BMS COLLEGE OF ENGINEERING BENGALURU
Autonomous Institute, Affiliated to VTU

A Technical Seminar Report based on review of Research Publication/Patent

"Stock Market Prediction using Machine Learning"

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering
in
Computer Science and Engineering

Submitted by:
ABHISHEK R
1BM19CS400

Work carried out at

Internal Guide

Sheetal V A
Assistant Professor
BMS College of Engineering

Department of Computer Science and Engineering
BMS College of Engineering
Bull Temple Road, Basavanagudi, Bangalore 560 019
2021-2022

BMS COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I,ABHISHEK R (1BM19CS400) student of 7th Semester, B.E, Department of Computer Science and Engineering, BMS College of Engineering, Bangalore, hereby declare that, this technical seminar entitled "STOCK MARKET PREDICTION USING MACHINE LEARNING" has been carried out under the guidance of **SHEETAL V A,** Assistant Professor, Department of CSE, BMS College of Engineering, Bangalore during the academic semester September - January 2022. I also declare that to the best of our knowledge and belief, the technical seminar report is not from part of any other report by any other students.

Signature of the Candidate

ABHISHEK R (1BM19CS400)

BMS COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Technical Seminar titled "**STOCK MARKET PREDICTION USING MACHINE LEARNING**" has been carried out by ABHISHEK R (1BM19CS400) during the academic year 2021-2022.

Signature of the guide
SHEETAL V A
Assistant Professor,
Department of Computer Science and Engineering
BMS College of Engineering, Bangalore

Abstract

In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume.

Chapter 1: Introduction

1.1 Overview

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements.

The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data.

Regression and LSTM models are engaged for this conjecture separately. Regression involves minimizing error and LSTM contributes to remembering the data and results for the long run. Finally, the graphs for the fluctuation of prices with the dates (in case of Regression based model) and between actual and predicted price (for the LSTM based model) are plotted.

1.2 Motivation

Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision.

However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data.

1.3 Objective

The objective of this is to compare the methods implemented in the past for the prediction of stock market value. The comparison between the results of the previous work help in determining the efficient algorithm and building the system model based on that selected algorithm.

Prediction of the stock market value has always been an area of interest for investors and researchers for a long time due to its complexity, intrinsic volatility and regularly changing nature. Thus, making reliable predictions are challenging. Both stock prices and twitter data have been employed in the prediction process. Previously various machine learning algorithms have been implemented to predict stock market variations.

Machine learning is defined as a technique of analyzing the data that automatizes the systematic model establishment. ML is a branch of artificial intelligence that works on the idea that systems can learn from data, patterns and make decisions. This method of learning requires minimum human intervention. Machine learning is described using 3 parameters that is P, E, T where T is the task learned, E is the experience through which T is learned and performance P varies with E.

Deep learning is a subset of machine learning in artificial intelligence (AI) that uses multiple layers to progressively extract higher-level features from raw input. In deep learning, each learning level transforms the input data given into more abstract and composite representation. Learning can be supervised, semi supervised or unsupervised. Deep learning has been widely used in financial areas such as stock market prediction, financial information processing, and trade execution strategies.

Chapter 2: LITREATURE SURVEY

[1] Survey on stock market prediction using machine learning techniques The objective is to predict the market performance with the help of an artificial neural network. The techniques of artificial neural networks classify the stock in mainly three categories that is buy, hold and sell, supported previous data. It's observed that the logistic regression model is employed by every individual to predict a stock in an exceedingly better way.

[2] Survey on stock market prediction using machine learning algorithms. The objective is to predict the stock movement with good accuracy. The realm of interest is that the dataset of the stock prices from past years. The raw dataset must be preprocessed for data analysis. After pre-processing the information, we are going to use machine learning techniques like random forest and support vector machines on the dataset to come up with the outcomes.

[3] Survey on predicting stock prices using LSTM Forecasting the stock prices has always been a difficult task for the analysts. The investors are highly curious about the stock prediction. For a successful investment, many investors have an interest in knowing the long-run condition of the stock market. The prediction models for the stock market helps the investors and also the analysts by providing the long run information of the stock market. Recurrent neural networks (RNN) and Long Short-Term Memory (LSTM) is the machine learning approaches used for stock price prediction.

[4] Survey on stock trend prediction using regression Analysis A data mining approach various organizations are collecting data, building large data warehouses to store the collected data. Discovering the knowledge out of the collected data is done by employing a technique called data mining. Data mining software tool is employed to extract values of the variables from the dataset to predict the long-run values of other variables with the utilization of time series data.

