

Stock Price Prediction using Time Series Analysis and Sentimental Analysis

A Major Project Synopsis Submitted to



**Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal
Towards Partial Fulfillment for the Award of**

**Bachelor of Technology
(Computer Science and Engineering)**

**Under the Supervision of
Prof.
Praveen Bhanodia**

**Submitted By
Suyash Jain(0827CS181221)
Yagvendra Khichi(0827CS181241)
Yashwant Patidar(0827CS181248)
Yogesh Sharma(0827CS181249)**



**Department of Computer Science and Engineering
Acropolis Institute of Technology & Research, Indore
Jan-June 2022**

1. 1. Abstract

Stock price analysis is in demand and an interesting topic in financial and academic research. For direct economic gain, stock forecasts have always attracted people interested in investing in stock markets and stock exchanges. It is also a great research topic in finance. Predicting stock market returns is a very complex issue and depends on many factors such as the company's financial position, national affairs, and public sentiments. Stock prices today are influenced by many reasons, including corporate news, and political and socio-economic conditions.

2. 2. Introduction of the Project

Stock Price Prediction is a complex and crucial task in the current machine learning domain. Multiple approaches have been proposed to solve this problem. Most of them are based on a Time Series analysis of a stock value, others are based on the news sentiments to be used to predict a particular stock value. The stock prediction problem is particularly very interesting to work on because there are several sides to this problem. The price value depends on people, sentiment, company performance, news, and regulators. Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to take calculated and research-based decisions to make significant gains.

3. 3. Objective

The aims of the project are as follows:

- To predict the future stock price.
- To generate patterns from the historical dataset for the prediction of stock price.
- To Analysis the public sentiment about a particular company.
- To provide a platform for investors from which they can select which company to invest.

The project will be useful for investors to invest in the stock market based on various factors. The project target is to create an application that can analyze the historical data of the companies and implement these values to the model created to determine the value of that particular stock will have in near future with suitable accuracy and also analyze public sentiments about that company on social media platform Twitter.

The main feature of the project is to generate approximate forecasting of stock price and create a general idea of future values based on previous data by generating a pattern and extracting public perception about that company from the social media platform Twitter.

4. 4. Scope

Stock price analysis of stocks using historical data helps new investors invest in stocks based on the time series considered by our software. A company's stock price depends on many factors, including supply and demand, company earnings or quarterly corporate performance, inflation, trader and investor behavioral factors, and the current state of the country's economy and its stocks. For those who try their luck on the stock market, "Stonks" is a general suggestion tool that helps you analyze your data to predict stock prices and select stocks with better returns.

5. 5. Study of Existing System

The Existing systems work on various methods from regression to classification. There are many drawbacks such as:

- The existing system fails when there are rare outcomes or predictors, as the algorithm is based on bootstrap sampling.
- The previous results indicate that the stock price is unpredictable when the traditional classifier is used.
- The existence system reported highly predictive values, by selecting an appropriate time period for their experiment to obtain highly predictive scores.
- The existing system does not perform well when there is a change in the operating environment.
- It exploits only one data source, thus highly biased.
- The existing system needs some form of input interpretation, and thus needs scaling.
- It doesn't exploit data pre-processing techniques to remove inconsistency and incompleteness of the data.
- The existing systems either predict stock prices using historical data or sentimental analysis.

There is a manual method too that can predict the price such as statistical, and sentiment analysis.

