JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

Comparative Risk Analysis and Price Prediction of Corporate Shares Using Deep Learning Models like LSTM and Machine Learning Models

Muhammad Mehdi[1*], Fawad Nasim[2], and Muhammad Qasim Munir[1]

[1]Department of Computer Science, The Superior University, Lahore, Pakistan.
[2]Faculty of Computer Science and Information Technology, The Superior University, Lahore, Pakistan.
[*]Corresponding Author: Muhammad Mehdi. Email: maahdi786@gmail.com

Received: May 13, 2024 **Accepted:** August 28, 2024 **Published:** September 01, 2024

Abstract: The prediction of share prices and risk analysis have always posed significant challenges for investors due to the influence of various economic, financial, and political factors. Inaccurate price predictions can lead to severe financial losses, particularly for investors with limited financial market knowledge. Recent research and advancements in Artificial Intelligence, Machine Learning, and Deep Learning models have greatly improved the accuracy of stock price predictions. This research focuses on applying the Long Short Term Memory model, a specialized Deep Learning technique, to predict the closing prices of Tech Industry stocks. The study calculates the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to evaluate the model's performance. Additionally, the results are compared with other models, including Feed Forward Neural Networks, Recurrent Neural Networks, and Machine Learning Models like Support Vector Machine and Gradient Boosting, using the Weighted Average metric. The Long Short Term Memory models showed the lowest weighted average error value of 0.0115, establishing it as one of the most effective models for predicting stock prices. The findings have significant implications for investors and risk analysts, particularly in the Tech Industry, offering a robust tool for improving stock price prediction accuracy.

Keywords: Feed Forward Neural Network (FNN); Gradient Boosting; Long Short Term Memory (LSTM); Recurrent Neural Network (RNN); Support Vector Machine (SVM).

1. Introduction

Analysis of risk of shares has always been an important and crucial concern for investors during investment in shares or stocks [1]. The nature of stock markets is very volatile and hence prices of shares keep on changing each new day. Behind this reason, there are several factors including economic, political instability, inflationary trends, the financial health of businesses, etc.

Hence investors are always at risk during the investment of their wealth in shares of various companies. Shares are considered the purest and basic type of corporate ownership and associated risk means that the desired outcome will not occur [2]. Before computer-based techniques, investors had to rely on their judgmental as well as accounting calculations to predict share prices and associated risks and returns. These calculations help them to lower the risk of loss of their money due to a sudden decrease in share prices or market crash.

However, with the advent of the latest computer-based techniques like Artificial Intelligence, Machine Learning, and Deep Learning models investors are guided to predict share prices and returns based on their historical data, risk assessments, and market conditions. These technologies tremendously assist them in picking a share or portfolio of shares with no or little risk of volatility and loss of returns [3].

The past data of share prices is numeric, sequential, or time series and resultantly we need to pick a suitable Machine Learning or Deep Learning model. In this regard, the Long Short Term Memory (LSTM) model is preferred and selected. It is a very simple model that produces reliable results on every kind of quantitative data so also it best suits and fulfills the requirement of corporate shares' risk analysis and price predictions.

Initially, the relationship between stock prices and their returns was considered to be linear but it is not always correct. This relationship has appeared to be non-linear being affected by many factors. Classical accounting formulas and ratio analysis cannot always help to predict share prices precisely and accurately. There is always a need for computer-based models based on Artificial Intelligence, Machine Learning (ML), or Deep Learning (DL) models to predict outputs of non-linear relationships, thus helping the investors to put their investments in portfolios of shares giving no or little loss [4].

In this research article, the LSTM model is used to predict the closing prices of Tech stocks (shares in companies in the Technology Industry). Their past performance and historical data are readily available with Yahoo Finance which is a very good repository of share price data of companies listed with different stock markets around the globe. For this research paper selected Tech companies include Google, Microsoft, Apple, and Amazon. The research focuses on contributing to the existing literature in many ways for example the application of ML and DL models in risk and price prediction of corporate stocks, a comprehensive comparative study highlighting LSTM's accuracy so also its advantages over other ML and DL models, and real-world application on Tech Stocks, which have unique volatility characteristics.

This research paper is structured into several sections. Firstly, the literature review covers past research on stock price predictions and machine learning models. Next, the methodology explains the data collection, preprocessing steps, and details of the model used which is LSTM. The results and discussion section presents a close and careful analysis and comparison of outcomes of other models FNN, RNN SVM, and Gradient Boosting with the results of the LSTM model. At last, the conclusion section summarizes the main findings and suggests future research dimensions and directions.

2. Literature Review

Computer-based predictive models can be cut into three broad areas, (1) Models based on Artificial Intelligence (AI) (2) Machine Learning (ML) Models (3) Deep Learning (DL) Models. There is numerous amounts of research, literature, and knowledge available for a researcher's review.

Stock markets provide lots of information; some of it helps investors, while other information can lead them astray and confuse them. AI models and algorithms like K Nearest Neighbor (KNN), Singular Value Decomposition (SVD), and Association Rule Mining (ARM) can provide investors with powerful and effective stock recommendation systems. This system aids them in selecting stocks that are likely to yield higher returns [5].

Digital Finance, also called E-Finance, is becoming very popular because of new advanced techniques using AI. These techniques help stakeholders make better decisions and transactions, allowing them to gain more and more benefits with less risk. They also assist in predicting bankruptcy, analyzing credit risk, and credit scoring [3].