[5] Survey on stock market prediction using machine learning developing an application for analyzing and predicting stock market prices increases the investor's interest in stock markets. First, we have to analyze the present and emerging methods of stock price prediction. The different approaches are fundamental analysis, technical analysis, and also the application of machine learning. Fundamental analysis and machine learning are accustomed to guide an investor's decisions. Whereas the technical analysis methodology provides limited useful information.

[6] Survey on stock market prediction: using historical data analysis the stock market plays a vital role in business and finance. Fundamental analysis is depleted using social media data with the help of Sentimental Analysis. Social media data includes a high impact because of its increased usage, and it helps predict the trend of the stock market. Technical analysis is depleted by applying machine learning algorithms on historical data of stock prices. The tactic usually involves gathering various social media data, news to extract sentiments expressed by individuals.

Chapter 3: METHODOLOGY/TECHNIQUES OR ALGORITHM USED

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions.

The dataset being utilized for analysis was picked up from Yahoo Finance. The dataset consisted of approximately 9 lakh records of the required stock prices and other relevant values. The data reflected the stock prices at certain time intervals for each day of the year. It consisted of various sections namely date, symbol, open, close, low, high and volume. For the purpose of simulation and analysis, the data for only one company was considered. All the data was available in a file of csv format which was first read and transformed into a data-frame using the Pandas library in Python. From this, the data for one particular company was extracted by segregating data on the basis of the symbol field. Following this normalization of the data was performed through usage of the sklearn library in Python and the data was divided into training and testing sets. The test set was kept as 20% of the available dataset.

Although machine learning as such has many models but this paper focuses on two of the most important amongst them and made the predictions using these.
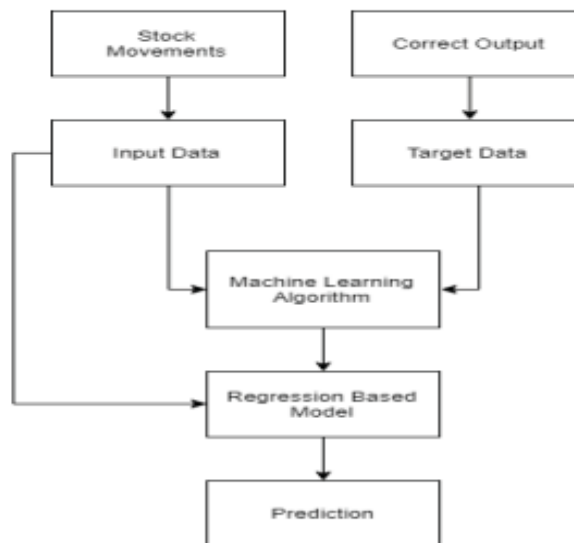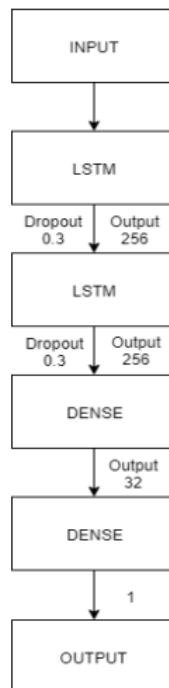
A. Regression Based Model

Fig. 1 Flow Chart for Regression Based Model

The paper utilizes the gradient descent linear regression algorithm for predicting correct values by minimizing the error function as given in Fig.1. Linear Regression [6]as governed by the above equation is performed on the data and then the relevant predictions are made.

B. Long Short Term Memory (LSTM) Network Based Model

The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long term history of the market [6]. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate [7]. Thus proving itself as much more reliable compared to other methods. Since stock market involves processing of huge data, the gradients with respect to the weight matrix may become very small and may degrade the learning rate. This corresponds to the problem of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them.

Long Short Term Memory (LSTM) Network Based Model

```
            ┌─────────────┐
            │    INPUT    │
            └─────────────┘
                   │
                   ▼
            ┌─────────────┐
            │    LSTM     │
            └─────────────┘
    Dropout  │  Output
      0.3    ▼   256
            ┌─────────────┐
            │    LSTM     │
            └─────────────┘
    Dropout  │  Output
      0.3    ▼   256
            ┌─────────────┐
            │    DENSE    │
            └─────────────┘
             │  Output
             ▼   32
            ┌─────────────┐
            │    DENSE    │
            └─────────────┘
             │   1
             ▼
            ┌─────────────┐
            │   OUTPUT    │
            └─────────────┘
```

Chapter 4: DESCRIPTION OF TOOL SELECTED

[1] Python Programming Language :

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry. Python Programming Language is very well suited for Beginners, also for experienced programmers with other programming languages like C++ and Java.