6. 6. Project Description

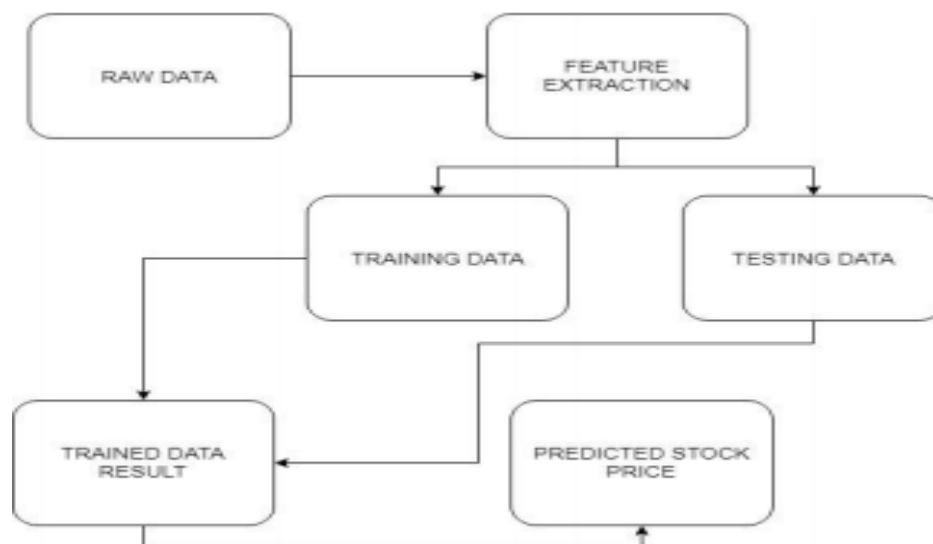
The Proposed system is “Stonks”. A web application through which users can predict the future price of a stock and can analyze the pattern from previous data.

The features of the project are as follows:

- Predict stock price
- Analyze the historical data
- Compare stocks over time using graphs.
- Analyze public sentiments of the company.

This project uses a deep learning algorithm (LSTM) model for prediction. Python is used for software development. Various Python libraries such as Pandas, Seaborn, Matplotlib, NumPy, and Plotly are also used for data cleaning and data visualization.

Keras and TensorFlow are used to develop machine learning models (LSTMs) and predict stock prices. Flair pre-trained model based on Pytorch is used to analyze the public sentiments using the NLP algorithm.



7. Methodology/Planning of the Project work

LSTM (LONG SHORT-TERM MEMORY)

Long short-term memory LSTMs have the property of storing information for a long period of time with default behavior. See the following figure. Each LSTM module has three gates, called oblivion gates, input gates, and output gates. Gate of long-term short-term memory.

1) Input gate — Check the value of the input used to change the storage. The sigmoid function determines the value passed by 0.1. The function then weights the passed values and determines their importance in the range of 1 to 1 input ports.

2) Forget gate — Find the details you want to discard from the block. This is determined by the sigmoid function. Examines the previous state (ht_1) and content input (X_t) and returns a number from 0 (omitted) to 1 (retained) for each number in cell state C_{t-1} .

Forget Gate Output Gate – The block's inputs and memory are used to determine the output. The sigmoid function determines the value passed by 0.1. The function weights the passed value by determining the importance in the range 1 to 1 and multiplying it by the output of the Sigmoid.

3) Output gate — The block's inputs and memory are used to determine the output. The sigmoid function determines the value passed by 0.1. The function weights the passed value by determining the importance in the range 1 to 1 and multiplying it by the output of the Sigmoid.

Sentimental Analysis

Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative, or neutral. Sentiment analysis is often performed on textual data to help companies monitor brand and product sentiment with customer feedback and understand customer needs.

Data Extraction and Preprocess

For building the predictive model we need well-defined data to extract meaningful insight and preprocess it to eliminate any error. For data extraction and preprocess, we use the following library:

PANDAS

Pandas is an open-source python library that is used to modify the data in a way that it can be used to develop a model without any missing or wrong value in the dataset.

SEABORN

Seaborn is a python library for data visualization. Seaborn is used to creating the graph for comparison and exploratory data analysis.

Plotly

Plotly Library is an open-source library that can be used for data visualization and understanding data simply and easily.

Flair

Flair allows you to apply state-of-the-art Natural Language Processing (NLP) models to your text. B. Named entity recognition (NER), part-of-speech tagging (PoS), dedicated support for biomedical data, disambiguation, and classification by rapidly increasing language support.

Web Technology

Streamlit

Streamlit is an open-source app framework for creating and deploying data science applications.