Using AI-driven strategies in stock price prediction focuses on two main methods: (1) Technical Analysis and (2) Fundamental Analysis. Technical Analysis uses "Regression Machine Learning Algorithms" to predict stock price trends by analyzing historical data from reliable sources like Yahoo Finance. This method provides a data-driven view, using past performance patterns to forecast future price changes. Whereas, Fundamental Analysis uses "Classification Machine Learning Algorithms" to classify investors' sentiments about the stock market. It relies heavily on news and social media resources to make predictions [6].

Since 1960, the Capital Asset Pricing Model (CAPM) has been a famous mathematical model for calculating expected returns. It describes a linear relationship between expected returns, market returns, risk-free returns, and asset Betas. Although useful, CAPM has many limitations, the major limitation being that the relationship between these variables is non-linear. ML and DL models are powerful in recognizing complex patterns and handling the growing amount of unstructured data [7].

ML models can help to predict the default risk of companies traded on the stock exchange. For example, analyzing financial data from 50 Iranian companies between the period 2016 and 2021 showed that these companies were declared bankrupt according to Iran's laws. In this context, Random Forest and Gradient Boosting Decision Trees offer a data-focused approach. These models use algorithms and computational techniques to analyze financial data and predict the default risk of different companies.

They are effective in handling complex financial situations and can accurately forecast a company's likelihood of default [8].

The use of ML models to predict stock market trends has gained significant attention in recent years. A review of 138 scholarly articles from 2000 to 2019 examines various market and stock indices. These studies also explored different variables used as inputs in ML models, with North American and Asian markets being the most studied. Common indicators like return rates, Simple Moving Averages, and Relative Strength Index are frequently used in these techniques. Neural Networks and Support Vector Machines are preferred algorithms due to their high accuracy rates. However, the precision of these models can vary indicating a need for further research to improve predictive abilities [9].

Stock market trends, which involve time series data, need systems capable of predicting profits. These markets are non-stationary, dynamic, and volatile, making predictions more challenging. A thorough review of 57 articles from 1991 to 2017 highlights the latest ML techniques used to forecast stock market trends [10]. In addition to traditional methods, using advanced technologies like SVM represents a significant change, offering new ways to better understand and predict stock market behavior [11].

Using ML algorithms to estimate "Market Betas" for stocks has proven to be better than traditional models, both economically and statistically. These ML estimators not only offer the lowest forecasting and hedging errors but also help to create market-neutral strategies and minimum variance portfolios [12]. Over the past two decades, the advantages of using computer-based methods to predict stock market trends have become visible and clear. Research on these methods is usually categorized into four main types (1) Artificial Neural Networks, (2) Support Vector Machines, (3) Generic Algorithms, and (4) Hybrid Approaches. Each category offers unique insights into the strengths and weaknesses of different prediction techniques [13].

Famous DL techniques such as LSTM, Gated Recurrent Unit (GRU), and Convolutional Neural Networks (CNN) are designed to forecast the closing prices of stock markets in developing countries like Nepal. These markets are more complex due to influences like macroeconomic factors that can cause disruptions [14]. Capital markets or stock exchanges can be accurately predicted using non-linear models. Multi-layer neural networks are highly effective at estimating complex target functions that have non-linear characteristics, involve multiple variables, and require extensive parameter tuning [15].

Investing in a single financial asset or a portfolio is challenging because financial markets are unpredictable, making it hard for simple models to accurately predict future asset values. To address this issue, Recurrent Neural Networks (RNN) especially the LSTM model are used. These advanced models can predict future stock market values more accurately than traditional models [16]. Likewise, CNN and LSTM can be used together to make predictions using historical time series data. In such models, CNN extracts feature from the dataset, whereas LSTM is used to predict stock prices with high accuracy [17].

Other DL models, such as Multilayer Perceptron (MLP) and Attention-Based Neural Networks can also be used to predict the next day's prices based on historical data. These methods are effective in all markets, whether highly developed, less developed, or developing markets [18]. Stock market performance can be evaluated using Market Capitalization ratios and other factors. This can also be achieved with modern tools like Artificial Neural Networks (ANN). A three-layer Feed Forward Neural Network with backpropagation algorithms, 5 nodes in a hidden layer, and a learning rate of 0.01 produces the best model. The errors estimated by ANN are lower than those from older methods [19].

3. Methodology

This research uses a structured approach to analyze and predict the closing prices of Tech companies' shares. The data used for this research is sourced from Yahoo Finance, a platform owned by Yahoo Inc., which offers comprehensive financial quotes, news, press releases, and reports. The proposed methodology starts with the extraction and preprocessing of the dataset. After that, the data is divided into training and testing sets. The LSTM model is developed, trained, and evaluated. The performance of the LSTM model is compared with models like FNN, RNN, SVM, and Gradient Boosting.

