**Mini Project Report on**



**Stock Price Prediction System using Machine Learning**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Vinay Pandey 2019229**

***Under the Mentorship of***

**Dr. Sachin Sharma**

**Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January 2023**

GEU logo

**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Stock Price Prediction System using Machine Learning”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Sachin Sharma, Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name University Roll No. **signature**

Vinay Pandey 2019229

**Table of Contents**

| **Chapter No.** | **Description** | **Page No.** |
| --- | --- | --- |
| Chapter 1 | Introduction | 1 |
| Chapter 2 | Literature Survey | 3 |
| Chapter 3 | Methodology | 5 |
| Chapter 4 | Result and Discussion | 8 |
| Chapter 5 | Conclusion and Future Work | 9 |
|  | References | 10 |

**Chapter 1**

**Introduction**

Stock Market is one of the most preferred kinds of investment, given the high amount of profit we can make from it, but also comes with a high risk factor. Many people in the past and current scenario have investigated tools, methods and techniques for increasing their profits and minimizing the risk involved, as the level of trading and investing grew.

* 1. **The US Stock Market Indexes**

Powerful indicators for global and country-specific economies around the world are Stock market indexes. In the United States of America S&P 500, Dow Jones Industrial Average, and Nasdaq Composite are the three most broadly followed indexes, followed by both the media and investors. In addition to which, there are approximately 5,000 others that make up the U.S. equity market. [1]

One of the most widely recognized US stock market indices include the Dow Jones Industrial Average (DJIA), the S&P 500, and the NASDAQ Composite. The DJIA is composed of 30 large-cap stocks, chosen to represent a broad range of industries in the US economy. The S&P 500 is made up of 500 of the largest publicly traded companies in the US and covers approximately 80% of the total market capitalization of the US equity market. The NASDAQ Composite is an index of more than 3,000 companies listed on the NASDAQ stock exchange, with a particular focus on technology and growth-oriented companies.

The two indexes on which stock from technological industries are traded are S&P 500 Index, and The Nasdaq Composite Index.

The Standard & Poor's 500 Index (known commonly as the S&P 500) is an index with 500 of the top companies in the U.S. Stocks are chosen for the index primarily by capitalization and other factors such as liquidity, financial viability, and trading history. The S&P 500 Index represents approximately 80% of the total value of the U.S. stock market. In general, the S&P 500 Index gives a good indication of movement in the U.S. market as a whole. The S&P 500 Index is a market-weighted index (also referred to as capitalization-weighted).

The Nasdaq Composite Index is a market-capitalization-weighted index of all the stocks traded on the Nasdaq stock exchange. This index includes some companies that are not based in the United States. Known for being heavily tech weighted, this index includes several sub sectors across the tech market including software, biotech, semiconductors, and more. Although this index is known for its large portion of technology stocks, it does include some securities from other industries as well.

The stock for which we will be predicting the values is of Apple Inc., which is listed under the name AAPL on the Nasdaq Composite Index and the S&P 500 Index.

**1.2 Objective**

Stock market prediction refers to the process of using various analytical and statistical tools to forecast the future performance of the stock market or individual stocks. This can involve analyzing past market data, economic indicators, company financial statements, and other factors that can influence the stock market.. This helps you to invest wisely for making good profits.

In this project, we will be utilizing a machine learning model, which is created using recurrent neural network (RNN). It is a deep learning model called LSTM (Long Short Term Memory). It is an improved version of recurrent neural networks, which have a drawback of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but are able to give more accurate predictions from the recent information, because of which as the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data.

**1.3 Applications**

This application will be helpful for stock investors, sellers, buyers, brokers, businesses, companies, and government agencies for doing tasks like risk management, economic forecasting, trading etc.

**Chapter 2**

**Literature Survey**

Stock Market Prediction Using Machine Learning Algorithms:

Publication Year: 2016

Author: P. Manisha, K. Aruna, and B. V. Kumar

Journal Name: Computational and Theoretical Nanoscience

Summary: In this paper [2], the authors evaluate the performance of several machine learning algorithms, including decision trees, random forests, and support vector machines, for stock market prediction using historical data.

Forecasting stock price in two ways based on LSTM neural network:

Publication Year: 2019

Author: Jingyi Du, Qingli Liu, Kang Chen, Jiacheng Wang

Journal Name: 2019 IEEE

Summary: The [3] LSTM neural network is used to predict Apple stocks by consuming single feature input variables and multi-feature input variables to check the forecast effect of the model on stock time series. The results of the experiment show that the model has a high accuracy of 0.033 for the multivariate input and is accurate, that is in line with the actual demand. For the univariate feature input, the predicted squared absolute error is 0.155, which is inferior to the multi-feature variable input.

Share Price Prediction using Machine Learning Technique:

Publication Year: 2018

Author: Jeevan B, Naresh E, Vijaya Kumar B P, Prashanth Kambli

Journal Name: 2018 IEEE

Summary: This paper is mostly [4] based on the approach of predicting the share price using Long Short Term Memory (LSTM) and Recurrent Neural Networks (RNN) to forecast the stock value on NSE data using various factors such as current market price, price-earning ratio, base value and other anonymous events. The model train for all the NSE (National Stock Exchange) data from the internet and recognize the input and group them and provide input according to the user configuration this RNN based architecture proved very efficient in forecasting the stock price by changing the configuration accordingly, which also use backpropagation mechanism while gathering and grouping data to avoid mixing of data.

