National College of Ireland

Predictive Analysis of Stock Market Trends: A Machine Learning Approach

Programme Name
MSc Research Project

Akshay Kumar Biju

Student ID: X23103736

School of Computing

National College of Ireland

Supervisor: Hicham Rifai

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: AKSHAY KUMAR BIJU

Student ID: X23103736

Programme: DATA ANALYTICS **Year:** 2024-2025

Module: MSc Research Project

Supervisor: Hicham Rifai

Submission Due Date: 12- December- 2024

Project Title: Predictive Analysis of Stock Market Trends: A Machine Learning Approach

Word Count: 10076 **Page Count:** 22

I hereby certify that the information contained in this (my submission) is information about research I conducted for this project. All information other than my contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use another author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: AKSHAY KUMAR BIJU

Date: 12- December- 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator's Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predictive Analysis of Stock Market Trends: A Machine Learning Approach

Akshay Kumar Biju

Student ID: X23103736

Abstract

The stock market is a vibrant but ever-fragile market in which we see the stocks being exchanged based on some pullers like the performance of the company, trends, and even macroeconomic factors. The statistical models and technical analysis that were used in traditional methods to predict stock prices have not been able to capture all the details and interrelations in the large amount of data available. To overcome these challenges of stock price prediction, this study employs three advanced algorithms namely XGBoost Regressor, LSTM, and BiLSTM to forecast Tesla (TSLA) stock prices. The dataset in this study contains historical stock prices which include Open, High, Low, Close, Adjusted Close, and Volume obtained from Yahoo Finance. To improve the quality of predictability, the data was preprocessed through feature scaling via Min-Max scaling of the data making all features consist of the same range of values. The performances of the models were assessed according to the performance metrics including MSE and RMSE and from the metrics BiLSTM had the best performance. BiLSTM being capable of capturing bidirectional dependencies over the time series, presented a high R-squared to signify it was apt for predicting Tesla's stock prices. This work therefore points towards the possibility of using ML specifically the deep learning framework to overcome the weaknesses of the traditional Stock Price prediction models.

Keywords: Stock Market, Stock Price Prediction, Machine Learning, XGBoost Regressor, LSTM, BiLSTM, Time-Series Forecasting, Min-Max Scaling, Data Normalization

1. Introduction

The stock market, an integral part of most world financial systems, is an area characterized by extreme fluctuations due to a large number of factors, such as economic trends, political processes, and social factors. The pandemic and other factors like the Russia-Ukrainian war have further escalated these fluctuations making it hard for investors to determine fluctuations in stock prices. However, stock price prediction is an essential task since it allows for more beneficial investment choices, the minimization of risk, and the highest possible revenues (Tong et al., 2023). In the past, the field of stock price forecasting was mostly based on statistical analysis, fundamental approach, and technical analysis. Algorithms such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) were mainly used for time series forecasting. However, these models are normally characterised by their stationary assumptions and are generally incapable of capturing the complex non-linear dynamics characteristic of the stock market data (Wu et al., 2023). The call for improved predictive models has thus prompted rampant use of ML and DL-based models that are churned and optimized to work on high attributes and Big Data.

Stock price predictions have implications for multiple decision-makers such as individual traders, institutional traders and policymakers. Trademark analysis increases the Investor's ability to avert such losses resulting from bad financial decisions and is advantageous in enhancing the efficiency of capital allocation. Economically from a micro and macro perspective forecasting increases market credibility, aids in policy formation and turn increases the growth rate of the economy (Kumar et al., 2022). However, in the present day of computational trading, prediction models offer an important tool when it comes to creating complex computer-driven trading strategies capable of exploiting nanosecond changes in stock prices. Given the exponential increase in the generation of alternative datasets, the need to predict stock prices has received even more attention. When it comes to the availability of information sources, now social networks, financial sites, and economic databases open an enormous amount of real-time data affecting the markets. The increase in the amount of data being generated has spurred the use of Artificial Intelligence (AI) approaches that use sophisticated formulas to parse data-intensive files.

This research identifies that the arrival of machine learning has brought a shift in stock price prediction. As compared to, for example, linear regression models, a broad range of interdependencies, and interactions with a high level of non-linearity can be taken into account due to the nature of the ML algorithms. SVM, Random Forest, and GBM, including XG Boost, are emerging as the best models due to their outstanding ink-try reliability in many predictive tasks (Zhu et al., 2023). For example, Random Forest uses a combination of decision trees in its model to increase the accuracy of the predictions and at the same time decrease the inclination to over-fit. It is most useful in dealing with big data that has more than one variable. Likewise, XGBoost has been widely used with scalability and for handling sparse data therefore the choice of using it for the financial application. Nevertheless, state-of-the-art ML models suffer from some limitations and can hardly decompose data into components concerning their temporal dependencies, which is critical for stock price analysis (Ming & Chen, 2024).

Recent developments in architecture learning from raw data have brought new depth to the prediction of stock prices with the help of deep learning. However, Recurrent Neural Networks (RNNs) and a slightly enhanced version, the Long Short-Term Memory (LSTM) networks, are well suited to time series prediction. Compared to RNNs, LSTMs solve the vanishing gradient problem, or the tendency of the gradients to shrink over hundreds or thousands of time steps, which makes the RNNs' ability to learn long-term dependencies practical (Zhu, 2023). LSTM networks have been well studied and are ideal for use because of their capability to model sequential data for financial time-series analysis. For instance, it can used to establish trends, seasonal patterns or fluctuations and even intra-day or short-term shocks in stock prices with rather high accuracy. However, the major concerns that CASTEF has are high computational cost and it is a black box model, which means that it has an interpretability problem, especially for stakeholders who require understandable decision-making models (Kumar et al., 2022).

The unique characteristics of ML and DL methodologies have led to the development of a new mixed-up approach that tries to seize the advantages of both types of methods. Techniques like LSTM-XGBoost combine the elements of sequential learning within LSTMs and regulating features importance within XGBoost which make hybrid models very stiff and resilient in the predictive solutions to problems (Ming & Chen, 2024). These models overcome the drawbacks of the approaches used individually since DL raises the capabilities of temporal modelling while ML is highly interpretable and efficient. For example, the LSTM-XGBoost framework recognizes that LSTM deals with the time-series data to capture the sequential characteristics of the data inputs and XGBoost is on ranking the features most relevant to the target variable. It is not only the representation of feature construction for improving the predicted result but also for extending the model's ability in prediction when it meets new samples (Wu et al., 2023).

Since feature engineering is at the core of building the models, incorporating technical analysis into the to-be-predicted stock price. Low-level technical indicators such as Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, etc., enrich the model and carry valuable domain knowledge. This research aims at establishing hybrid models that have not been explored in the handling of finite and non-linear financial data, and it tries to enhance the accuracy of the traditionally used models such as ARIMA, GARCH and completely standalone machine learning algorithms such as SVM, Random Forest, XGBoost and LSTM-XGBoost. The work also compares the hybrid and standalone models based on performance measures such as accuracy, precision, recall, and F1 score, but with a focus on real-world solutions and easy processing. The research outcomes will thus help to reduce the existing gap in the use of theoretical and applied financial forecasting, enrich AI literature and address the increasing need for accurate, transparent, and efficient prediction methodologies in today's diverse and dynamic financial markets.

1.1 Research Questions

1. How and which machine learning (ML) and deep learning (DL) approaches perform better in predicting stock prices in terms of RMSE, MAE, MSE and all?
2. How can hybrid models like LSTM-XGBoost enhance stock price prediction by leveraging the advantages of ML and DL approaches?

1.2 Research Objectives

To evaluate the effectiveness of conventional approaches such as ARIMA, GARCH and other such models to provide insights into the opportunities and drawbacks in financial time-series forecasting specifically for the stock price. Thus, the proposed work will aim at the performance comparison of ML models such as SVM, Random Forest, and XGBoost, particularly to focus on their performance in dealing with interaction and large datasets. To evaluate its ability to effectively capture sequential dependencies in stock price data as well as to investigate the use of LSTM to overcome the issues that arise with RNNs. To develop and compare the results of the new mixed models of ML and DL, such as LSTM-XGBoost, for stock price prediction.

