STOCK MARKET PREDICTION USING MACHINE LEARNING

Department: Computer Engineering

WARAD GAJANAN DALAL
Student of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology, Nagpur

SHREYAS DILIP NIMJE
Student of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology, Nagpur

VEDANT CHANDRASHEKHAR BOPANWAR
Student of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology, Nagpur

Project Guide:- DR. SAMIR AJANI
Professor of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology, Nagpur

Abstract--
With the prevalence of big data, deep learning has become an increasingly popular method for forecasting stock market trends and prices. We gathered two years of data from the Indian stock market and developed a deep learning model that incorporates a thorough approach to feature engineering in order to predict stock market price trends. Our approach includes a range of preprocessing techniques and multiple feature engineering methods, combined with a customized deep learning system specifically designed for predicting stock market price trends. We compared our proposed solution to frequently used machine learning models and found that our approach outperforms them due to the comprehensive feature engineering we employed. Our system achieves high levels of accuracy in predicting stock market trends. This work contributes to the research community by providing detailed information on prediction term lengths, feature engineering, and data preprocessing methods for stock analysis in both financial and technical domains.

I. INTRODUCTION

Stock market forecasting means predicting the current trends of a company and predicting the value of the stock, regardless of whether it will go up or down. A stock market is a place where shares of a company are traded. A stock is an investment in an institution where it represents ownership of a company. The stock market is where these shares are bought. Buying shares in a company is owning a small part of the institution. We forecast stock prices using a machine learning algorithm to develop a model that effectively predicts the stock price based on current market trends. We used LSTM recurrent neural networks to accurately predict stock prices. You will find two types of stocks, one of which was intraday trading, which is known to us as day trading. Intraday trading means that all positions are assigned then and there before the market closes and ownership is not changed after the end of the day. LSTMs are very important because they are very powerful for sequence prediction problems because they can store past or past data. This is very important in stock forecasting because we need to store and read past stock data and accurately forecast stock prices in the future.

It is one of the most popular investments due to the high returns of the stock market. People are looking for methods and tools that will increase their profits and minimize their risks while trading and investing. Two stock exchanges namely the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) where most of the business in the Indian stock market takes place. Sensex and Nifty are two major Indian market indices. Because stock prices are dynamic, forecasting the stock market is difficult. For this purpose, forecasting models have been gradually developed in recent years and applied to money market forecasting. This

classification is usually made:

1. Time series analysis
2. Comprehensive analysis
3. Technical analysis

II.   MOTIVATION:

Businesses mostly care about client satisfaction and product reviews. It has been demonstrated that changes in social media mood are correlated with changes in the stock markets. Recognising client complaints and subsequently resolving them increases both customer satisfaction and a company's credibility. Therefore, a neutral automated method is required to categorise client reviews of any issue. Companies may have accumulated mountains of consumer input in today's climate, where we are understandably suffering from data overload (although this does not necessarily translate into better or deeper insights), but it is still hard for us mere mortals to manually analyse it without error or bias. Companies with the best of intentions frequently find themselves in a lack of insights. You are aware that you need insights to guide your decisions and that you currently lack them, but you are unsure of the best way to acquire them. Sentiment research offers some insights into the most crucial concerns, at least as seen by the customers. Decisions can be based on a considerable amount of data rather than simple intuition because sentiment analysis can be automated. The primary driver behind stock price predictions is the expected price of a stock in the future. Environmental science, finance, and economics motivation can be helpful in a variety of situations, including as business and industry. Determine the company's stock's potential future value.

