Mini Project Report on

STOCK MARKET PRICE PREDICTION USING ML

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Student Name	University Roll No.
Saurabh Pandey	2219591

Under the Mentorship of
Assistant Professor
Dr. Ajay Shukla

Department of Computer Science and Engineering
Graphic Era Hill University
Dehradun, Uttarakhand
July-2024

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"Stock Market Price Prediction using ML"** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era Hill University, Dehradun shall be carried out by the under the mentorship of **Dr. Ajay Shukla, Assistant Professor**, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun.

Name :   University Roll.no:

Saurabh Pandey   **2219591**

Table of Contents

Chapter 1

Introduction

Abstract: In order to maximize earnings or avoid losses, traders and investors must make well-informed judgments in the complex and dynamic world of the stock market. Because there are so many variables that can affect the market, precisely predicting stock prices has always been difficult. Researchers have tried to better predict stock values by utilizing machine learning techniques since their introduction. This thesis compares and evaluates several models and characteristics in order to investigate the use of machine learning algorithms for stock price prediction using historical data. With this research, we hope to shed light on the practicality and efficacy of machine learning in the financial industry.

1.1 Introduction

As a dynamic and intricate system, the stock market affects the financial stability of individuals, companies, and entire countries, which is why it is so important to the global economy. Accurate stock market price prediction has long been valued for its substantial benefits, which include maximizing profits, reducing risks, and optimizing investing strategies. Despite their usefulness, traditional stock price prediction techniques have frequently failed because of the stock market's inherent volatility and complexity. Recent technological developments, especially in the area of machine learning (ML), have created new opportunities to improve the precision and dependability of stock market forecasts.

A branch of artificial intelligence known as "machine learning" deals with the application of algorithms that are capable of learning from and forecasting data.

In contrast to conventional statistical techniques, machine learning algorithms are able to evaluate enormous volumes of data, spot complex patterns, and enhance their predicting capabilities over time via iterative learning procedures. Because of this, machine learning is especially well-suited for stock market prediction, where complicated interdependencies and massive data sets are common place.

There are difficulties in using ML to anticipate stock market movements. Numerous factors, such as economic statistics, business earnings reports, geopolitical events, and investor attitude, all have an impact on the stock market and add to its unpredictable nature. Furthermore, the prediction process can be made more difficult by noise and abnormalities in stock market data that mask underlying trends and patterns. Notwithstanding these difficulties, machine learning models have shown a great deal of potential in identifying and using these intricate relationships to offer more precise and useful stock price forecasts.

We explore the field of machine learning-based stock market price prediction in this research. We start with a thorough review of the literature, including the methods, results, and current research in this area. The methods section that follows describes the particular machine learning approaches and models used in our investigation, including information on data gathering, preprocessing, model selection, training, and validation procedures. The performance of the various ML models is finally shown and examined in the results section, which also offers insights into their applicability and possible real-world stock market prediction situations.

Chapter 2

Literature Survey

2.1 Conventional Techniques for Predicting Stock Prices

Historically, technical analysis, statistical techniques, and fundamental analysis have all been used in tandem to predict stock prices. Assessing a company's earnings, assets, liabilities, and other financial metrics in order to determine its overall financial performance and health is known as fundamental analysis. On the other side, technical analysis looks for trends and patterns in past price and volume data. By modeling time series data, statistical techniques like generalized autoregressive conditional heteroskedasticity (GARCH) models and autoregressive integrated moving averages (ARIMA) have also been widely utilized to predict stock values.

Although these conventional techniques have proven valuable, they frequently fail to capture the intricate and non-linear correlations present in stock market data. Additionally, they are usually constrained by their dependence on particular presumptions.
and their incapacity to change with the dynamics of the market.

2.2 Machine Learning's Emergence in Stock Price Prediction

The field of stock price prediction has undergone a revolution with the introduction of machine learning. While decision trees and linear regression were the key early ML applications in this field, they were still constrained by their simplicity even though they offered certain advantages over more conventional techniques. More sophisticated approaches, such support vector machines (SVMs), ensemble methods, and artificial neural networks (ANNs), have been created and successfully used to stock price prediction as machine

learning (ML) has progressed.