Detailed steps of the proposed methodology are as under;

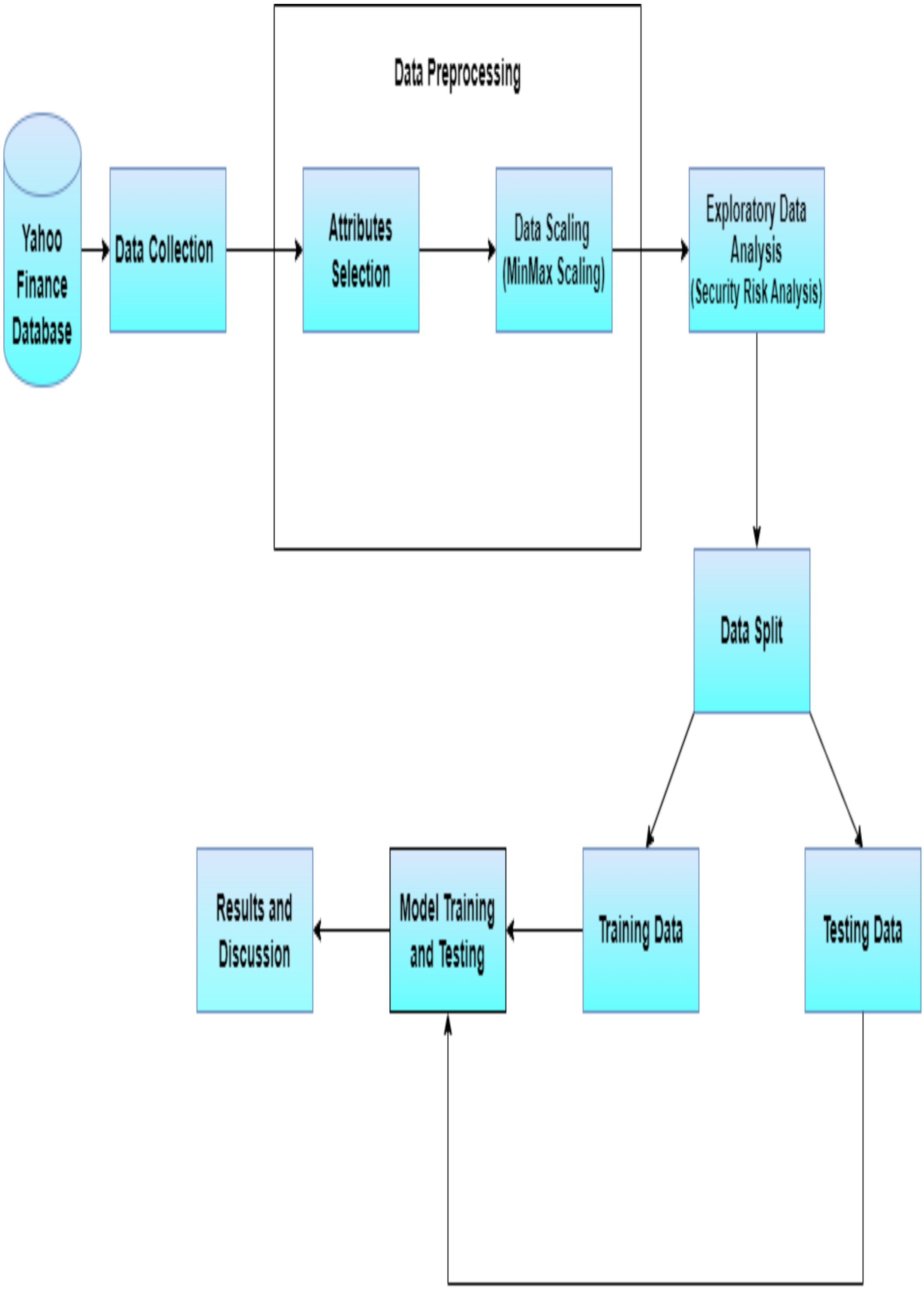

Figure 1. Steps of proposed methodology

3.1. Data Collection and Preprocessing

One can easily find stock price data for popular companies on Yahoo Finance. For this research article, the Technology Industry is chosen and famous companies like Apple, Microsoft, Google, and Amazon are selected. For risk analysis purposes, the extracted dataset has 8 features/columns that is Date, Open Price, High Price, Low Price, Close Price, Adj. Close Price, Volume, and Company Name, covering the period from 10.07.2023 till 10.07.2024. Python programming language is selected for this task as it is quite user-friendly and offers a variety of libraries for data extraction, preprocessing, visualization, model creation, and making predictions.

The historical values of closing prices of these Tech stocks are as under;

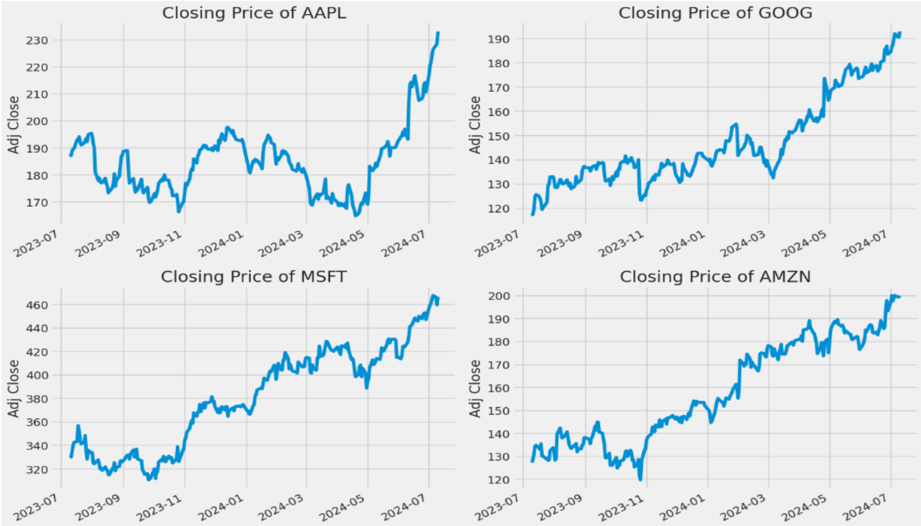

Figure 2. Graph showing closing prices of chosen Tech Stocks (Apple, Microsoft, Google and Amazon)

For this extracted data, based on 10, 20, and 50 days, Moving Averages are calculated. The graph so plotted shows that Moving Averages calculated based on 50 days started to move away from the adjusted close price whereas Moving Averages based on 10 and 20 days are more stable.

