Stock Market Prediction System

A COURSE PROJECT REPORT

Submitted by

RIJUL SARJI (RA2011027010005)

SHUBHANKAR KAMLESH BHOSLE (RA2011027010057)

Under the guidance of

Dr. M. PRAKASH

In partial fulfillment for the Course

of

Machine Learning I (18CSE392T)

in

DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS

SCHOOL OF COMPUTING

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)
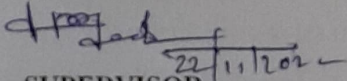
KATTANKULATHUR - 603 203

November, 2022

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this mini project titled **"Stock Market Prediction System"** is the bonafide work of **Rijul Sarji (RA2011027010005)**, **Shubhankar Kamlesh Bhosle (RA2011027010057)**, who carried out the project work under my supervision.
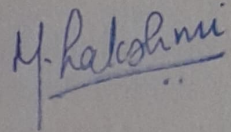
SUPERVISOR

Dr. M. PRAKASH

Associate Professor

Department of Data Science and Business Systems

SRM Institute of Science and Technology

Kattankulathur – 603 203

HEAD OF THE DEPARTMENT

Dr. M. LAKSHMI

Professor & Head

Department of Data Science and Business Systems

SRM Institute of Science and Technology

Kattankulathur – 603 203

ABSTRACT

In today's financial world stock exchange has become one of the most significant events. The world's economy today is widely dependent on the stock market prices. The Stock Market has been very successful in attracting people from various backgrounds be it educational or business .The nonlinear nature of the Stock Market has made its research one of the most trending and crucial topics all around the world.. People decide to invest in the stock market on the basis of some prior research knowledge or some prediction. In terms of prediction people often look for tools or methods that would minimize their risks and maximize their profits and hence the stock price prediction takes on an influential role in the ever challenging stock market business. Adopting traditional methodologies such as fundamental and technical analysis doesn't seem to ensure the consistency and accuracy in the prediction. As a result the machine learning technologies have become the recent trend in the stock market prediction whose prediction is based on the existing stock market values eventually as an outcome of training on their previous values. This paper focuses on RNN (Recurrent Neural Networks) and LSTM (Long Short term memory) technologies in predicting the ongoing trend of the stock market.

ACKNOWLEDGEMENT

We express our heartfelt thanks to our honorable **Vice Chancellor Dr. C. MUTHAMIZHCHELVAN**, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our **Registrar Dr. S. Ponnusamy,** for his encouragement

We express our profound gratitude to our **Dean (College of Engineering and Technology) Dr. T. V.Gopal,** for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing **Dr. Revathi Venkataraman,** for imparting confidence to complete my course project

We wish to express my sincere thanks to **Course Audit Professor** and **Course Coordinator** for their constant encouragement and support.

We are highly thankful to my Course project Faculty Dr. **M. Prakash, Associate Professor, Department of Data Science and Business Systems,** for his assistance, timely suggestion and guidance throughout the duration of this course project.

We extend our gratitude to our **HoD, Dr. M. Lakshmi, Professor, Department of Data Science and Business Systems,** and my Departmental colleagues for their Support.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

TABLE OF CONTENTS

1. INTRODUCTION

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. The rest of the paper consists of following: Section II discusses the related work. Section III puts forward the two models used and the methods used in them in detail. Section IV discusses the results produced with different plots for both the models in detail. While Section V consists of conclusion and the last section involves the references.

2. LITERATURE SURVEY

By **Anusha J Adhikar**
Department of Computer Science and Engineering
Sir M. Visvesvaraya Institute of Technology,
Bengaluru, Karnataka, India

Research on Stock Price Prediction Method Based on Convolutional Neural Network, IEEE 2019- Sayavong Lounnapha et al. This paper intends for a prediction model for stock price which is centered at the convolutional neural networks, that has exceptional capability of learning on its own. The data set is taught and tested relating the behaviors of both Convolutional Neural Networks and Thai stock market The result shows that the model on grounds of Convolutional Neural Networks can effectually recognize the altering trend in stock market price and envisage it which provides significant allusion for stock price forecast. The accuracy of the prediction is found to be elevated, and it could also be promoted in the field of finance.

Enhancing Profit by Predicting Stock Prices using Deep Neural Networks, IEEE 2019-Soheila Abrishami, et al., The prediction of economic time series is quite a herculean task, which has fascinated the attentiveness of many scholars and is extremely vital for investors. This paper focuses on presenting a deep learning system, which makes use of a range of facts for a part of the stocks on the NASDAQ exchange to predict the value of the stock. This model has been trained on the smallest of data for a particular stock and accurately estimates the concluding value of that stock for multi-stepahead. It consists of an auto encoder in order to remove noise and makes use of time series data engineering to syndicate the advanced features with the original features. These new features are given to a Stacked LSTM Autoencoder for multistep-ahead estimation of the stock concluding value. Further, this estimation is used by a profit maximization approach to offer assistance on the right time for buying and selling a particular stock.