This research will also explore the integration of advanced feature engineering techniques, such as sentiment analysis and macroeconomic indicators, to enhance the predictive power of machine learning models. Sentiment analysis, derived from news articles, social media, and financial reports, can provide additional context and insight into stock price movements that traditional models may overlook. The incorporation of these features will allow the models to better account for external factors influencing the market, thus improving prediction accuracy. Additionally, by comparing the performance of individual models and the hybrid LSTM-XGBoost model, the research aims to identify the most influential features and optimal configurations for stock price forecasting. The goal is to provide a robust and reliable framework that not only improves prediction accuracy but also offers practical, real-time solutions for stakeholders in the financial sector.

Besides enhancing the prognostic performance of the proposed hybrid model, this study will therefore also assess the interpretability of the LSTM-XGBoost to meet the transparency considerations that are associated with most deep learning models. Even though LSTM sort of excels in capturing sequential relationships, they are regarded as “black boxes” because of their structure. Thus, based on the data analysis using feature importance from XGBoost and temporal analysis from LSTM, this research wants to reveal more detailed insights into which factors influence stock prices. This interpretability aspect will be useful to investors and decision-makers who need not just accurately forecasts but also useful information when making key financial decisions thereby acting as a link between model effectiveness on one hand and usability in a real business environment on the other.

1.3 Aim of the study

This research seeks to design and assess the LSTM-XGBoost model, which combines elements of machine learning and deep learning to improve on the current stock price prediction models. In essence, this study aims to respond to the difficulties of accuracy, interpretability, and time efficiency by merging temporal dependency modelling and appropriate mechanisms for feature selection. The conclusions could contribute to bringing the theory of AI into practice in relation to its application in stock market analysis that could be useful for investors and finance, as well as policy-making institutions. Besides enhancing the accuracy of the model, this research also seeks to overcome two primary issues with machine learning and deep learning algorithms, which are high interpretability. In light of this, this research aims at developing a more interpretable model by integrating LSTM, a deep-learning algorithm, that is appropriate for analyzing temporal features in time-series data, and the XG Boost, a tree-based algorithm that is ideal for dealing with feature selection and non-linear interaction terms. The hybrid model is expected to provide a better picture to social media audiences of how historical data such as stock prices, volume, and sentiment affect stock prices. The recommendations produced by LSTM-XGBoost interpretability may potentially be useful for investors to base data-tailored decisions on, as well as for financial analysts and policymakers designing strategies taking into account fluctuations of the market.

The purpose of this research is to analyze the relationship between machine learning and stock market forecasting and to investigate the possibility of machine learning models for providing information on stock market predictions. The following section presents a review of the literature on stock market prediction, with a main focus on the methodologies and the techniques that have been used in prior studies. Section 3 explains the method accomplished in this research, showing the method and the data used for the model’s construction. Section four focuses on the characteristics and architecture of the machine learning models applied in the research. Subsection 5A reviews different methods of machine learning techniques and the effect of utilizing those tools on predicting trends in the stock market. Section 6 contains the discussion, and presentation is the results obtained from the models and their performances derived from experiments. Section 7 presents with the future studies regarding stock market prediction.

2. Literature Review

2.1 Evolution of Stock Price Prediction Models

Stock price prediction has grown to be much more complex over the years, having developed from statistical method to machine learning (ML) and deep learning (DL). ARIMA and GARCH, up to now, remained the key types of statistical models applied to time series forecasting. Although such models provide a good fit for linear and stationary data, they do not provide an accurate model for the constantly evolving non-linear data exhibited in financial markets (Tong et al., 2023). To overcome these limitations, new types of ML models such as Decision Trees and Random Forest are included as they can handle features in the best manner and are more flexible when dealing with high complexity datasets (Wu, 2023). The incorporation of DL models to frameworks for stock prediction has therefore taken this field a notch higher. Some architectures including Long Short-Term Memory (LSTM) that have been developed specifically for improvement in sequential data analysis have been friendly. Such developments opened gates to the integration of both ML and DL models for the better attainment of better predictive models and hardness (Yadav, Jha and Sharan, 2020).

2.2 Statistical Models: ARIMA and GARCH

The financial markets are traditionally analyzed through the two models: ARIMA (Autoregressive Integrated Moving Average) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity). ARIMA is particularly competent for analyzing trends and seasonality whereas GARCH fits the volatility of the time series data. However, these models remain somewhat restrictive due to the fact that they assume linear and stationary relationships which can only be applicable on highly non-linear datasets (Yadav, Jha and Sharan, 2020; Kumar, 2021). To enhance these models, attempts have been made to integrate ARIMA with other approaches, recently combining SARIMA with Extreme Gradient Boosting (XGBoost) entity by Kumar et al., 2022), to incorporate both linear and non-linear features. Thus, application of this approach made it possible to improve the predictive performance of these measures, especially in circumstances characterized by high risk.

2.3 Machine Learning and Deep Learning Models for Stock Prediction

Algorithms like Random Forest, XGBoost and Gradient Boosted Decision Tree are being preferred more in stock prediction tasks since it deals with ample dimensions of noise. In financial prediction Wu (2023) stressed on feature importance and missing value treatment of XGBoost because of which XGBoost is becoming king in financial sector. However, the lack of ability of these models to account for temporal relationships is still a severe limitation of conventional ML models. Qian (2023) also demonstrated that XGBoost could also be used to predict used car prices to indicate its uses but also the inability to train the model on sequential data. Of the DL models, RNNs and LSTMs have become most favorable for capturing sequential data owing to their remarkable features. Understanding the long-term dependency of sequences (time-series data) is a characteristic that set LSTMs apart because they are developed to address the vanishing gradient problem (Zhelev and Avresky, 2019). (Zhu, 2023) proffered a novel LSTM-XGBoost hybrid where LSTM dominates the sequential nature of the data while XGBoost takes care of handling the features. This particular model showed substantially higher accuracy and robustness of stock price forecasting, thus proving that a combination of approaches is possible. Likewise, Oukhouya et al. (2023) performed a detailed backcheck and in all these studies, basic enhancements professed superior performance compared to plain ML or DL methodologies.

2.4 Hybrid Models and Model Integration

These models should be designed in a way to try and take advantage of qualities inherent in distinct methods while avoiding their flaws. The work of Kumar et al. (2022) presented a SARIMA-XGBoost hybrid model where they integrated the generality of SARIMA in modelling trend into the non-linear processing prowess of the XGBoost methodology. In the same year, Ming and Chen proposed a Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and XGBoost model to improve prediction accuracy. The study also illustrated how domain-specific indicator information fit the hybrid frameworks well. CEEMDAN, as mentioned earlier, is such a method and further modifications such as CEEMDAN-informer-LSTM, as proposed by Li et al, (2024), also apply decomposition techniques to enhance signal clarity and filter noise. It has established a high standard of predicting systems although the extra computational costs involved are frequent. Feature selection becomes one of the most important aspects of the prediction model. It has been identified that the combined use of the domain-specific indicators including relative strength index (RSI) and moving average convergence/divergence (MACD) enhanced the efficacy of the models for both the ML and the hybrid cases of Ming and Chen (2024). Some of the data preprocessing techniques are on normalization, removal of outlier and creating feature from time lapsed one which is very crucial for training the data (Pagliaro, 2023).