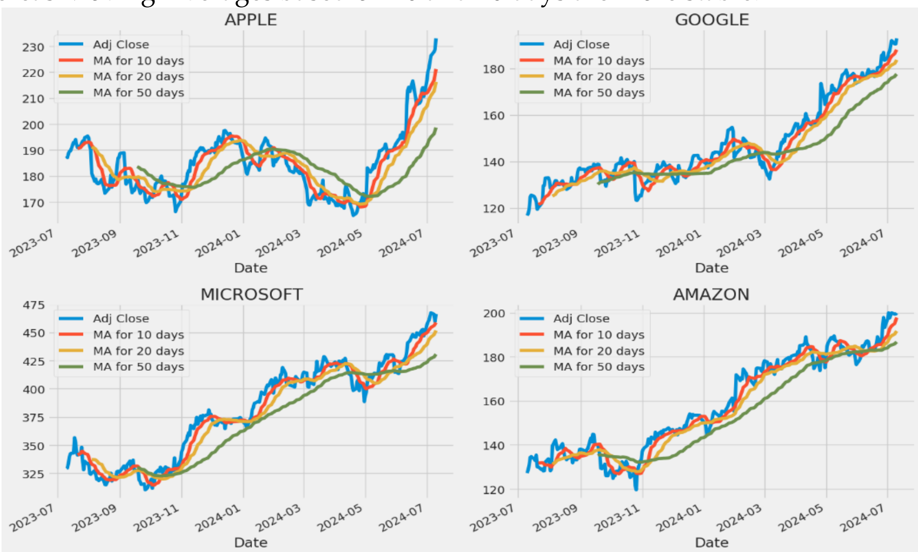

Figure 3. Graph showing Moving Averages based on 10, 20, and 50 Days

Daily returns of these Tech stocks can also be extracted from Yahoo Finance and can also be graphed as under.

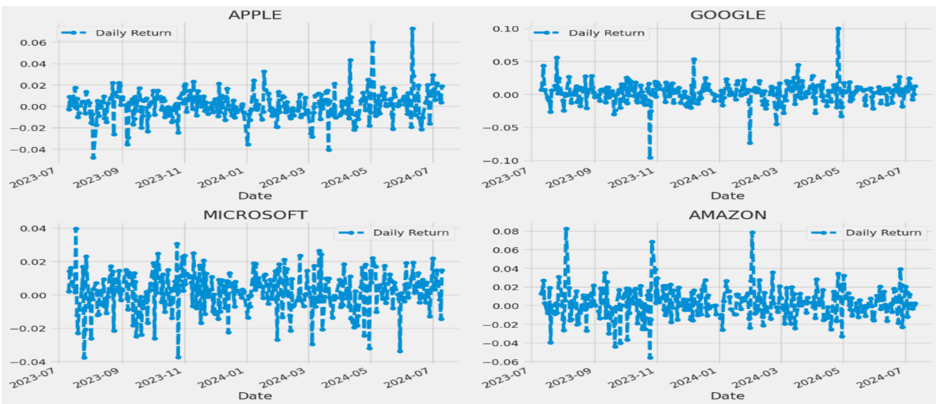

Figure 4. Graph showing daily returns of selected Tech Stocks

For more understanding, Histogram for daily returns can also be visualized.

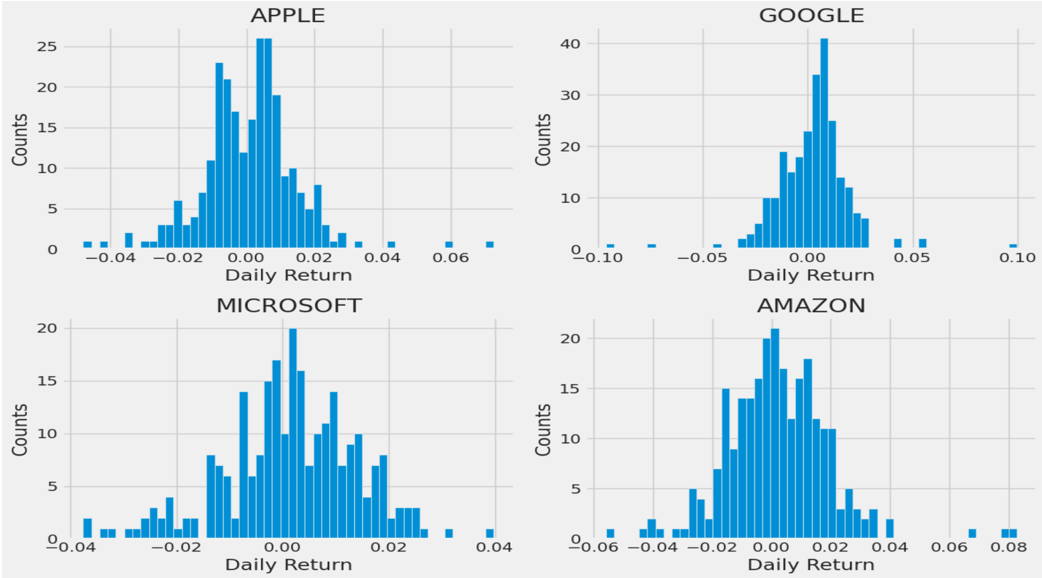

Figure 5. Histogram showing daily returns

These shares belong to the same industry that is Tech Industry, there are huge chances that the prices and returns of these shares are positively correlated. For example, a graph between the returns of Microsoft and Google is drawn to check whether their returns are correlated positively or not.

