of LSTM(Long – short term memory) algorithm.



Support vector machine(SVM) algorithm is used for checking fluctuation between actual data and predicted value. This type of analysis is useful for checking accuracy of prediction.



FUTURE SCOPE:

Financial analysts, investors can use this prediction model to take trading decision by observing market behavior. More work on refining key phrases extraction will definitely produce better results. Enhancements in the preprocessor unit of this system will help in improving more accurate predictability in stock market. Future research includes using other machine learning techniques such as Relevance Vector Regression, which promises to have better accuracy and fewer vectors in classification. Another worthwhile approach would be to 27 test a model based on article terms and percentage of stock price change. While our models relied on fixed stock prices that traded within a consistent range, penny stocks with wild fluctuations may prove worthy of further research.

CONCLUSION:

Thus, as we can see above in our proposed method, we train the data using existing stock dataset that is available. We use this data to predict and forecast the stock price of n-days into the future. The average performance of the model decreases with increase in number of days, due to unpredictable changes in trend. The current system can update its training set as each day passes so as to detect newer trends and behave like an online-learning system that predicts stock in real-time.

REFERENCES:

[1] ASM Shakil Haider, "Forecasting Dhaka Stock Exchange (DSE) return: An Autoregressive Integrated Moving Average (ARIMA) approach", 2009, North South Business Review, Volume 3, Number 1, ISSN 1991- 4938.

[2] Tarun Kanti Bose, Md. Reaz Uddin, and Md. Wahidul Islam, "Measuring and Comparing the Efficiency of Dhaka Stock Exchange and Chittagong Stock Exchange", 2014, International Journal of Scientific and Research Publications, Volume 4, Issue 3,1 ISSN 2250-3153

[3] Md. Kamruzzaman, Md. Mohsan Khudri, and Md. Matiar Rahman, "Modeling and predicting stock market returns: A case study on Dhaka stock exchange of Bangladesh", 2017, Dhaka Univ. J. Sci. 65(2): 97-101

[4] Ayodele A. Adebiyi, Aderemi O. Adewumi, and Charles K. Ayo, "Stock Price Prediction Using the ARIMA Model", 2014, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.

[5] Ashish Sharma, Dashish Sharma, Dinesh Bhuriya and Upendra Singh, "Survey of Stock Market Prediction Using Machine Learning Approach", 2017, International Conference on Electronics,Communication 2017,doi:10.1109/iceca.2017.8212715

[6] Poonam Somani, Shreyas Talele, and Suraj Sawant "Stock Market Prediction Using Hidden Markov Model", 2014, IEEE 7th Joint International Information Technology and Artificial Intelligence Conference. doi:10.1109/itaic.2014.7065011

[7] Md. Rafiul Hassan and Baikunth Nath, "Stock Market Forecasting Using Hidden Markov Model: A New Approach",2005, 5th International Conference on Intelligent Systems Design and Applications (ISDA'05)".

[8] Kumar abhishek, Anshul Khairwa, Tej Pratap, and Surya Prakash "A stock market prediction model using Artificial Neural Network", 2012, Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12). doi:10.1109/icccnt.2012.6396089