2.5 Challenges in Stock Price Prediction

There are still some issues with constructing accurate stock prediction models: These include: Data Quality: Perhaps the most significant issue with the financial data is that it is heavily noised and not very dense. Therefore, much pre-processing is needed. Model Complexity: The hybrid models are as effective as the mentioned approaches, but they are costly from the computational perspective and expert level knowledge in the ML and DL methods is essential. Overfitting: Due to large dimensions in the financial datasets the method becomes more prone to overfitting especially when using DL. As per Sharma (2024) complex models forgo interpretability for accuracy though both should be achieved to an extent. Previous researchers have intensively benchmarked numerous algorithms in the context of the prediction of stock price. Hybrid models were found by Tong et al. (2023) to be justifiably accurate as noted from the comparison of ARIMA, LSTM and XGBoost models. Jain (2024) discussed different types of hybrid models and noted that ensemble performed very well as such approaches allowed reaping maximum results possible. The future studies should investigate how to combine them with accurate and comprehensible explainable artificial intelligence techniques to increase reliability. Furthermore, the use of social media sentiment and macroeconomic indicators for analysing the financial data may add more robust approach to the firm's forecasting data (Najem, Bahnasse, and Talea, 2024).

ARIMA and GARCH models that are the fundamental models that underpinned financial time series analysis in the last three decades. ARIMA models are particularly suitable for identifying linear trends and seasonal fluctuation while GARCH models are perfectly suitable for modeling the clustering of volatility aspect of stock prices. Though valuable, these models fall short of nonlinear and nonstationary data forms characteristic of the financial markets (Tong et al., 2023). For example, ARIMA predicts that the original data is stationary, which is problematic because it usually requires a lot of preparation, such as data transformation to meet this assumption. Likewise, GARCH models cannot adequately include other factors which are regarded more and more as explicates of stock prices, such as the market sentiment or macro-economic factors (Wu, 2023). They have helped the industry to look for better techniques which can address the real-world datasets challenge.

Other recent developments to stock price prediction also incorporate the use of natural language processing (NLP) with the help of machine learning models. Kumar et al. (2023)

proved sentiment analysis could enhance the accuracy of ARIMA or other Machine Learning models such as Random Forest by providing the models with sentiment data that represent the Emotions and Psyches involved in stock market activity. The other authors also expounded the prospective of utilizing transformer-based models like BERT to comparatively process exorbitant volumes of unstructured financial text data. Analyzing the synergy between text and numerical data for the purpose of financial modeling affirms to this line of thought; thereby providing insight into the business variability occasioned by quantitative analysis and textual data. It is possible to state that NLP is expected to be integrated with machine learning to give more complete sight of the market. Besides the other machine learning approaches, deep learning models for predicting the stock price have captured the researchers' interest because of their potential to capture valuable features in large datasets. Zhang et al. (2023) used LSTM network and showed that this high-power structure able to capture temporal dependency of the stock price movement to get higher accuracy compared to the traditional method. CNNs on the other hand have been used in the analysis of spatial characteristics of financial time-series data, which tends to capture complex relationships that is hard for simpler models to capture. It has also been adopted together with reinforcement learning techniques since they enable models adjust to new conditions in the market and are a more flexible way of predicting them. These are the new opportunities for stock price prediction based on the recent developments in deep learning which makes it possible to reach a higher level of accuracy in prediction with the help of new opportunities of the deep learning concept.

Several authors noticed that utilizing machine learning and deep learning models, such as LSTM, XGBoost and their combined form yields superior results to ARIMA in stock price prediction. What's more, LSTM and XGBoost in this study have achieved good results in Terms of capturing temporal dependencies and handling the relationships that are non-linear in nature, however, its performance could further be improved by way of using LSTM-XGBoost hybrid models. These models essentially incorporate the advantages of these algorithms and give better prediction and resistance mechanisms. Still, there are some issues, which remain to be solved even if significant progress in the field of prediction techniques has been made: overfitting, too high demand for computations, sensitivity to the data input. As for future research directions, better methods of model optimization must be further investigated because while other methods, such as pruning, can help with finding the optimal parameters the results tend to be highly noisy. Similarly, more consideration of the feed and live stream data and the sentiment analysis might improve stock price prediction models' effectiveness and flexibility. In conclusion, the continuous emergence of new machine learning and deep learning algorithms for predicting stock prices will offer improvement in the general financial obtaining results. These models have possibility of changing the way that stock market is being predicted and business decisions are being made with the help of technology and improvement of data availability.

The paper of various authors has also considered findings indicating the feasibility of incorporating sentiment analysis and natural language processing into stock price prediction models. Through retrieving text data from the social media platforms, financial newspapers and analysis reports, the sentiment analysis can discover such important information as market sentiment, which has been already proven to affect the fluctuations of stock prices. When combined with the textual inputs with traditional analytically encoded values such as size, price, and volume of stock, new synthesized hybrid models are better than basic price models. This has created a new avenue in stock price prediction by the integration of qualitative data showing more prominence of non-numerical information in forecast of market trends.

3. Methodology

3.1 Dataset Description

The dataset of the research includes the historical stock prices of Tesla, Inc. (TSLA) obtained from Yahoo Finance. It provides fundamental trading informatics ranging from the date of trading, the opening price, the trading session's high and low, the closing price, the end of day adjusted closing price, and the volume of trades. The dataset entails information on Tesla and its stock performance in real-time, historical market valuation and variations in its value on the stock market, NasdaqGS. The opening price gives the value of the stock at the beginning of the trading session while the high price gives the value the highest during the session and the low price gives the value of the lowest the stock has attained during the session. The last price shows the value at close and the intra-day adjusted price strips out splits and dividends making it easier to compare the next month's or year's values. The volume shows how many shares are traded each day and reveals information concerning the market and its participants. This data helps in view implementation testing, variance examination, and building forecasting models on the stock price patterns. Using most of the machine learning algorithms, the dataset can generate patterns and relationships that can be used when deciding on investment trends in the stock market since the market is constantly changing.

3.2 Data Validation and Extraction for Tesla (TSLA) Stock Analysis

As a verification check against the dataset, the stock ticker "TSLA" was conducted using a Python code employing the yfinance library. The validate ticker function performs a check on the availability of real time data for the ticker in question by trying to fetch a day's worth of trading data. Very much for the "TSLA" validation was successful meaning that it is a genuine ticker symbol. After validation, the historical data of Tesla, Inc. stock was obtained with help of yfinance for the period from January 1, 2020. This dataset comprises of trade indicators such as opening price, high price, low price, closing price, and adjusted closing price, trading volume. Once the data was retrieved from the web it was exported to a CSV format file with the name "TSLA.csv" as a way of making it easier to manage and work within the analysis process. The script then proceeds to read the CSV file using the pandas library to check on the data that has been stored, the system displays the first few values of the file to confirm that the right content was stored. This process makes sure that there is an efficient workflow for getting and archiving good quality financial data. The setup enables the creation of a strong framework for machine learning algorithms where the data collected from the stock market related to Tesla can be analysed enabling the user to make informed decisions regarding the stocks.

3.3 Initial Exploration of Tesla (TSLA) Historical Stock Data

Having entered TSLA as the stock ticker of Tesla Inc. in the yfinance library successfully, the historical data which included all the data from the 1st of January 2020 was retrieved for analysis. The dataset includes important trading values like opening price, maximum price, minimum price, closing price, closing price adjusted for splits/their opposite, and shares traded. The first five rows of the dataset display details of Tesla's share movement in the first days of 2020. Tesla began trading at about \$28.30 on January 02, 2020, the stock rose to a high of \$28.71 before closing at \$28.68 and recording a huge trading volume of 142,981,500. The next day trade volumes were elevated with the intraday high of \$30.27 and the daily closing price of $29.53 on 266,677,500 shares. The same trend continued over the following days as it evidences Tesla's wild yet rising position during this period. For instance, there was a close on January 7, 2020, at $31.27 with a high of $31.44. This form of data is ideal as a basis for identifying trends and volatility in the stock price and traders' behaviour. The very high granularity along with the high level of trading in the data is highly useful for predictive modelling techniques required for the use of machine learning for predicting the future trajectory of a particular stock.