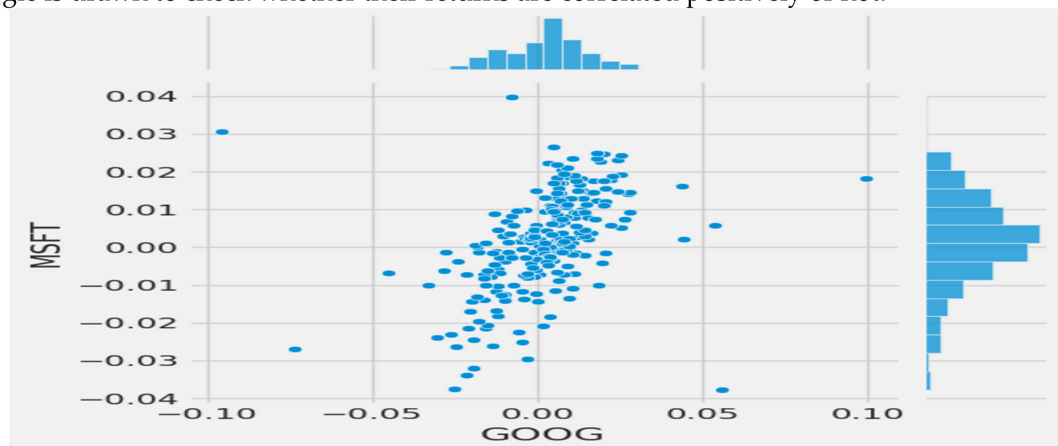

Figure 6. Graph displaying the positive correlation between returns of Microsoft and Google

Similarly, a correlation plot drawn to get numerical values of stock returns and prices is also visualized.

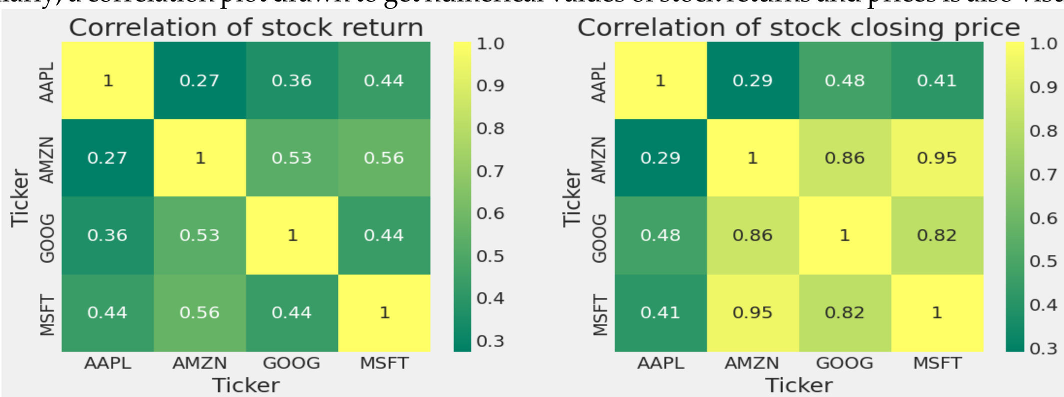

Figure 7. Graph showing the Correlation of numerical values of selected Tech stocks

The above graph shows that if two stocks have a positive correlation in their prices, their returns also tend to be positively correlated. For instance, the correlation between the returns of Microsoft and Amazon is 0.56 and their prices it is 0.95. Now this type of portfolio can be riskier because a drop in price and return of one stock might lead to a drop in the other stock as well. This decline could be due to factors such as economic, and political reasons. Hence to avoid losses, an investor should create a diversified portfolio of shares that is investment to be made in shares of companies from different industries.

The risk of these Tech stocks is also assessed, the graph between expected returns and associated risk of each stock is as under;