[2] Jupyter Notebook:

The Jupyter Notebook is an open-source web application that allows data scientists to create and share documents that integrate live code, equations, computational output, visualizations, and other multimedia resources, along with explanatory text in a single document.

[3] Datasets:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

Chapter 5: DETAILED DESCRIPTION OF MODULES IMPLEMENTED and OUTPUT

A. Regression Based Model

In general, the Regression based Model is used for predicting continuous values through some given independent values. The factors considered for the regression were low, open, high, close and volume. The R-square confidence test was used to determine the confidence score and the predictions were plotted to show the results of the stock market prices vs time.

Output:
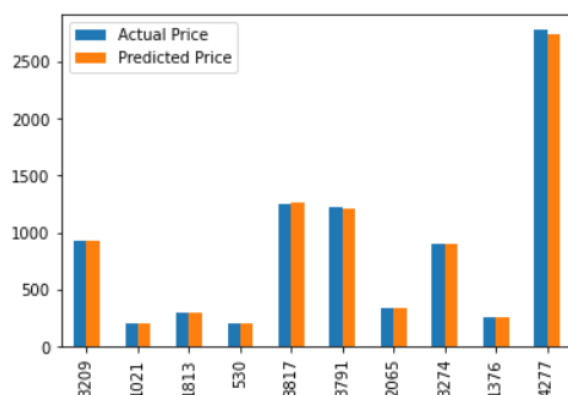
B. Long Short Term Memory (LSTM) Network Based Model

LSTM is the advanced version of Recurrent-NeuralNetworks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM.

Output:
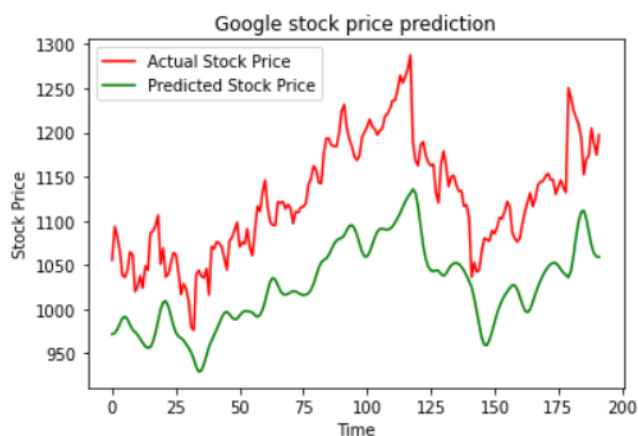
Chapter 6: NEW LEARNINGS FROM THE TOPIC

This paper was an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning techniques. The primary contribution of the researchers being the application of the novel LSTM Model as a means of determining the stock prices.

Both the techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results with the LSTM model proving to be more efficient. The results are quite promising and has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them. Sentiment analysis though Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning based models can also be used for prediction purposes.

REFERENCES AND ANNEXURES

[1] Saurav Agrawal, Dev Thakkar, Dhruvil Soni, Krunal Bhimani, Dr. Chirag Patel, "Stock Market Prediction using Machine Learning Techniques". International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 5, Issue 2, 2019. ISSN: 2456-3307, DOI: https://doi.org/10.32628/CSEITI952296.

[2] K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal, "Stock Market Prediction Using Machine Learning Algorithms". International Journal of Engineering and Advanced Technology, Volume 8, Issue 4, April-2019. ISSN: 2249-8958.

[3] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, "Predicting Stock Prices Using LSTM". International Journal of Science and Research 2017. ISSN: 2319-7064.

[4] S Abdul salam Sulaiman Olaniyi, Adewole, Kayode S, Jimoh, R. G, "Stock Trend Prediction Using Regression Analysis – A Data Mining Approach". ARPN Journal of Systems and Software, Volume 1, Issue 4, 2011. ISSN: 2222-9833.

[5] Gareja Pradip, Chitrak Bari, J. Shiva Nandhini,"Stock market prediction using machine learning". International Journal of Advance Research and Development, Volume 3, Issue 10, 2018.

[6] Vivek Kanade, Bhausaheb Devikar, Sayali Phadatare, Pranali Munde, "Stock market prediction: Using historic data analysis". International journal of advanced research in computer science and software engineering, volume 7, issue 1, 2017. ISSN: 2277 128X. DOI: 10.23956/ijarcsse/V711/0112.