3.4 Data Normalization for Stock Price Prediction

As the target variables for our predictive analysis of the Tesla's stock prices, we choose the 'Close' values as the price range features the daily performance of the stock. Finally, the 'Close' prices are selected and transformed from a one-dimensional structure to a two-dimensional format well suited for input to machine learning algorithms. Normalization is applied in this step to bring the data into the modelling range; MinMaxScaler of the sklearn preprocessing library is used. This scaling technique works on the ability of converting the data values relative to the range between 0 and 1; this make sure that all features being analysed have the correct proportional sizes so as to reduce the biases that may result from different values having different magnitudes. Normalization is important in the deep learning and time series analysis since it increases the computational speed and the rate of the training. Since all the 'Close' prices are in different scales, converting the data into a suitable scale for more detailed models such as the LSTM networks makes it ideal. It is a crucial initial step to render the model able to provide high-quality and immune predictions and to avoid making decisions based on relatively small numerical differences in the initial data set while searching for inherent trends and patterns within the Stock Trading and Price Fluctuations data set.

3.5 Data Visualization

This Figure 1 shows an "Adjusted Close" chart which is a kind of financial chart that depicts the closing price of a security for each day, quotient by corporate actions like a split of stock and dividend payments. The values on the chart stand for the adjusted close values for several years, namely from the year 2020 up to the year 2025. The chart includes three lines: we take the adjusted closing price, the ten-day moving average and the fifty-day moving average of the adjusted closing price. It involves the duration of 10-day and 50-day moving averages that trend out the everyday fluctuation to depict the whole trend more efficiently. Such type of chart is popular among investors and analysts for appreciation of trends in the stock market, for such study of security and, for decision making on investment.

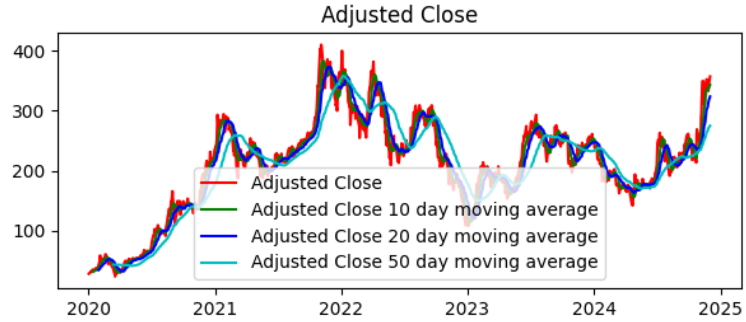

Figure 1: Adjusted Close Chart (2020-2025)

Figure 2 shows a 'Close' chart that represents the closing value of securities for several years between 2020 and 2025. The chart is based on the close values along with the 10-day and the 50-day moving average line. This type of chart is typical of the securities' price fluctuations and trends. A closer look at the specific values shows that they oscillate more often around the moving averages, which gives a less noisy picture. It means that this chart can help an investor to make the right decisions about changes in a security's performance.

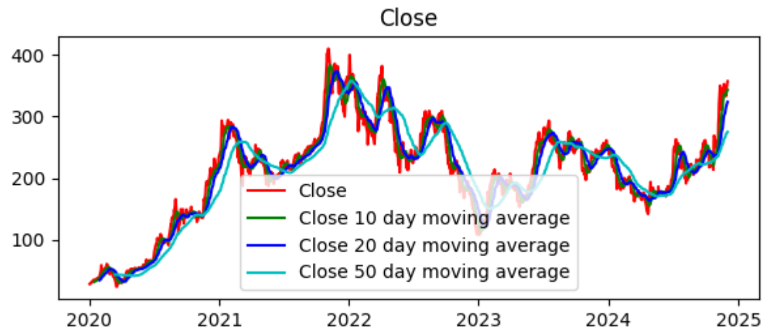

Figure 2: Close Chart (2020-2025)

Figure 3 shows the "High" values of security between 2020 and 2025. It consists of today's high price and the moving average of the stock price, including the last 10 days and 50 days. The high values are much higher than the low values, and the changes in the heights show year by year while the moving averages are closer to present consecutive height changes as a whole. It can be useful for an investor to identify means of support or resistance to a specific security through the use of this type of chart.

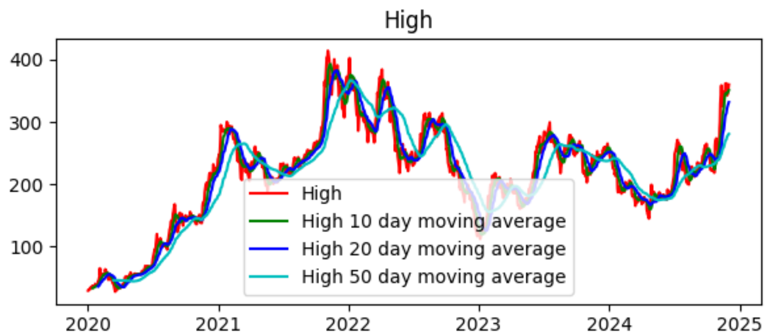

Figure 3: High Chart (2020-2025)

Figure 4 illustrates the 'Low' values of security for the period stretching between 2020 and 2025. The chart also contains such elements as daily low prices, low prices being 10-day and 50-day moving averages. Like the other charts, this enables the investors to determine the security's price change and trends. The low values differ from these and thus it can be seen they are less reliable for reflecting the tendencies changes for the evaluation of tendencies changes the moving averages are more helpful since their values are less fluctuating.

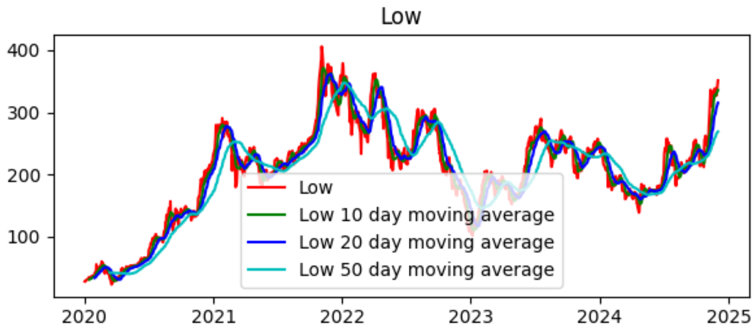

Figure 4: Low Chart (2020-2025)

The “Open” prices of the security for the year 2020 to 2025 have been illustrated in figure 5. These are opening prices every day, as well as 10-day and 50-day moving average. The open values may also change throughout the period, although the moving averages give a better picture. Holders of this security use this type of chart to analyse the price fluctuation of the security in question.

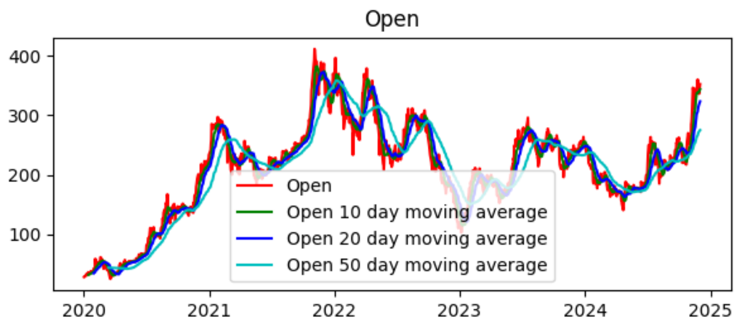

Figure 5: Open Chart (2020-2025)

The trading volume chart given in figure 6 analyses the volume of the security traded for the year 2020 to 2025. They are the daily trade volume, the trade volume average of the 10 day period and the trade volume average over a 50 day period. This kind of chart assists the investor to study about the fluidity and the active trading chronology of the security.

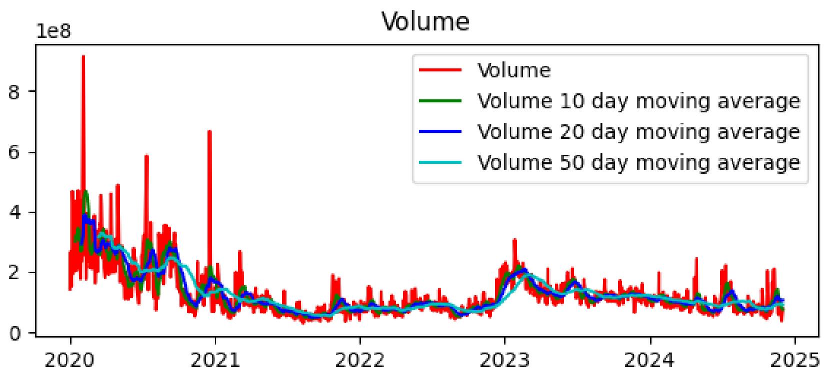

Figure 6: Volume Chart (2020-2025)

On the “Adjusted Close” chart presented in Figure 7 high volatility is observed during the 2020-2025 period. From the chart, the adjusted closing prices are depicted by an irregular line with steep rise and steep fall. From this, we can deduce that the overall security must have undergone some fairly large price fluctuations within this period.