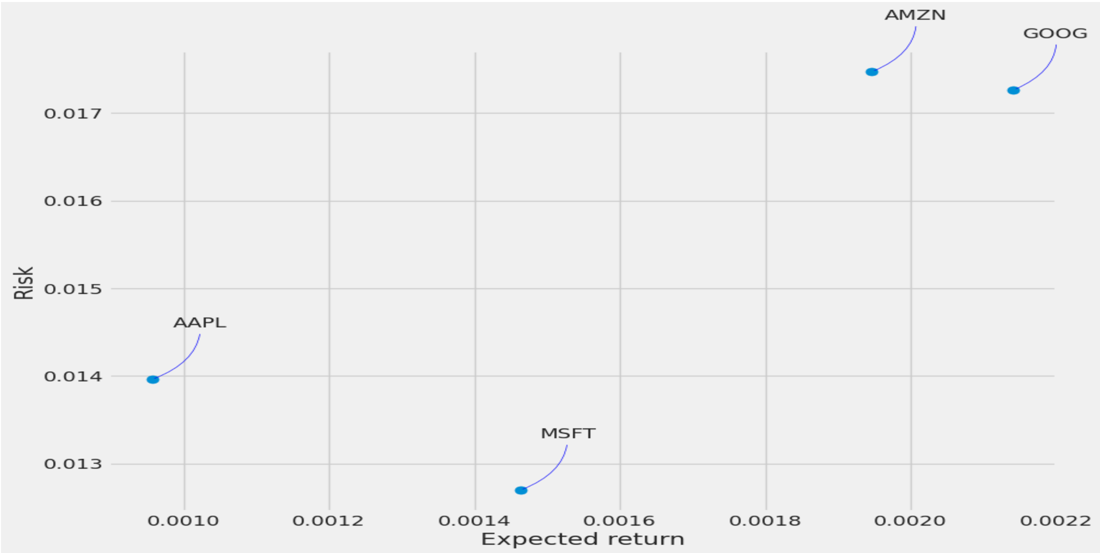

Figure 8. Risk Assessment: Graph of risk and expected returns

From the graph above, it is visible that Amazon and Google are the most riskier but at the same time, their returns are higher than those of other selected stocks. Endorsing this concept, the higher the risk, the higher the returns.

3.2. Data Split and Model Development

For this purpose, stock data for Apple Company from 01.01.2012 till 10.07.2024 have been collected from Yahoo Finance. The dataset has 3,149 rows and includes 6 columns that are Open Price, High Price, Low Price, Close Price, Adjusted Close Price, and Volume. The extracted data can be graphed as shown below;

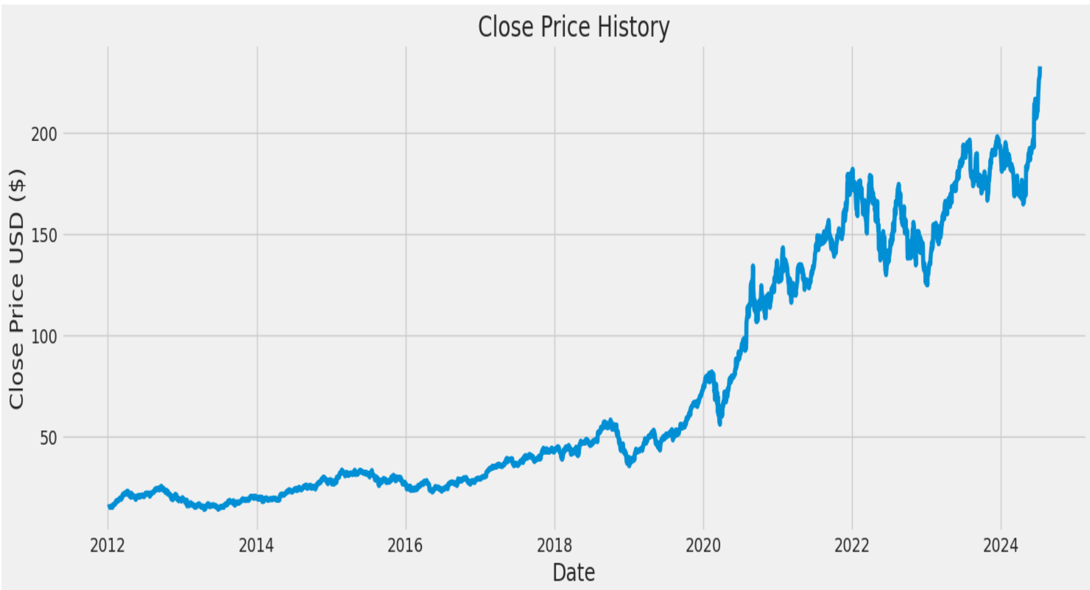

Figure 9. Closing Price Historical data shows a rising trend in Apple Prices

Divide the dataset into training and testing sets, using 2,992 out of 3,149 rows (95% of the data) for training and validation, and the rest for testing. The data is scaled into values between 0 and 1 (by using the MinMax Scaling Method). This scaled data helps to improve the performance and speed of the deployed model.

Next LSTM model (the focused model of this research article) is developed by using the Keras library in Python programming, specifically using Keras.Model and Keras.Layers. The model includes two layers of LSTM, one layer with 128 memory cells and the other with 64 memory cells, thus helping the model to learn long-term dependencies in the input data.

Furthermore, two Dense (fully connected) layers with 25 neurons and 1 neuron respectively are added to the model. These layers help to learn complex patterns from the output of previous layers. The model is then compiled using the Adam optimizer, with Mean Square Error as the loss function. The model is trained for 5 epochs. After training, the model is tested using the test data to predict closing price values.

4. Results and Discussion

4.1. Results

To estimate the errors and accuracy, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error are calculated. These errors belong to Regression Metrics, most suitable for sequential or time series data types. The quality of RMSE is that it gives errors in units [20]. So, it is easy to understand.

RMSE calculates the square root of averaged squared deviations between the predicted and actual values, the formula is as under;

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2}$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2}$$

Where $\mathbf{A_i}$ is the actual value and $\mathbf{P_i}$ is the predicted value, the lower the RMSE the better the model's performance.

MAPE stands for Mean Absolute Percentage Error. It is used to measure the accuracy of the forecasting model. MAPE calculates the accuracy of the model as a percentage, the mathematical expression is as under;

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|A_i - P_i|}{Ai}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|A_i - P_i|}{Ai}$$

where $\mathbf{A_i}$ is the actual value and $\mathbf{P_i}$ is the predicted value. Both RMSE and MAPE, are equally important and their use depends upon the type of dataset.