Summary of Literature Survey:

Here, I have reviewed various approaches for Stock price prediction. All approaches have their own advantages and disadvantages. CNN & LSTM is a most popular algorithm to predict the stock price but there are some challenges in this method like needing a lot of training data, high computational cost, without GPU data quite slow to train, depending on any previous information for prediction. Deep learning is able to provide highly accurate prediction results using standard tools and also outperforms all standard prediction methods.

**Chapter 3**

**Methodology**

**3.1 Tools and Technologies Used**

The tools and technologies used are:

**Python –** The language used for this project is Python. This is because of its numerous community based features like ,that it has an abundance of powerful tools ready for scientific computing Packages. Packages like NumPy and Pandas are freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast and efficient.

**Jupyter Notebook** – Jupyter Notebook allows users to write and run code interactively in a web browser, making it an excellent tool for data exploration, data analysis, scientific computing, and machine learning. Users can write code in individual cells and execute them one at a time or all at once, making it easy to test and iterate on code.. [5]

**TensorFlow -** TensorFlow [6] is a popular open-source framework for building and training machine learning models. It is useful for a wide range of tasks, from image and speech recognition to natural language processing and reinforcement learning Made by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well.

Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platforms together with iOS and Android.

Other packages used include NumPy, Pandas, matplotlib and more.

**3.2 Proposed Workflow**

**Diagram

Description automatically generated**

**Figure 3.2.1** Proposed System Architecture and Workflow [7]

The system presented here composes of five modules:-

1. Input as Dataset

2. Pre-processing

3. Data splitting

4. Build & Model train a Stacked LSTM Model

5. Output as Predicted Result

Data of the Apple stock prices have been taken from a website named Tiingo by using a user-allocated API key. The data is then scaled accordingly, and then split into training and testing datasets, after which a Stacked LSTM model is created using the data.

Table

Description automatically generated

**Figure 3.2.2** Summary of the Stacked LSTM Model

**3.3 LSTM Architecture**

**Diagram

Description automatically generated**

**Figure 3.2.1 LSTM Architecture**

Long Short Term Memory is a type of recurrent neural network that addresses the issue of long-term dependencies faced by RNN. In traditional RNN, the output from the previous step is used as input for the current step. However, in situations with long-term dependencies, RNN may not be able to predict accurately over time. LSTM addresses this problem by retaining information over longer periods of time, resulting in more precise predictions based on recent information.As the length of the gap between relevant information increases, traditional RNN may not perform efficiently. However, LSTM is designed to retain relevant information for extended periods of time, making it a powerful tool for processing, predicting, and classifying time-series data.

Chart, line chart, histogram

Description automatically generated

**Figure 3.3.1** Prediction of model on testing data

**Chapter 4**

**Result and Discussion**

By following the described methodology, a stacked LSTM model was developed and trained on a dataset of Apple stock prices obtained from a .csv file and Tiingo data. Utilizing this model for forecasting the stock price for the next 30 days results in the output shown below.

Chart, line chart

Description automatically generated

**Figure 4.1** Prediction by model for next 30 days

**Chart, line chart

Description automatically generated**

**Figure 4.2** Overall graph of stock data with the prediction

**Chapter 5**

**Conclusion and Future Work**

Through this project, I have developed a system that can forecast the stock value for up to the next 30 days. This system has the potential to aid investors in making informed decisions and increasing their overall profitability in the stock market. However, it is important to note that financial decisions should not be solely based on machine learning models as the stock market is highly unpredictable and volatile. It is necessary to conduct thorough research before investing.

Although numerous machine learning models with high accuracy rates are available, this model can be further enhanced in future work. Improvements can be made by expanding the input dataset, increasing the number of epochs, and implementing bi-directional LSTM instead of stacked LSTM to achieve better accuracy.

**References**

[1] Bloomberg. "There Are Now More Indexes Than Stocks [Online]. Accessed on 21st January 2023: <https://web.archive.org/web/20170602050244/https://www.bloomberg.com/news/articles/2017-05-12/there-are-now-more-indexes-than-stocks>

[2] P. Manisha, K. Aruna, B. V. Kumar et al., “Stock market prediction using machine learning algorithms” in 2016 in Journal of Computational and theoretical Nueroscience

[3] J. Du, Q. Liu, K. Chen, and J. Wang, “Forecasting stock prices in two ways based on

LSTM neural network,” in 2019 IEEE 3rd Information Technology, Networking,

Electronic and Automation Control Conference (ITNEC), Mar. 2019, pp. 1083–1086, doi:

10.1109/ITNEC.2019.8729026

[4] B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, “Share Price Prediction using

Machine Learning Technique,” in 2018 3rd International Conference on Circuits,

Control, Communication and Computing (I4C), Bangalore, India, Oct. 2018, pp. 1–4, doi:

10.1109/CIMCA.2018.8739647

[5] Project Jupyter.” https://www.jupyter.org (accessed Feb. 23, 2023).

[6] J. Brownlee, “Introduction to the Python Deep Learning Library TensorFlow,” Machine

Learning Mastery, May 04, 2016. https://machinelearningmastery.com/introductionpython-deep-learning-library-tensorflow/ (accessed Jan. 23, 2023).

[7] Pramod, & Pm, Mallikarjuna. (2021). Stock Price Prediction Using LSTM. Test Engineering and Management. 83. 5246-5251.