8. 8. Expected Outcome

The built project will be able to clean up and extract data using pandas. The processed data is now ready to build a predictive model. Predictive models are developed using LSTM algorithms built using the Keras and TensorFlow libraries. Sentimental Analysis was done using the NLP algorithm which shows the public sentiments about the particular company whether it is neutral, negative, or positive using Flair library. The front end (GUI) is created with Streamlit. The project can predict stock prices with reasonable accuracy and compare stock prices from time to time. This project will help new entrants and analysts in the investment sector predict and compare stock prices.

9. 9. Resources and Limitations

The resources needed for the creation of the software are:

- Python-Python is an interpreted programming language.
- Pandas-It is a python open-source for data manipulation and data cleaning.
- Seaborn-Seaborn is a data visualization library for creating graphs.
- Keras-Keras is deep learning for creating an artificial neural network.
- LSTM-Long Short-Term Memory is a deep neural network algorithm.
- Sklearn-Scikit-learn is an open-source machine learning library.
- Streamlit - open-source app framework for creating and deploying data science applications.
- Plotly- An open-source library that can be used for data visualization and understanding data simply and easily

- NumPy- library for the Python programming language, adding support for large, multi-dimensional arrays and matrices.
- Flair- NLP library for sentimental analysis.
- Anaconda-Anaconda is a distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment.
- Spyder-Spyder is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language.

These are the resource for the development of the working project.

The limitations of the project are below:

- The training of data is slow.
- Internet connection is required.
- System (operating system) is required to execute.
- Software excludes the minor factors (unpredictable events) that may impact the outcome.

10. 10. Conclusion

Currently, there are several models for predicting the stock market, but they are less accurate. In order to predict stock price movements more accurately, we proposed a model using RNNs and LSTMs. LSTM has introduced a memory cell, which is an arithmetic unit that replaces the conventional arithmetic unit artificial neurons in the hidden layer of the network. In this task, increasing the epoch and batch size will improve the prediction accuracy. The proposed method uses the test data used to predict which results will be more accurate than the test data. The proposed method can track and forecast the stock market, and the forecast provides higher and more accurate results. The proposed system also provides the Sentimental analysis that gives users information about public sentiments about the company using the flair library. The above model gives accurate results with more convenient for equity analysts, business analysts, and stock market investors.

11. 11. References

- [1]. Shriram.S, Dr. K. Anuradha, Dr. K.P. Uma. “Future Stock Price using Recurrent Neural Network, LSTM and Machine Learning”, 2021
- [2]. Usmani S, Shamsi JA. “News sensitive stock market prediction: literature review and suggestions”, 2021
- [3]. Dev Shah, Haruna Isah and Farhana Zulkerine. “Stock market Analysis: A Review and Taxonomy of Prediction Techniques”, 2019
- [4]. K. Hiba Sadia, Aditya Sharma, Adarrsh paul, Sarmishtha Padhi, Saurav Sanyal. “Stock Market Prediction Using Machine Learning Algorithms”, 2019.
- [5]. Prit Modi,Shaival Shah,Himani Shah. “Big Data Analysis in Stock

Market Prediction”, IJERT 2019.

[6]. Yanmin Zhu, Yanyan Shen, Hao Li. “Stock Price Prediction Using Attention-based Multi-Input”, ACML 2018.

[7]. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.

[8]. Kordonis, John & Symeonidis, Symeon & Arampatzis, Avi. ”Stock Price Forecasting via Sentiment Analysis on Twitter”, 2016

[9]. Kalyani Joshi, Bharthi H.N, Jyothi Rao. “Stock Trend Prediction using News Sentiment Analysis”, 2016

[10]. Anshul Mittal, Arpit Goel. “Stock Prediction using Twitter Sentimental Analysis”, 2011

**[11]. Available at
<https://www.analyticsvidhya.com/blog/2021/10/machine-learning-model-deployment-using-streamlit/>**

[12]. Available at <https://github.com/Rajat-dhyani/Stock-Price-Predictor>

**[13]. Available at
<https://medium.com/codex/stock-price-prediction-a-modified-approach>**