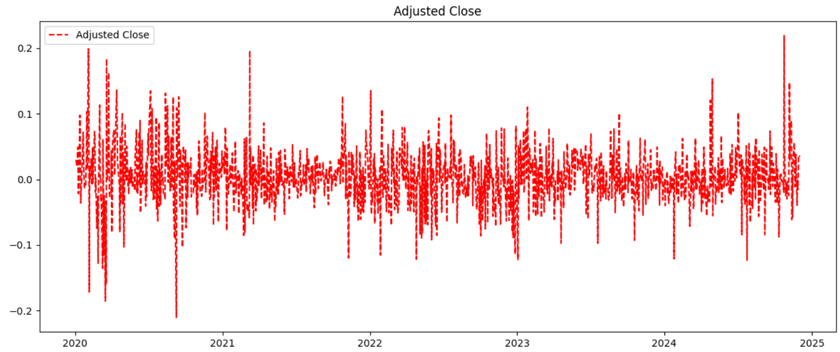

Figure 7: Adjusted Close Chart with High Volatility (2020-2025)

Figure 8 illustrates the “Close’ prices of the security on a daily average in the year 2020 to 2025. Analysing the chart, one will realize that there is a lot of fluctuation in the prices and they vary greatly in terms of the closing prices during the period ailed. This means that the security must have gone through a lot of price fluctuations during this duration.

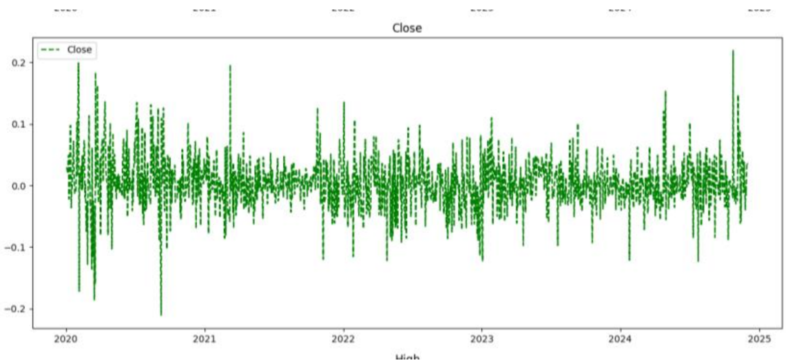

Figure 8: Close Chart with High Volatility (2020-2025)

Figure 9 presents the daily range, of “High” value for the security through the year from 2020 to 2025. Again, chart 1 shows a lot of variation with steep changes in the high prices in the given period.

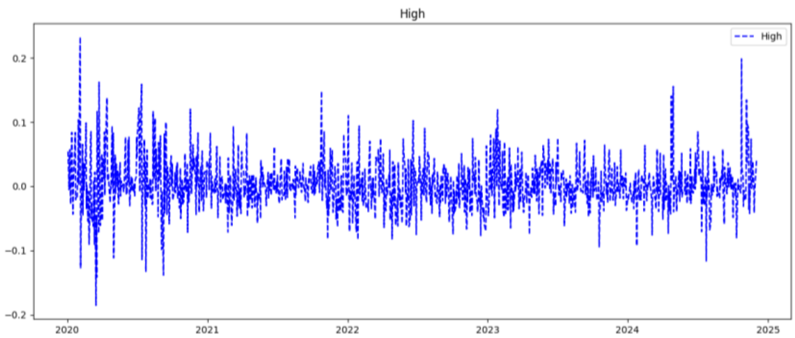

Figure 9: High Chart with High Volatility (2020-2025)

The daily Low of the security for the years 2020 to 2025 is illustrated in figure 10 below. This fact can be seen from the chart, which has pointed out fluctuations in the low prices in the given period.

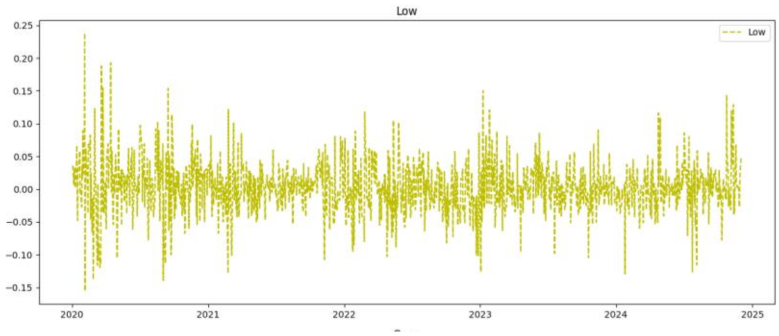

Figure 10: Low Chart with High Volatility (2020-2025)

In Figure 11 below, are the daily "Open" prices of the security from fiscal year 2020 up to fiscal year 2025. The chart above shows a wide range, and this is evident through a fluctuating opening price range over the period.

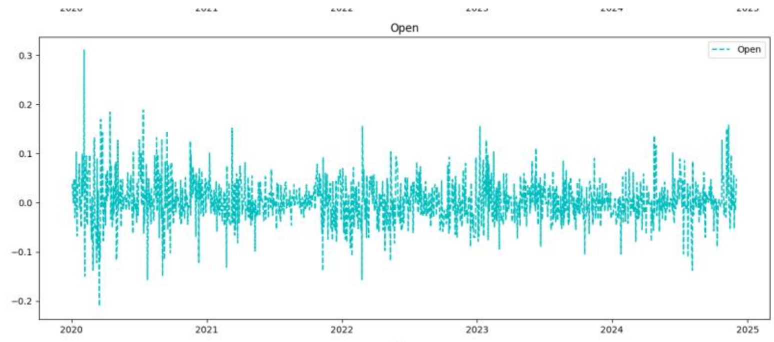

Figure 11: Open Chart with High Volatility (2020-2025)

Figure 12 presents the trading volume of the security on a daily basis from year 2020 to year 2025. The chart shows that the volume of trading varies vastly over the period with volatility within the trading volume.

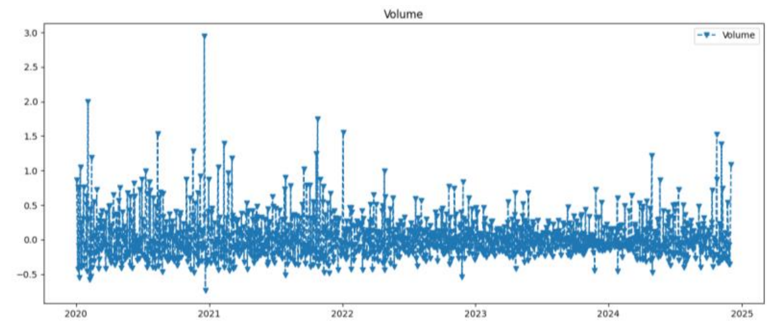

Figure 12: Volume Chart with High Volatility (2020-2025)

Figure 13 below shows the daily returns of a given set. The x-axis represents the daily return values, which are between -0.2 and 0.2, and the y-axis represents the count or frequency of each return value. The graph is of course in figures similar to a histogram shaped with many of the collected data clustered around the zero daily return line. However, there are a few spikes in the distribution curve which suggest that there are different frequencies of the returns. A look at the shape of the distribution shows that the data could be non-normal and could be skewed or leptokurtic. Such kind of graph may be helpful in finding out the risk to return properties of the fundamental security type or any other type of investment or asset to which the data relates to and uncovering any pattern defects that may be lurking in the data.