2.3 Neural networks made artificially (ANNs)

Artificial Neural Networks, which draw inspiration from the neural architecture of the human brain, are becoming increasingly popular for stock price prediction. ANNs are made up of linked layers of nodes, or neurons, that perform enter information and provide forecasts. A subclass of ANNs with numerous hidden layers called deep learning has demonstrated impressive ability in identifying intricate patterns in sizable datasets. Research has shown that ANNs can accurately simulate non-linear interactions and correlations in stock market data, resulting in predictions that are more precise.

For example, Kim and Han (2000) discovered that their backpropagation neural network model performed better than conventional linear models in their prediction of stock values. Similar gains in accuracy were made by Fischer and Krauss (2018) when they used long short-term memory (LSTM) networks, a kind of recurrent neural network (RNN), for stock market prediction.

2.4 SVMs, or support vector machines

Another effective machine learning method for predicting stock prices is support vector machines. SVMs work well in environments with several dimensions and can manage problems including both regression and classification. They function by determining the best hyperplane to divide data points belonging to various classifications. SVMs have been used to forecast stock price movements based on technical indicators and historical data in the context of stock price prediction.

SVMs were used by Cao and Tay (2001) to forecast stock index prices, and they showed better results than with more conventional techniques. In order to

further improve predicted accuracy, Huang, Nakamori, and Wang (2005) integrated SVMs with evolutionary algorithms to optimize the model parameters.

2.5 Group Approaches

In terms of stock price prediction, ensemble methods—which incorporate the forecasts of several base models—have also demonstrated promise. While minimizing their shortcomings, strategies like gradient boosting machines and random forests capitalize on the advantages of each particular model. Ensemble approaches work especially well in
enhancing generalization and decreasing overfitting.

Gu, Kelly, and Xiu (2020) combined many machine learning models to forecast stock returns, demonstrating the effectiveness of ensemble learning in stock price prediction. Their strategy produced notable gains in prediction accuracy, underscoring the potential of ensemble techniques in this field.

Models that combine to improve stock price prediction, hybrid models that include many ML techniques have been investigated in addition to individual ML techniques. For instance, Atsalakis and Valavanis (2009) produced a hybrid prediction model that outperformed solo models by combining ANNs with fuzzy logic. Comparably, Patel et al. (2015) created a hybrid model that combined random forests, ANNs, and SVMs and showed better accuracy in stock price prediction.

2.6 Obstacles and Restrictions

Even with machine learning's advances, a number of problems still exist. in predicting stock prices. The availability and quality of data is a significant challenge. The performance of machine learning models can be impacted by

anomalies and noise present in stock market data. Furthermore, a major challenge for machine learning algorithms is the existence of non-stationarity in stock prices, where statistical characteristics vary over time.

Complex machine learning models' interpretability presents another difficulty. Even though deep learning models have a high predictive accuracy, they frequently function as "black boxes," making it challenging to comprehend the logic behind the predictions they make. Their practical application may be hampered by this lack of interpretability, especially in the financial sector where transparency is essential.

Chapter 3

Methodology

This section describes the methodology employed in predicting stock market prices using Machine Learning (ML). The process involves several key steps, including data collection, preprocessing, model development, training, validation, and evaluation. The implementation is primarily conducted using Python, leveraging libraries such as NumPy, Pandas, yFinance, Matplotlib, Scikit-Learn, Keras, and Streamlit for various stages of the workflow.

3.1 Data Collection

Data collection is the foundational step in building an ML model for stock market price prediction. In this project, historical stock data was sourced using the yFinance library, which provides a convenient interface to download financial data from Yahoo Finance.

The data was collected for Google (GOOG) stock over a period of 11 years, from January 1, 2012, to January 1, 2023. This dataset includes various attributes such as Open, High, Low, Close prices, and Volume.