The results indicate that the suggested framework outclasses the state of the art time series forecasting methodologies with respect to analytical accuracy and effectiveness.

3. REQUIREMENTS ANALYSIS

- **Numpy**

  NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It is open-source software.

- **Pandas**

  Pandas is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library. Pandas is fast and it has high performance & productivity for users.

- **Sklearn**

  Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering, and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

- **Mathplotlib**

  Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. It is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

4. DATA SET DESCRIPTION

In this notebook we will be looking at data from the stock market, particularly some technology stocks. We will learn how to use pandas to get stock information, visualize different aspects of it, and finally we will look at a few ways of analyzing the risk of a stock, based on its previous performance history. We will also be predicting future stock prices through a Long Short Term Memory (LSTM) method!

We'll be answering the following questions along the way:

```python
company_list = [AAPL, GOOG, MSFT, AMZN]
company_name = ["APPLE", "GOOGLE", "MICROSOFT", "AMAZON"]

for company, com_name in zip(company_list, company_name):
    company["company_name"] = com_name

df = pd.concat(company_list, axis=0)
df.tail(10)
```

Date	Open	High	Low	Close	Adj Close	Volume	company_name
2022-02-03	2834.750000	2884.949951	2766.659912	2776.909912	2776.909912	11276600	AMAZON
2022-02-04	3112.129883	3224.000000	3012.159912	3152.790039	3152.790039	12640500	AMAZON
2022-02-07	3170.399902	3253.820068	3145.000000	3158.709961	3158.709961	5131200	AMAZON
2022-02-08	3135.010010	3235.850098	3111.010010	3228.270020	3228.270020	3802000	AMAZON
2022-02-09	3257.469971	3276.689941	3205.000000	3223.790039	3223.790039	3439300	AMAZON
2022-02-10	3167.000000	3214.330078	3155.000000	3180.070068	3180.070068	3413400	AMAZON
2022-02-11	3162.689941	3180.000000	3054.679932	3065.870117	3065.870117	3851600	AMAZON
2022-02-14	3035.020020	3168.969971	3033.000000	3103.340088	3103.340088	4022000	AMAZON
2022-02-15	3152.110107	3160.139893	3092.750000	3130.209961	3130.209961	2822000	AMAZON
2022-02-16	3115.810059	3161.645020	3095.000000	3155.254883	3155.254883	1675095	AMAZON

The dataset consists of 1024 individuals data. There are 14 columns in the dataset, which are described below.

1. **Date:** displays the date of data

2. **Open**: displays the price at the time of opening.

3. **High**: displays the highest point reached that day.

4. **Low**: displays the lowest point reached that day.

5. **Close**: displays the price at time of closing.

6. **Adj. Close**: displays the serum cholesterol in mg/dl (unit)

7. **Volume**: displays the volume of stocks of the company

8. **company_name**: displays the name of the company

5. ALGORITHM USED

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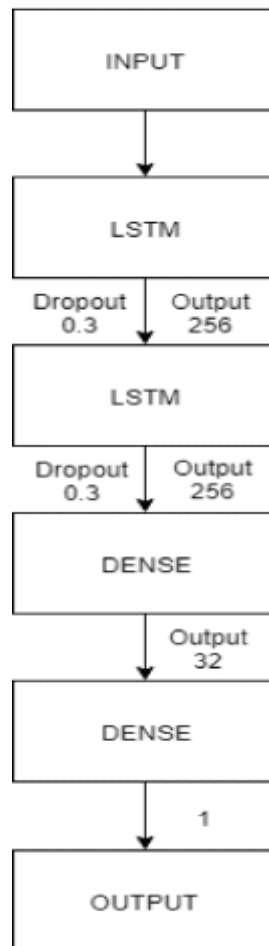
LSTM is the advanced version of Recurrent-Neural-Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs work on finding the relationship between the recent and the current information. This indicates that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

The main purpose behind using this model in stock market prediction is that the predictions depend on large amounts of data and are generally dependent on the long term history of the market. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate. Thus proving itself as much more reliable compared to other methods.