III.   LITERATURE REVIEW:

Utilising the CNN-sliding window model, LSTM, RNN, and stock price prediction: [3] Three distinct deep learning models were subjected to the experiment. Compared to the other two models, CNN provides results that are more precise. This is because CNN does not rely on any historical data to make predictions. It just makes predictions using the open window. This makes it possible for the model to comprehend the dynamic patterns and changes present in the current window. RNN and LSTM, on the other hand, make predictions about upcoming events using data from past lags. Using news on online social networks to train neural networks for stock market prediction: [5] Although RNNs can theoretically use the data in arbitrary long sequences, in reality they are only capable of going back a small number of steps. The inability of typical RNNs to learn long-term dependencies is due to the vanishing gradient problem. GRUs, or gated recurrent units, were suggested. LSTM neural networks are used to anticipate price movement on the stock market. [1] This project's goal is to investigate how recurrent neural networks, namely LSTM networks, can be used to anticipate changes in stock market prices. Examine their effectiveness in terms of accuracy and other metrics using trials on actual data, and determine whether they offer any advantages over more conventional machine learning techniques.

Recurrent neural networks are applied to the Sri Lankan stock market to forecast daily stock prices: [4] Recurrent neural networks (RNN) are a subclass of neural network that use feedback connections. Several RNN models are used to predict financial time series. This study's objectives included developing models utilising the Recurrent Neural Network (RNN) Approach to estimate daily stock values of specific listed companies on the Colombo Stock Exchange (CSE), as well as assessing the models' accuracy and, if necessary, highlighting any shortcomings. a technique for predicting stock return based on LSTM The Chinese stock market is examined [2] One sequence was defined as a sequential collection of the daily dataset of any one stock in a specific time (N days) in our LSTM model for stock prediction. The closing price and trade volume on a certain day during these N days are two sequence learning features that are used to explain the performance of the stock in the daily dataset.

IV.   METHODOLOGY:

The two broad groups of prediction techniques are statistical techniques and artificial intelligence techniques. The logistic regression model, the ARCH model, and other statistical techniques are examples. Multi-layer perceptrons, convolutional neural networks, naive Bayes networks, backpropagation networks, single-layer LSTMs, support vector machines, recurrent neural networks, etc. are examples of artificial intelligence techniques.
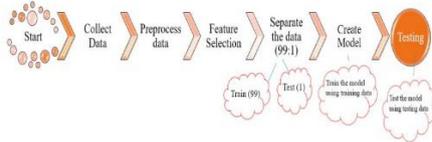
A. Activity Diagram:
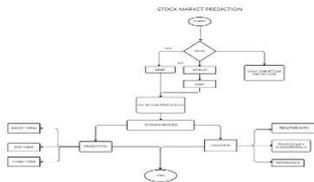
B. Requirements Gathering:

One of the most important stages in creating a stock market prediction app is requirements gathering. It is crucial to compile a thorough list of criteria in order to make sure the app is effective at forecasting market behaviour and fits the needs of its users. Here are some important criteria to take into account:

C. System Design:

The user interface (UI) need to be intuitive and simple to use. Users should have little trouble entering data and viewing the app's recommendations in a straightforward manner.

We created the system architecture, backend API, and frontend UI using the requirements gathered. To develop the system, we used Python, the Streamlit Library, Machine Learning, and Pandas. To verify

the system's operation, we additionally added third-party APIs like Yahoo Finance and Web Scraping.

D. Implementation:

Raw Stock Price Dataset: The official website of the Mumbai Stock Exchange (BSE) is where day-by-day historical stock prices of chosen companies are gathered.

The system was put into place in stages, beginning with the backend API (Yahoo Finance).

After that, we used Python and the Streamlit Libraries to create the front-end interface, adding functions for user authentication, stock searching, and timeline filtering. In order to facilitate user contact, we have integrated email services.

E. Testing:

After the system was put into place, we tested it repeatedly to make sure it complied with the specifications and was error-free. We used manual testing to confirm the system's usability and user experience while using automated testing techniques to assess the system's operation. In order to get user feedback and incorporate their suggestions and criticism into the system, we also conducted user acceptance testing.

In conclusion, we used a methodical process that included requirements collecting, system design, implementation, and testing to create our stock market prediction system. We built the system using Python's Streamlit Libraries and third-party APIs, implementing the functionality discovered throughout the requirements collecting process.