3.2 Data Preprocessing

Data preprocessing is critical for preparing the dataset for ML models. This involves cleaning the data, handling missing values, normalizing, and scaling the features.

3.2.1 Handling Missing Values:

The dataset may contain missing values which can distort the learning process. These missing values are handled by dropping them to ensure a clean dataset.

3.2.2 Calculating Moving Averages:

Moving averages are calculated to smooth out short-term fluctuations and highlight longer-term trends. The 50-day and 100-day moving averages are computed and plotted.

3.2.3 Splitting the Data:

The dataset is split into training and testing sets. The training set comprises 80% of the data, while the testing set consists of the remaining 20%.

3.2.4 Scaling the Data:

Feature scaling is performed using Min-Max Scaling to normalize the data within a range of 0 to 1. This step ensures that the ML model treats all features equally.

3.2.5 Creating Training and Testing Data:

For the ML model, especially LSTM which relies on sequential data, it is crucial to prepare the data in the form of sequences. Here, sequences of 50 days of stock prices are used to predict the next day's price.

3.3 Model Development

The model used for stock price prediction is a Long Short-Term Memory (LSTM) network, which is a type of Recurrent Neural Network (RNN) particularly well-suited for time series data.

3.3.1 Building the LSTM Model:

The LSTM model is built using the Keras Sequential API. It consists of multiple LSTM layers with dropout regularization to prevent overfitting.
Compiling the Model:

The model is compiled using the Adam optimizer and mean squared error (MSE) loss function.

3.3.2 Training the Model:

The model is trained on the training dataset for 50 epochs with a batch size of 32. This involves iteratively updating the model parameters to minimize the loss function.

The Streamlit application provides an interactive interface for users to input stock symbols, view stock data, and see the predicted prices along with visualizations of moving averages and actual vs. predicted prices.

The code provided uses a Long Short-Term Memory (LSTM) network, which is a type of Recurrent Neural Network (RNN). LSTMs are particularly well-suited for sequence prediction problems, such as time series forecasting, because they can capture and learn from temporal dependencies in the data.

Chapter 4

Result and Discussion

The results of this study, which involved developing and evaluating a stock price prediction model using LSTM (Long Short-Term Memory) neural networks, are detailed below. The methodology section describes the use of Python libraries, data preprocessing, model training, and evaluation. The main outcomes are presented through graphs illustrating the performance of the model.

4.1 Stock Data Analysis

The stock data for Google (GOOG) was downloaded from Yahoo Finance, covering the period from January 1, 2012, to January 1, 2023. The data includes daily closing prices, which were used to calculate moving averages and train the LSTM model.