Since the 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.
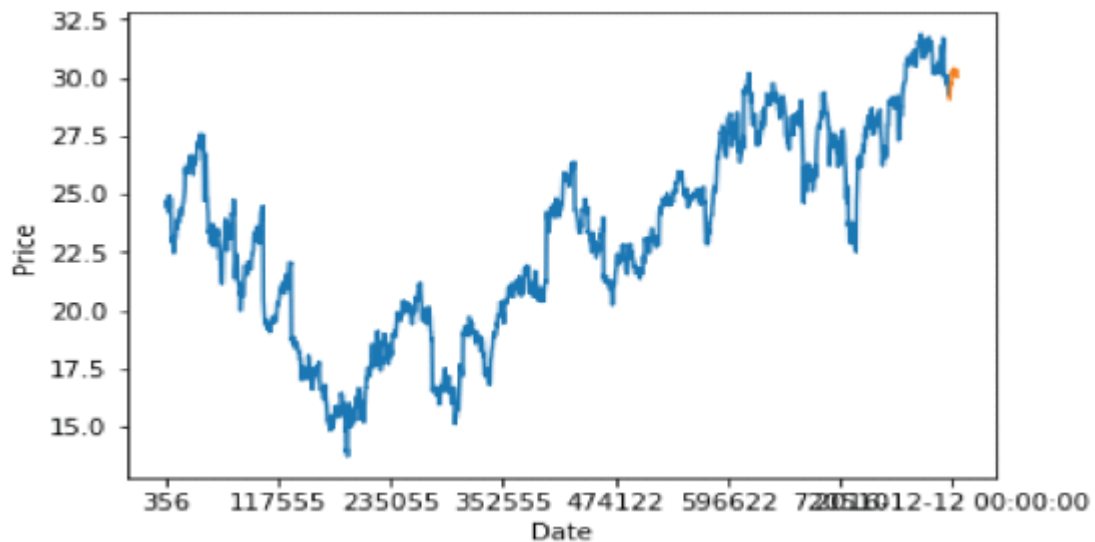
● Linear Regression

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).
Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study.

6. RESULTS AND DISCUSSION

The proposed system is trained and tested over the dataset taken from Yahoo Finance. It is split into training and testing sets respectively and yields the following results upon passing through the different models:

The plot is the result of application of linear regression algorithm on the dataset to predict varying prices with respect to the time.
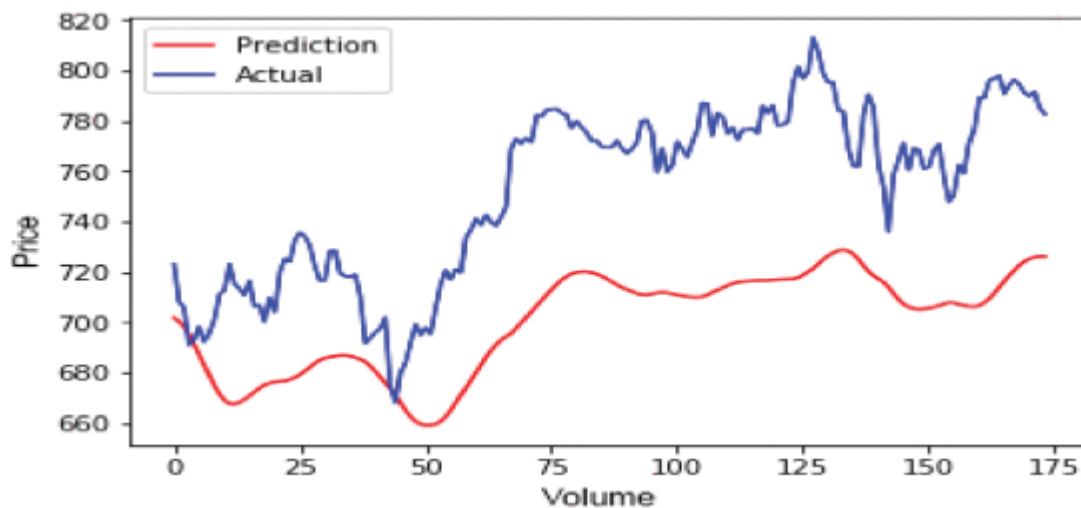
The prediction is shown by the red line and the actual trend is shown by blue. The proximity of these two lines tells how efficient the LSTM based model is. The prediction approximates the real trend when a considerable amount of time has passed. The model resulted in a Train Score of 0.00106 MSE (0.03 RMSE) and a Test Score of 0.00875 MSE (0.09 RMSE). The more the system is trained and the greater the size of the dataset utilized the greater the accuracy which will be attained. The LSTM Model offered more accuracy than the Regression based Model.

7. CONCLUSION AND FUTURE ENHANCEMENT

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.

Future Enhancements:
- We wish to add more parameters in the data set. That way, we can get more accurate results and better predictions.
- We will use more effective classification algorithms to give better performance and accurate results.
- Use of more training data ensures the higher chances of the model to accurately predict the value of the stock .

8. REFERENCES

- Numpy documentation
  https://numpy.org/doc/stable/

- Pandas documentation
  https://pandas.pydata.org/

- Sklearn documentation
  https://scikit-learn.org/stable/

- Stock Market Prediction System

  https://www.ijeast.com/papers/258-262,Tesma508,IJEAST.pdf