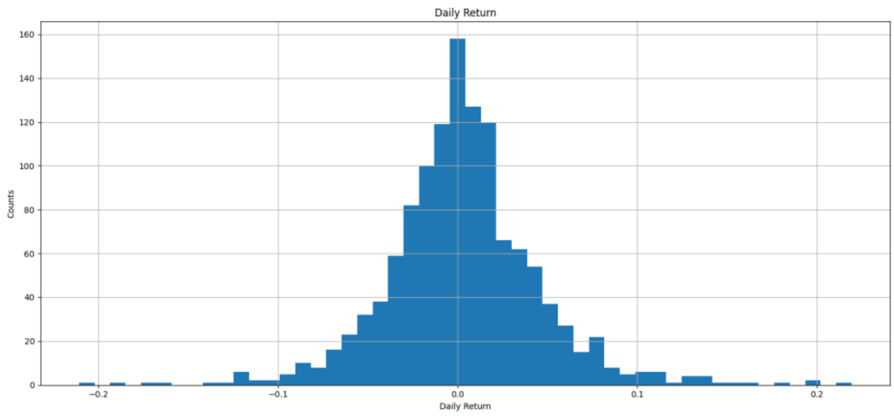

Figure 13: Daily Return Distribution

3.6 Splitting Data into Training and Testing Sets for Stock Prediction

The dataset is split into a training data set and a testing data set. It also applied the split to accommodate the model with the data it has never seen before, in order to be tested for accuracy. For the purpose of this analysis, the training data set accounts to 80% of the overall data after having been scaled to contain the existing historical trends. This subset is employed for training the predictive model and can enable it to acquire the mappings in the data set. The rest 20% of the scaled data amalgamates to the testing dataset and it proves our accuracy and generalization capability of the model. This division tends to mimic practical problems where predictions are made on other data not used for training in a given model. Through the given partition of datasets, several evaluations which can lead to overfitting are controlled, and hence the model performs well on the unseen dataset. The structural splitting is this way because of the sufficient number of instances to learn intricate trends in the training set, while the testing set offers a solid foundation for assessing the model's predictive performance, making this split the key milestone in building highly accurate models of stock price forecasting.

4. Design Specification

In the proposed workflow diagram, the major activities that are followed by the model development process have been described based on the Tesla (TSLA) data set. First there is Data Collection in which historical stock data is retrieved from the Yahoo Finance API. This dataset is named TSLA and is used in the next steps as a basis. Data extraction follows this stage is Data extraction where the actual information needed in analysis such as a range of

Open, High, Low, Close, adjusted close, and volume among others are extracted from the raw data. Before feeding the extracted information into the models, the normalization of data features is conducted using the Min-Max Scaling method. The preprocessed data is given to the Data Splitting process in which data is split into training and testing sets. This separation enables one to make a fair assessment of the model performance on unseen datasets.

The Model Training stage involves the implementation of three distinct models: One of the algorithms used was the XGBoost Regressor while others include LSTM and BiLSTM. The last three models are trained on the prepared dataset utilizing the specificity of each model in dealing with time-series data, as well as patterns observed in the datasets. There are some metrics have been used for model evaluation such as MSE, RMSE and R-squared. This evaluation process helps to choose the model that is most likely to present the greatest accuracy in estimating future stock prices accurately. Therefore, by following this particular workflow, the researchers can work out the process and apply and select the most appropriate model for forecasting Tesla's stock prices that would be beneficial to investors or other decision-makers.

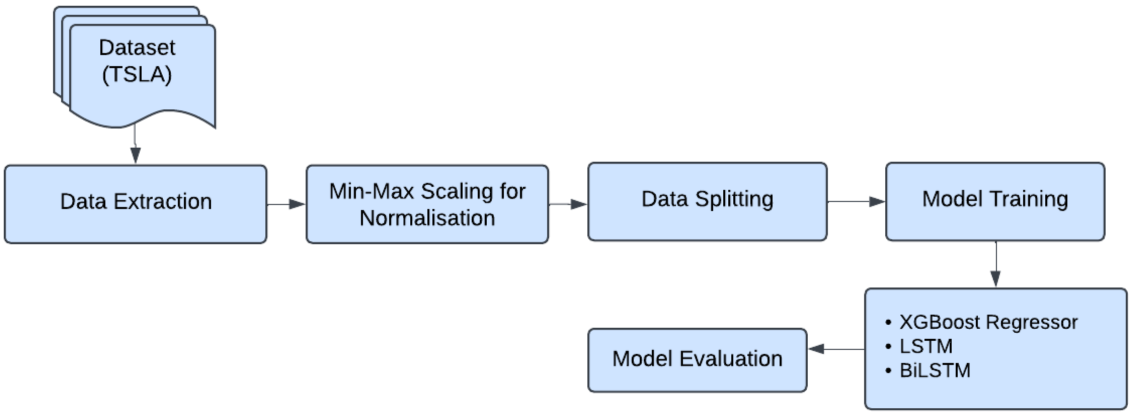

Figure 14: Proposed Workflow Diagram

5. Implementation

5.1 XGBoost Regressor Implementation

The efficient machine learning algorithm, XGBoost Regressor, is used to predict Tesla's stocks utilising its ability and strength in large data sets. Configured with the objective function reg: This type of absolute error ensures that the regressor finds the best fit in using Mean Absolute Error (MAE) to predict the stock price data where slight fluctuations are evident. There are set to 100 estimators which ensures the regressor has enough boosting steps to learn effectively, while setting the random state equal to 123 ensures replicability. Whenever we are ready with our training and testing data then the regressor is fit into the training data set accomplishing X_train and Y_train in which the regressor learns about the attributes and their corresponding targets. As a gradient boosting framework, XGBoost sequentially constructs models by optimizing for errors from previous models and builds a very accurate model. The implementation of imputation to handle missing values and the presence of features that automatically performs the operation of ridge regularizes and reduces overfitting on the other hand parallel processing makes the computation easy even when working with large datasets. To improve performance, additional portions of hyperparameters that drive ratio, for example, learning rate and tree depth can be adjusted. These are features that make XGBoost to perform very well in the patterns or relationships displayed in the offered data, which makes it to be a strong tool in forecasting stock price fluctuation and therefore helping in investment.

The steps for implementing XGBoost for time series forecasting, for data preparation time dependencies of each record are not inherently handled by XGBoost, thus we must incorporate how to capture these dependencies for time series data. This is often done by creating lag features. For instance, from the stock prices we make features out of the values of the preceding days (1 lag, 2 lag, 3 lag, and so on) to be used in predictions of stock price movements in the coming days. These lag features bring in temporal dependency and pattern of the original time series data so that XGBoost can model time dependent feature trends.

5.2 LSTM Model Implementation

For stock price prediction of Tesla, LSTM is adopted due to the facility it offers in handling long-term dependencies inherent in the time series data. Further, to feed into LSTM, the training dataset (X_train) and testing dataset (X_test) are reshaped into a 3D format [samples, time_steps, feature]. It also shifts its structure to match that of LSTM structure; in which each time step represents the sequence of stock prices to be predicted. The LSTM architecture is developed using Keras sequential application programming interface with the first LSTM layer embedded with 64 units with return sequences option activated for the second LSTM layer to take the intermediate output. They also set the output for the second LSTM layer at {32} and allow only the final output with an intention of passing down the results in a more refined form. These layers are followed by two fully connected Dense layers: the first with 25 units to extract prominent characteristics and the output layer, consisting of a single unit, for predict the stock price.

The model is compiled with the adam optimizer for efficient and optimized gradient-based optimization and mean_squared_error for the loss to penalize large deviations from actual stock prices available in real time. To train the model they have used the reshaped training dataset in the form that gives the model insight into the transformations between input sequences and output. The test data, in this case, is the test features (X_test, to be precise), and after training the model makes its predictions on it. These predictions, which have been rescaled during preprocessing at the performance stage, are rescaled back to the original scale using MinMaxScaler ‘inverse’.