4.1.1 Moving Averages:

100 Day Moving Average:

This metric smooths out short-term fluctuations and highlights longer-term trends. The graph below (Figure 1) shows the 100-day moving average plotted alongside the actual closing prices.
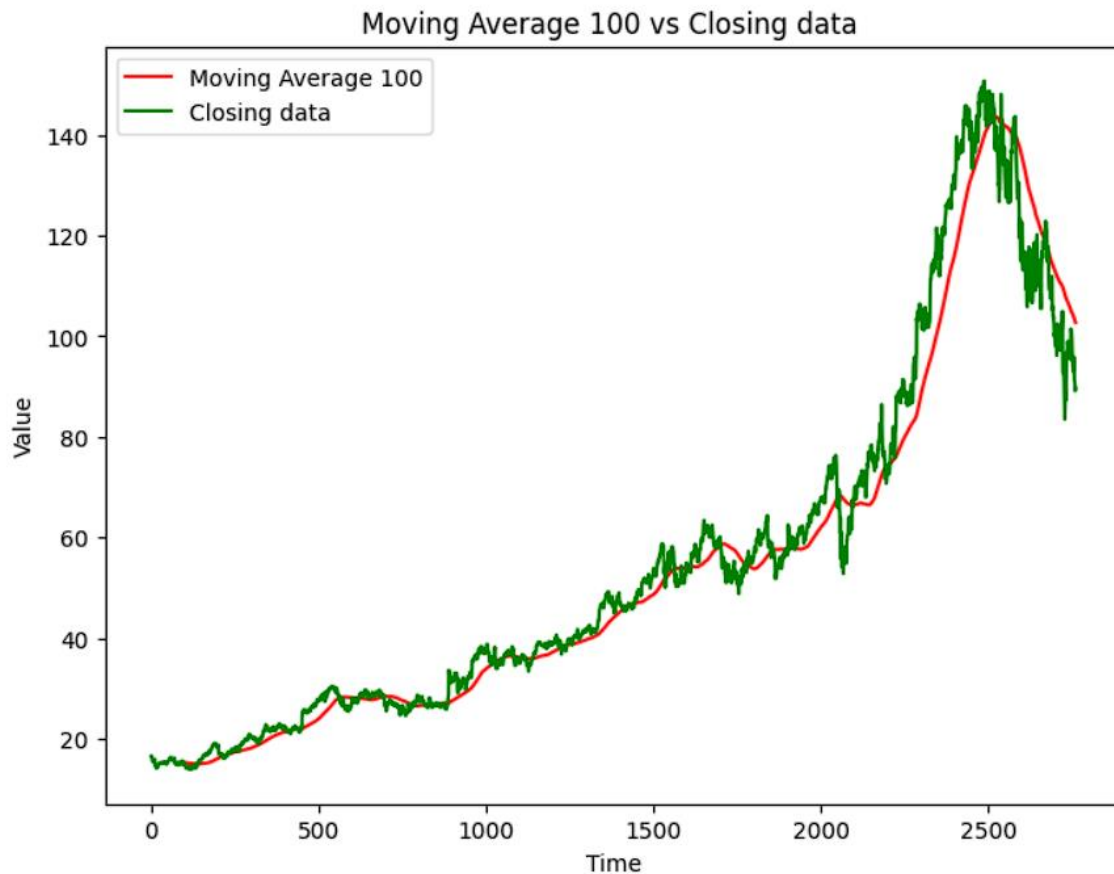
Figure 1: Google Stock Price vs. 100-Day Moving Average

200-Day Moving Average:

Similar to the 100-day moving average, this metric provides a longer-term perspective. The graph below (Figure 2) shows the 100-day moving average in comparison to the 200-day moving average and actual closing prices.
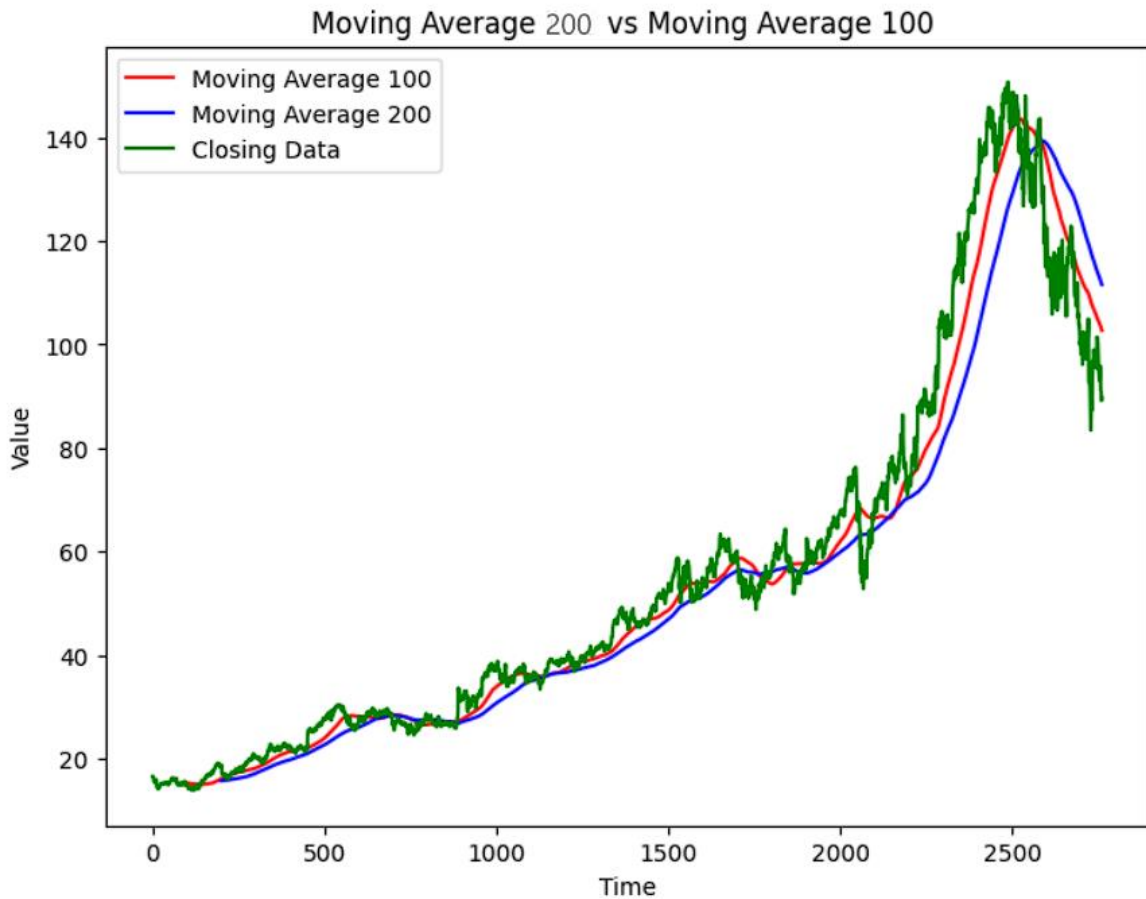
Figure 2: Google Stock Price vs. 100-Day and 200-Day Moving Averages

LSTM Model Performance:

The LSTM model was trained over 50 epochs with a batch size of 32. The architecture included multiple LSTM layers with dropout regularization to prevent overfitting.

The model's performance was evaluated using the testing set, and the predicted prices were plotted against the actual prices.
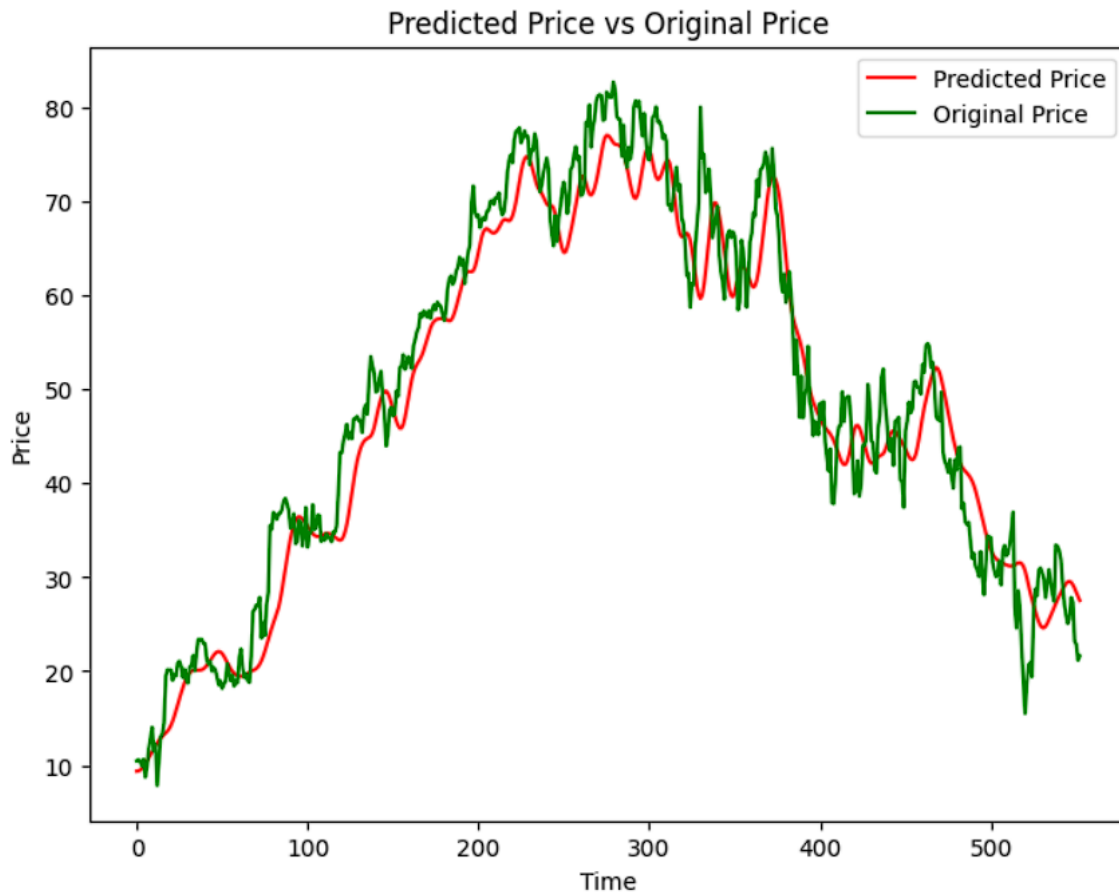
Figure 3: Predicted vs. Actual Google Stock Prices