V. RESULT, DISCUSSION:

Tables and graphs have been utilised to show the findings after the project was completed successfully.

For this study, the dataset was split into training and testing sets in a 70:30 ratio. This ratio was chosen

because it would be adequate for the demands of the project's predictive analytics, which incorporates a lot of variable data.

Here, we show the value of the training dataset's close price vs. time chart.

The real and predicted values are found to be close to each other using this method, providing testing accuracy that is acceptable. This graph displays the predicted and actual closing prices for a certain company's stock.

Due to data dependencies and the fact that the research involves time series forecasting, it is vital to train the model primarily using historical data. Additionally, a subset of the data was reserved for cross-validation. In order to ensure random sampling, which improves the model's performance during testing, random rows were also assigned to the training data and the testing data. The dataset's division into training and testing data is handled by the Scikit Learn framework. The ratio used to divide training and testing data is crucial since it affects how well models function. Similar training and test data can cause models to be overfitted, whereas vastly different dataset values can cause models to be underfitted. Therefore, it is essential to have a suitable train-to-test dataset ratio in order to acquire a true picture of the performance of different models, which in turn would produce trustworthy findings.

VI. CONCLUSION:

In this essay, we examine the development of businesses across a range of industries in an effort to determine the ideal time frame for forecasting share prices in the future. The crucial inference from this is that businesses in a particular industry have similar dependencies and growth rates. If the model is trained using a larger sample size of the data set, the prediction may be more accurate.

Furthermore, there may be room for unique business analysis in the projection of different shares.

To improve accuracy, we may examine the various share price trends of various industries and analyse a graph with a wider range of time periods. This approach largely aids in market analysis and growth projections for various companies over time. The prediction accuracy may be improved by include additional factors (such as investor sentiment, election results, and geopolitical stability) that are not directly connected with the closing price.

The machine learning models LSTM and GRU, which are contemporary iterations of recurrent neural networks, are used for stock market prediction and analysis.

The LSTM and GRU models are trained by feeding them historical datasets and statistical data, upon which it has learned and adapted to the pattern and forecasted the future stock price value, which is roughly and closely to the original value.

VII. REFERENCES:

[1]F. a. o. Eugene, "Efficient capital markets: a review of theory and empirical work," Journal of finance, vol. 25, no. 2, pp. 383-417, 1970.

[2]Z. A. Farhath, B. Arputhamary and L. Arockiam, "A Survey on ARIMA Forecasting Using Time Series Model," Int. J. Comput. Sci. Mobile Comput, vol. 5, pp. 104-109, 2016.

[3]S. Wichaidit and S. Kittitornkun, "Predicting SET50 stock prices using CARIMA (cross correlation ARIMA)," in 2015 International Computer Science and Engineering Conference (ICSEC), IEEE, 2015, pp. 1-4.

[4]D. Mondal, G. Maji, T. Goto, N. C. Debnath and S. Sen, "A Data Warehouse Based Modelling Technique for Stock Market Analysis," International

Journal of Engineering & Technology, vol. 3, no. 13, pp. 165-170, 2018.

[5]G. Maji, S. Sen and A. Sarkar, "Share Market Sectoral Indices Movement Forecast with Lagged Correlation and Association Rule Mining," in International Conference on Computer Information Systems and Industrial Management, Bialystok, Poland, Sprigner, 2017, pp. 327-340.

[6] M. Roondiwala, H. Patel and S. Varma, "Predicting stock prices using LSTM," International Journal of Science and Research (IJSR), vol. 6, no. 4, pp. 1754-1756, 2017.

[7] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," PloS one, vol. 14, no. 2, p. e0212320, April 2019.

[8] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics, 2017.

[9] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen," Diploma, Technische Universität München, vol. 91, no. 1, 1991.

[10] Y. Bengio, P. Simard, P. Frasconi and others, "Learning long-term dependencies with gradient descent is difficult," IEEE transactions on neural networks, vol. 5, no. 2, pp. 157-166