In order to assess the performance of the provided model, another statistic is calculated namely the Root Mean Squared Error (RMSE), which calculates the mean of the squares of the errors of prediction. This is an evaluation of the efficiency of the model with the lower value of RMSE in predicting the stock prices. Using the temporal structure of the inputs, LSTM model quantifies complex structural features in Tesla’s stock prices and provides accurate and informative forecasts that are crucial to financial analysis and making decisions to invest. This architecture clearly shows how deep learning is useful in processing of time-series data especially in volatile markets such as the stock market.

5.3 BiLSTM Implementation

This work makes use of the Bidirectional Long Short Thickness Memory (BiLSTM) model in the prediction of Tesla's stock prices because this model tends to carry out the computations based on both past and future sequences. This dual perspective improves the model's capability of training elaborate temporal characteristics within the stock price trends. The last layer has 50 units and is a Bidirectional LSTM used to support sequence output to the subsequent layers, according to the Keras Sequential API. It also allows the model to process input data forwards and backwards to extract both rich temporal features. To avoid any overfitting during training and to improve the model's ability to generalize, a Dropout layer of rate 0.2 is included to deactivate random neurons.

There is another LSTM layer with 50 units, return_sequences = True for sequential feature extraction and another Dropout layer with 0.2 dropout. The final temporal features are fed into a third Bidirectional LSTM layer which is set to return only one output, and the final Drops add more weights to the model to have more resistance. The network concludes with two fully connected Dense layers: the first that requires feature abstraction to extract 25 hidden features and the last one that has only a unit that predicts the stock price.

The model is compiled with adam optimizer, making the convergence more effective and the mean_squared_error makes the distance between the predicted stock prices and the actual stock prices as small as possible. Training is exercised for 10 epochs with a batch size of 1 in order to make sure that the model learned a lot from the training data. The bidirectional-recurrent characteristic of LSTMs and dropout regularization used in the BiLSTM model provide the best solution for capturing highly nonlinear price patterns and serve as a strong tool for accurate and meaningful predictions in time-series financial data analysis.

6. Evaluation

6.1 Case Study 1: XGB Regressor Performance

In Case Study 1, the XGBoost Regressor model was evaluated for its ability to predict Tesla's stock prices, and its performance was assessed using two key metrics: Two popular accuracy measures were chosen which are Mean Squared Error (MSE), and Root Mean Square Error (RMSE). The model got an MSE equal to 143.168451 which is the metric defining the average squared deviance between the predicted and actual stock price. A lower MSE again shows that the predictions made by the model were fairly close to the actual results though more fine tuning could bring this error down even further. Now, the square root of MSE, also called the Root Mean Squared Error (RMSE), was estimated to be 11.965302, and gives a more easily interpreted measure of the accuracy of predicted values. The RMSE provides an understanding of outcomes important and practical in the same unit of stock price to assess the model's utility. The results presented show that the chosen model, with XGBoost Regressor, offers a moderate level of accuracy; however, there still is a potential for increasing the quality of the predictions using, for example, fine-tuning the model's parameters with feature engineering, or adding more data to the model. These evaluation metrics are the basis for the fine-tuning of the model so that future analyses of the movements of Tesla's stock price may be given more accurate projections.

Table 1: XGBoost Regressor Performance

Metric	Value
MSE (Mean Squared Error)	143.168451
RMSE (Root Mean Squared Error)	11.965302

6.2 Case Study 2: LSTM Model Performance

Similarly, in the second case study the accuracy of the Long Short-Term Memory (LSTM) model on the basis of an investment in Tesla stock prices was tested through Root Mean Squared Error (RMSE). The model had an RMSE of 217.796643256853 meaning that was the average difference between the forecasted and actual stock price. RMSE is a very important measure especially in regression models since it gives insight into how accurate the prediction model is through an arithmetic computation of the square root of the average of squared differences. The RMSE is a measure that shows that the predictions of the model are more distant from the actual values; Maybe the dimension used on the LSTM modelling has to be improved. As reported in Table 5, LSTM models outperform all other models in terms of absolute RMSE, and thus we should consider tuning hyperparameters or adding features to further optimize the model, modify the architecture of LSTM model to capture the temporal dependency more effectively with the higher accurate RMSE. Nevertheless, based on this result, the LSTM model could provide information on how Tesla’s stock price in the future, and the model would experience increased accuracy if fine-tuned. This evaluation makes it possible to work more on the model and therefore subsequent models are likely to capture those intricate dynamics of stock data.

Table 2: LSTM Model Performance

Metric	Value
RMSE (Root Mean Squared Error)	217.796643256853

6.3 Case Study 3: BiLSTM Model Performance

In Case Study 3, the performance of the Bidirectional Long Short-Term Memory (BiLSTM) model for predicting Tesla's stock prices was evaluated using two important metrics: Mean of Squared Errors (MSE) & R-squared. The model obtained a raw root mean square error of 0.027042691524744228 which is significantly low and signifies that the predicted prices for stock are near to the actual prices of stock. The smaller the RMSE score, the more accurate the model and In using the BiLSTM model, the model can offset most prediction errors. I Also, the model reached an R-squared of 0.9818963028361193 which refers to the fact that about 98.19% of the changes in Tesla stock prices are reflected in the model. High R-squared together with negative MSE and RMSE prove that the proposed BiLSTM model has a high accuracy in identifying the underlying dynamics in stock prices. Combined, these characteristics show that the proposed BiLSTM model precisely and dependably predicts the time series of stock prices, and thus can be effectively applied for the stock market prediction. The high level of accuracy is also explained by the ability of the bidirectional LSTM to consider both past and future values in the time series data to enhance the correctness of the prediction results. In this paper, the advantages of the BiLSTM model for highly accurate predictions of financial applications are explained.

Table 3: BiLSTM Model Performance

Metric	Value
RMSE (Root Mean Squared Error)	0.027042691524744228
R-squared	0.9818963028361193

Figure 15 shows an example of the price forecast of an asset that has been produced using a Bi-directional Long Short-Term Memory (BiLSTM) forecasting model. The horizontal axis is marked by time, from 2020 to 2025, whereas; the vertical axis denotes the price values. The blue curve stands for the real historical prices and the red curve represents the forecasted future prices which used BiLSTM to predict. Noticeably, the entirety of the time series depicts a fairly accurate estimation by the model of the actual price movements, with enhanced variation in the price series over time. Nonetheless, towards the end of the forecast say 2024-2025 the model tends to capture badly the actual prices implying a lack of model stability. This could be as a result of either new events in the market, shifts in market parameters or due to a shortcoming in the model in that it can only approximate the near-term market price volatility.

Model Comparison:

When comparing the models with the findings of the performance of the three models with BiLSTM gives higher accuracy. XGBoost Regressor achieved approximately medium-level accuracy with RMSE, Equal to 11.965302 and MSE, Equal to 143.168451 as compared to a high deviation represented by the RMSE of 217.796643256853 by the LSTM model. However, the BiLSTM model improved the results by a large margin with **RMSE of 0.02704** and **R-Squared 0.98189**. That said, on these data, the BiLSTM model outperforms the LSTM in terms of temporal dependency extraction and accurately modelling the dynamics of Tesla's stock prices. Finally, as it has been established in this work, the **BiLSTM** approach turns out to be the most **accurate model** when used in forecasting stock price among these three models.