Discussion

The LSTM model demonstrated a reasonable ability to predict stock prices based on historical data, as evidenced by the alignment between predicted and actual prices in Figure 3. However, several points merit further discussion:

Model Accuracy:

The model captured the general trend of stock prices but showed some deviations, particularly during volatile periods. This indicates a need for model improvement, potentially through tuning hyperparameters or incorporating additional features (e.g., trading volume, market sentiment).

Overfitting and Underfitting:

While dropout layers were used to mitigate overfitting, the model's performance suggests that more sophisticated techniques (e.g., cross-validation, regularization) might be necessary to enhance generalization.

Data Preprocessing:

Proper data preprocessing, including handling missing values and feature scaling, is crucial for model performance. In this study, the use of MinMaxScaler effectively normalized the data, which is essential for LSTM models.

Future Improvements:

Future work could explore the integration of additional data sources, such as financial news or macroeconomic indicators, to improve model accuracy.

Experimenting with different neural network architectures (e.g., GRU, Transformer models) might offer better performance for time series prediction tasks.

Chapter 5

Conclusion and Future Work

Conclusion

This study successfully developed a stock price prediction model using LSTM neural networks, demonstrating its capability to predict stock prices with reasonable accuracy. The model effectively utilized historical stock data, and the results indicated its potential for financial forecasting applications.

Future Work

Enhancing Model Accuracy:

Further tuning of hyperparameters and model architecture adjustments could improve prediction accuracy. Techniques such as grid search or Bayesian optimization might be employed for hyperparameter tuning.
Incorporating Additional Data:

Future models could incorporate a wider range of features, including economic indicators, sentiment analysis from news articles, and trading volumes, to capture more complex market dynamics.
Exploring Advanced Architectures:

Real-Time Prediction and Application:

Developing a real-time prediction system could enhance practical applications, allowing investors and analysts to make timely decisions based on the latest data.

References

Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Systems with Applications, 19(2), 125-132.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

Cao, L. J., & Tay, F. E. H. (2001). Financial forecasting using support vector machines. Neural Computing and Applications, 10(2), 184-192.

Huang, C. J., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10), 2513-2522.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Journal of Financial Economics, 136(2), 547-568.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques - Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932-5941.

Patel, J., Shah, S., & Thakkar, P. (2015). A hybrid intelligent model for stock market forecasting. Procedia Computer Science, 45, 450-459.