The comparison of actual closing prices and predicted prices by the LSTM model are shown as under;

Table 1. Comparison of Actual Closing Prices and Predicted Prices by LSTM

Date	Actual Closing Prices	Predicted Closing Prices
22.11.2023	191.31	191.23
24.11.2023	189.97	192.28
27.11.2023	189.79	190.67
28.11.2023	190.40	190.86
…		
05.07.2024	226.34	220.31
08.07.2024	227.82	225.43
09.07.2024	228.68	225.76
10.07.2024	232.00	226.65

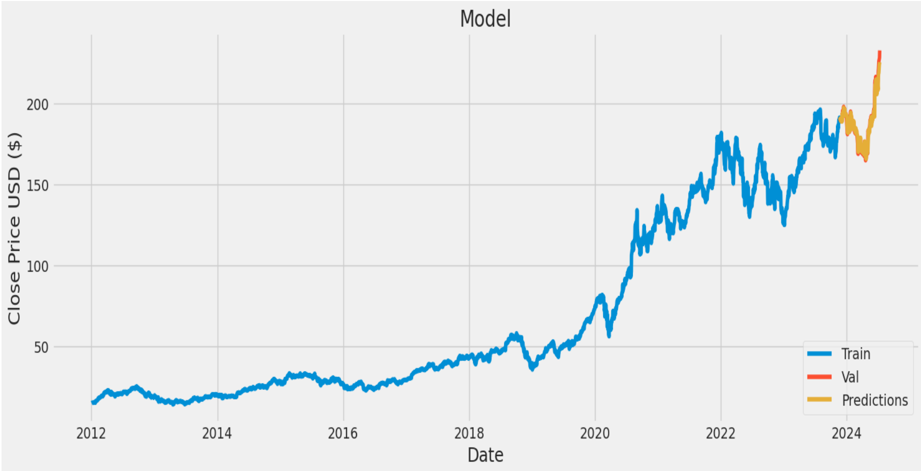

Figure 10. Visualization of Actual Closing and Predicted Prices by LSTM

The calculated values of RMSE and MAPE for our LSTM model are 2.9322 and 98.8889 % respectively, which means that there is a difference of 2.9322 Dollars in predicted values and model accuracy is 98.8889%.

4.2. Discussion

LSTM model comes up with one of the best models to predict time series datasets of share prices. To validate this result let us compare the performance of the LSTM model with two other famous DL models (1) Forward Neural Network (FNN) and (2) Recurrent Neural Network (CNN) and two ML models (1) Support Vector Machine (SVM) and (2) Gradient Boosting. The calculated values of RMSE, MAPE, and Weighted Average (giving equal weights (50%) to each error) are available in the below table.

Table 2. Comparison of RMSE, MAPE, and Weighted Averages

Model	RMSE	MAPE	Weighted Average
LSTM	2.9322	98.8889	0.0115
FNN	187.307	99.5916	0.4999
RNN	187.2846	99.5807	0.4999
SVM	187.3608	99.6176	0.5000
Gradient Boosting	10.451	0.02861	0.5277

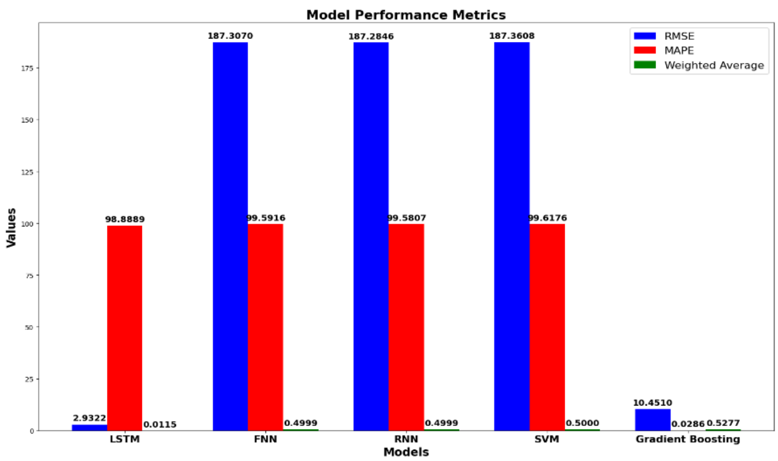

Figure 11. Graph showing performance metrics of models

According to the above table, the RMSE of LSTM is far better than both other four selected models, the RMSE for LSTM is 2.9322 whereas for FNN, RNN, SVM, and Gradient Boosting it is 187.3070, 187.2846, 187.3608 and 10.4510 respectively. Hence, keeping in view the results of RMSE, LSTM is a better model for Time Series Predictions. If MAPE is taken into account, the results of FNN, RNN, and SVM are the same that is 99% (approximately), and hence well performed the LSTM model whose accuracy is 98.8889%. However, Gradient Boosting has a very low level of accuracy that is 0.028%

In this situation where one cannot make a decision about which model is the best predictor, the technique of Weighted Averages, giving equal weights to each metric, can be used. The model with a lesser value of Weighted Average will be declared as the best model. Doing so, it is discovered that the combined score or Weighted Average of LSTM is 0.0115 which is less than all other selected models.