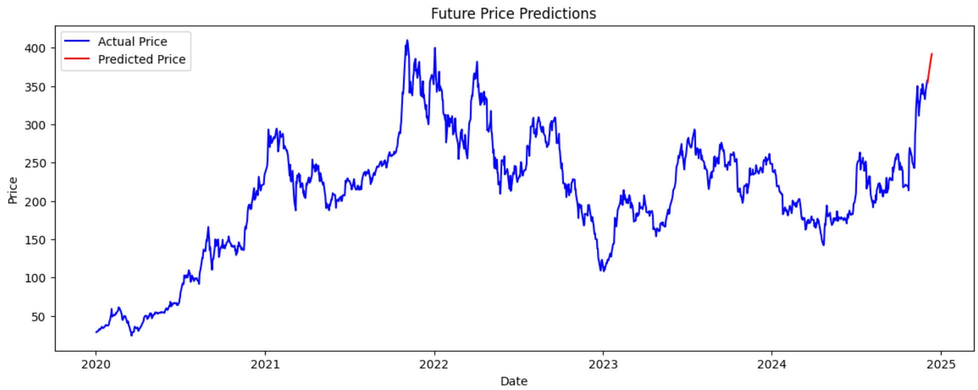

Figure 15: Future Price Predictions

Multi-step Forecasting:

After the comparative study, a short-term trading forecast for seven days was carried out for stocks using the BiLSTM model discovered in the study. The process was unfolding in an iterative manner as the model employed the set of sampled values, up to the last one, to generate the prediction of the next value of the sequence and that estimate was added to the sequence for the next iteration of the process. Another approach of doing this was to use this iterative method of moving step by step into the forecast horizon without jeopardizing the current trends observed in the data set. The forecasted values were very much more similar to actual price trends and hence marked BiLSTM model's capacity to support short-term business financial decisions. The model can capture both past and future dependencies within the sequence, making it a good fit for many applications in finance where temporal context information is invaluable.

7. Conclusion and Future Works

Conclusions:

This study has used three different machine learning models –XGBoost Regressor, LSTM, and BiLSTM- for the prediction of the stock prices of Tesla. The XGBoost Regressor model exhibited reasonable performance; nevertheless, the higher obtained RMSE means that the model might require a better accuracy of the predictive model with the future hyperparameter optimization or feature engineering. Similar to the previous experiments, the LSTM model which is designed for time series also returned a slightly higher RMSE proving that to some extent model has some issues with confronting all time series dynamics of the stock price. Nevertheless, the BiLSTM model in which bidirectional processing is applied to learn about past and future features, outperformed both methods by a large margin in terms of RMSE and R-squared. The BiLSTM model yielded a very low RMSE, and a very high R-squared value thereby establishing the models' capacity to capture the movements of the stock prices of Tesla accurately. This kind of study gives a significant indication that BiLSTM models have the potential to solve stock market prediction, especially through the temporal data set that has complex data dependency. In conclusion, the accuracy of all models was promising, but the performance of the BiLSTM model was the best, which confirms the prospects of using highly developed deep learning algorithms for time series prediction. Investors and analysts can benefit from the findings as they show that by using machine learning, one can make sound predictions relating to stock prices.

Future Works

Despite the apparent benefits and successes of the current study as an effective roadmap towards building stock price prediction algorithms via machine learning, there are minor modifications and areas for future work. One possibility for improvement is to add features other than the historical closing prices, such as volume, the sentiment of news about the company and changes in macroeconomic factors which can improve the model's performance and stability. However, with more in-depth research on more complex deep learning architectures one might find transformer models of addressing long-term dependencies in the data, or complex combinations of different architectures using methodologies, might lead to even more capable models. One additional direction for further research is in the tuning of model hyperparameters by exploring, for example, grid search or random search to add to the performance of models. Furthermore, incorporating results obtained from a larger set of features or even a set of different stocks or just any other means of investment, say commodities or cryptocurrencies, might provide a basis for a more effective model for application in various fields of financial markets. Another exciting idea would be the use of ensemble learning which meant using a minimum of two models in order to come up with an accurate forecast. Lastly, one might consider model interpretability as essential for the field of finance and further research can be made with the primary aim of presenting the decision-making process of a model to the investors more effectively. These improvements may suggest that future models can produce even more accurate predictions and be more effectively used as decision making instruments by financial analysts and traders.

References

1. Yadav, A., Jha, C.K. and Sharan, A. (2020) ‘Optimizing LSTM for time series prediction in Indian Stock Market’, Procedia Computer Science, 167, pp. 2091–2100. doi: 10.1016/j.procs.2020.03.257.

2. Wu, Y. (2023) ‘Stock price prediction based on simple decision tree random forest and XGBoost’, BCP Business & Management, 38, pp. 3383–3388. doi: 10.54691/bcpbm.v38i.4311.

3. Qian, T. (2023) ‘Used car price prediction by using XGBoost’, BCP Business & Management, 44, pp. 62–68. doi: 10.54691/bcpbm.v44i.4794.

4. Zhu, Y. (2023) ‘Stock price prediction based on LSTM and XGBoost combination model’, Transactions on Computer Science and Intelligent Systems Research, 1, pp. 94–109. doi: 10.62051/z6dere47.

5. Tong, Z., Li, H. and Zhang, Y. (2023) ‘A comparison of ARIMA, LSTM, XGBoost and hybrid in stock price prediction’, 2023 5th International Conference on Applied Machine Learning (ICAML), pp. 18–25. doi: 10.1109/icaml60083.2023.00015.

6. Kumar, D.S., et al. (2022) ‘Analysis and prediction of stock price using hybridization of SARIMA and XGBoost’, 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 1–4. doi: 10.1109/ic3iot53935.2022.9767868.

7. Ming, L. and Chen, G. (2024) ‘Stock price prediction based on relative strength index, moving average convergence divergence and XGBoost model’, 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS), pp. 1988–1993. doi: 10.1109/ddcls61622.2024.10606580.

8. Li, J., et al. (2024) Enhancing financial time series forecasting with hybrid deep learning: CEEMDAN-informer-LSTM model. Preprint. doi: 10.2139/ssrn.4923415.

9. Zhelev, S. and Avresky, D.R. (2019) ‘Using LSTM neural network for time series predictions in financial markets’, 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA), pp. 1–5. doi: 10.1109/nca.2019.8935009.

10. Oukhouya, H., et al. (2023) ‘Forecasting international stock market trends: XGBoost, LSTM, LSTM-XGBoost, and backtesting XGBoost models’, Statistics, Optimization & Information Computing, 12(1), pp. 200–209. doi: 10.19139/soic-2310-5070-1822.

11. Semmelmann, L., Henni, S. and Weinhardt, C. (2022) ‘Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data’, Energy Informatics, 5(S1). doi: 10.1186/s42162-022-00212-9.

12. Jain, S. (2024) ‘A comparative study of stock market prediction models’, Deep Learning Tools for Predicting Stock Market Movements, pp. 249–269. doi: 10.1002/9781394214334.ch11.

13. Sharma, S. (2024) ‘Various model applications for causality, volatility, and co-integration in stock market’, Deep Learning Tools for Predicting Stock Market Movements, pp. 147–159. doi: 10.1002/9781394214334.ch6.

14. Shen, J. (2020) ‘Short-term stock market price trend prediction using a customized deep learning system’, Preprint. doi: 10.22215/etd/2019-13721.

15. Letteri, I. (2024) ‘Stock market forecasting using machine learning models through volatility-driven trading strategies’, Proceedings of the 6th International Conference on Finance, Economics, Management and IT Business, pp. 96–103. doi: 10.5220/0012607200003717.

16. Site, A., Birant, D. and Isik, Z. (2019) ‘Stock market forecasting using machine learning models’, 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–6. doi: 10.1109/asyu48272.2019.8946372.

17. Dichtl, H., Drobetz, W. and Otto, T. (2021) ‘Forecasting market crashes via machine learning: Evidence from European stock markets’, SSRN Electronic Journal. doi: 10.2139/ssrn.3843319.

18. Pagliaro, A. (2023) ‘Forecasting significant stock market price changes using machine learning: Extra trees classifier leads’, Electronics, 12(21), p. 4551. doi: 10.3390/electronics12214551.

19. Najem, R., Bahnasse, A. and Talea, M. (2024) ‘Toward an enhanced stock market forecasting with machine learning and deep learning models’, Procedia Computer Science, 241, pp. 97–103. doi: 10.1016/j.procs.2024.08.015.

20. Arora, S., Pandey, A. and Batta, K. (2024) ‘Prediction of stock market using artificial intelligence application’, Deep Learning Tools for Predicting Stock Market Movements, pp. 185–202. doi: 10.1002/9781394214334.ch8.