Hence, LSTM is proven to be one of the best options for predicting stock market data (time series-based prices and returns)

5. Conclusion

The nature of share market data is dynamic and volatile. Investors get extremely afraid of sudden price fluctuations in negative directions. These reductions in prices can cause them severe losses. Although there are many traditional financial management and accounting tools available, these are not free from errors and misreporting. With the arrival of recent computer-based technologies like Artificial Intelligence, Machine, and Deep Learning models, the fluctuations and volatility in stock markets can be predicted more accurately and precisely. Models like LSTM, FNN, RNN, SVM, etc. can be really helpful in predicting stock prices. However, it should be kept in mind that there is always a need for improvement as the above provided results have clarified that these models are also not 100% accurate. These models can be used for predictions, but it is important to remember that the chances of errors are still present. Efforts should be focused on creating such models that ensure a reduction in the likelihood of errors.

References

1. Kevin, Sanjay. Security analysis and portfolio management. PHI Learning Pvt. Ltd., 2022.
2. Gitman, Lawrence J., Roger Juchau, and Jack Flanagan. Principles of managerial finance. Pearson Higher Education AU, 2015.
3. Najem, Rihab, Meryem Fakhouri Amr, Ayoub Bahnasse, and Mohamed Talea. "Artificial intelligence for digital finance, axes, and techniques." Procedia Computer Science 203 (2022): 633-638.
4. Pan, Shuiyang, Suwan Cheng Long, Yiming Wang, and Ying Xie. "Nonlinear asset pricing in Chinese stock market: A deep learning approach." International Review of Financial Analysis 87 (2023): 102627.
5. Gonzales, Reyes Michaela Denise, and Carol Anne Hargreaves. "How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system." International Journal of Information Management Data Insights 2, no. 2 (2022): 100130.
6. Mokhtari, Sohrab, Kang K. Yen, and Jin Liu. "Effectiveness of artificial intelligence in stock market prediction based on machine learning." arXiv preprint arXiv:2107.01031 (2021).
7. Ndikum, Philip. "Machine learning algorithms for financial asset price forecasting." arXiv preprint arXiv:2004.01504 (2020).
8. Peykani, Pejman, Mostafa Sargolzaei, Negin Sanadgol, Amir Takaloo, and Hamidreza Kamyabfar. "The application of structural and machine learning models to predict the default risk of listed companies in the Iranian capital market." Plos one 18, no. 11 (2023): e0292081.
9. Kumbure, Mahinda Mailagaha, Christoph Lohrmann, Pasi Luukka, and Jari Porras. "Machine learning techniques and data for stock market forecasting: A literature review." Expert Systems with Applications 197 (2022): 116659.
10. Henrique, Bruno Miranda, Vinicius Amorim Sobreiro, and Herbert Kimura. "Literature review: Machine learning techniques applied to financial market prediction." Expert Systems with Applications 124 (2019): 226-251.
11. Paiva, Felipe Dias, Rodrigo Tomás Nogueira Cardoso, Gustavo Peixoto Hanaoka, and Wendel Moreira Duarte. "Decision-making for financial trading: A fusion approach of machine learning and portfolio selection." Expert Systems with Applications 115 (2019): 635-655.
12. Drobetz, Wolfgang, Fabian Hollstein, Tizian Otto, and Marcel Prokopczuk. "Estimating stock market betas via machine learning." Journal of Financial and Quantitative Analysis (2023): 1-56.
13. Strader, Troy J., John J. Rozycki, Thomas H. Root, and Yu-Hsiang John Huang. "Machine learning stock market prediction studies: review and research directions." Journal of International Technology and Information Management 28, no. 4 (2020): 63-83.
14. Pokhrel, Nawa Raj, Keshab Raj Dahal, Ramchandra Rimal, Hum Nath Bhandari, Rajendra KC Khatri, Binod Rimal, and William Edward Hahn. "Predicting nepse index price using deep learning models." Machine Learning with Applications 9 (2022): 100385.
15. Fallahgoul, Hasan, Vincentius Franstianto, and Xin Lin. "Asset pricing with neural networks: Significance tests." Journal of Econometrics 238, no. 1 (2024): 105574.
16. Moghar, Adil, and Mhamed Hamiche. "Stock market prediction using LSTM recurrent neural network." Procedia computer science 170 (2020): 1168-1173.
17. Bilal, A., Imran, A., Baig, T. I., Liu, X., & Abouel Nasr, E. (2024). Breast cancer diagnosis using support vector machine optimized by improved quantum inspired grey wolf optimization. Scientific Reports, 14(1), 10714. Nature Publishing Group.
18. Lu, Wenjie, Jiazheng Li, Yifan Li, Aijun Sun, and Jingyang Wang. "A CNN-LSTM-based model to forecast stock prices." Complexity 2020, no. 1 (2020): 6622927.
19. Sajjad, R., Khan, M. F., Nawaz, A., Ali, M. T., & Adil, M. (2022). Systematic analysis of ovarian cancer empowered with machine and deep learning: a taxonomy and future challenges. Journal of Computing & Biomedical Informatics, 3(02), 64-87.
20. Gao, Penglei, Rui Zhang, and Xi Yang. "The application of stock index price prediction with neural network." Mathematical and Computational Applications 25, no. 3 (2020): 53.
21. Rubi, Maksuda Akter, Shanjida Chowdhury, Abdul Aziz Abdul Rahman, Abdelrhman Meero, Nurul Mohammad Zayed, and KM Anwarul Islam. "Fitting multi-layer feed-forward neural network and autoregressive integrated moving average for Dhaka Stock Exchange price predicting." Emerging Science Journal 6, no. 5 (2022): 1046-1061.
22. Hodson, Timothy O. "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not." Geoscientific Model Development Discussions 2022 (2022): 1-10