Journal of Innovative Computing and Emerging Technologies

Security Risk Analysis and Price Predictions with Machine and Deep Learning Models (LSTM)

Muhammad Mehdi[1], Inam Ul Haq[2], Atif Ikram[2]

Received: 17 April,2023; Accepted: 20 Jan, 2024; Published: 30 March, 2024

Abstract: Risk analysis and price predictions of securities, shares and stocks, have been a challenging problem for investors. Many factors, Economic, Political etc., can disturb stock returns and their prices. However, with the advent of Artificial Intelligence and Machine/Deep Learnings techniques, predictions of returns and prices have become very easy with accuracy and precision. In this research paper focus is given to Long Short Term Memory (LSTM) model, closing prices of Tech stocks/shares of companies in technology industry are predicted, RMSE and accuracy (MAPE) is calculated and then compared these values with the predicted and calculated values of Feedforward Neural Network (FNN) model and Recurrent Neural Network (RNN) model. After using weighted average method, LSTM is proven to be one of the best models to predict securities prices and returns.

Keywords: Artificial Intelligence, Deep Learning, Feedforward Neural Network (FNN), Long Short-Term Memory (LSTM), Machine Learning, Recurrent Neural Network (RNN), Security Risk Analysis, Stocks, Shares

1- Introduction

In the past "Security Risk Analysis" is an important and very challenging concern for investors who are going to invest in the securities which include share, stock and corporate bonds(Kevin, 2022). Stocks markets are very much volatile and price of stocks keep on changing each day. This happens due to several reasons which includes Economic factors, political instability, inflations, company's financial health etc.

So investors are at risk while investing their money in stocks of different companies. Risk means that desired outcome will not happen. In the past, to overcome this issue, they used their judgmental as well as some Accounting calculations to predict stock prices and returns, in such a way they tend to minimize their risk of loss of money due to sudden drop in stock prices or market crashes.

[1]*Department of Computer Science COMSATS University Islamababd, Lahore 54000, Pakistan*
[2]*Department of Information Technology, University of Okara Pakistan, inamulhaq@uo.edu.pk*
[3]*Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Malaysia, aikram4u@gmail.com*
[*]*Corresponding Author maahdi786@gmail.com*

However, with the arrival and advancement of latest computer related technologies like Artificial Intelligence and Machine/Deep learning models, these investors can be assisted in stock prices predictions based on past data of stock prices, market conditions security risk assessments etc. This ultimately help them to select a stock or portfolio of stock with less risk of volatility and loss of money Najem et al., 2022 [1].

The historical data of stock prices is a numerical, time series or sequential data, hence a suitable Machine Learning or Deep Learning models to be used. Focus and preference is given to Long Short Term Memory (LSTM) model. LSTM method suits and fulfills the requirement of Security Risk Analysis and Price Predictions so also it is a very simple model that support all kind quantitative data.

Earlier the relationship between stock price and returns was considered to be a linear relationship, however it is not always true as this relationship is discovered to be more non-linear. We cannot predict stock prices and returns precisely/accurately using only classical Accounting formulas and ratios. There is also a need of intelligent computerized models based on AI, Machine Learning or Deep learning models for the predictions of outputs of these non-linear relationships, thus facilitating the investors to invest in stocks with confidence of no or little loss (Pan et al., 2023).

For this research paper LSTM model is used to calculate and predict correlations, returns and closing prices of Tech Stocks in Technology Industry whose historical data is readily available on Yahoo Finance, a very good repository of financial data of various companies listed with different stock markets around the world. We have selected companies like Google, Apple, Microsoft and Amazon for this research paper.

2-Literature Review

Computer based predictive models can be divided in 3 broad categories, (1) Models based on Artificial Intelligence (2) Machine Learning Models (3) Deep Learning Models. Numerous amount of knowledge and literature is available for a researcher's review.

2.1 Models based on Artificial Intelligence

Stock markets are full of information; some assist the investors while other distract their decisions. Various models and algorithms based on Artificial Intelligence like K Nearest Neighbor (KNN), Singular Value Decomposition (SVD) and Association Rule Mining (ARM) can give an investor a powerful and effective stock recommender system which helps them to select stocks with growing returns (Gonzales & Hargreaves, 2022).

Digital Finance also known as E-Finance is witnessing a surge in its popularity due to arrival of various advance techniques backed by Artificial Intelligence helping stakeholders to take more informed transactions and proactive decision making helping them to get maximum benefits while taking less amount of risk. These techniques also support to predict bankruptcy, credit risk analysis/modelling and credit scoring etc. (Najem et al., 2022).

Employing AI strategies in domain of stock price prediction with a focus on two basic Analytical Methodologies (1) Technical Analysis and (2) Fundamental Analysis. Technical Analysis uses Regression Machine Learning Algorithms to predict stock price trends. Fetching historical data from any reliable source like Yahoo Finance, this technique provides data driven perspective, enhancing past performance patterns to predict future price fluctuations. on the other hand, Fundamental Analysis employs classification Machine Learning Algorithms to encompass sentiments of investors on stock market forecast. News, Social Media resources play vital role in this regard (Mokhtari et al., 2021).

2.2 Machine Learning Models

Since 1960, Capital Asset Pricing Model (CAPM), which is a famous model to calculate expected returns, it is a linear relationship among expected returns, market returns, risk free returns and asset Betas. Although it is useful to calculate expected returns but accompanied by many limitations. One of the identified limitation is relationship among above mentioned variables is non-linear and hence can be best explained by Machine Learning algorithms like neural networks and gradient boosting as these are powerful model for recording non-linear relationship and effectively taking into account expanding realm of unstructured data (Ndikum, 2020).

Machine Learning models are useful to predict default risk of companies being traded on stock exchange for example financial data of 50 Iranian companies from 2016 to 2021 showed that these companies were declared bankrupt as per prevailing laws in Iran. In this regard Random Forest and Gradient Boosted Decision Tree provides a data centric approach. These models rely on algorithms and computational techniques to utilize financial data to predict default risk of different companies, these models are capable to solve complex financial scenarios and help to predict default risk of a company [2]

The use of ML models to forecast and predict stock markets trends has attracted considerable attention in the past few years. In an attempt of drawing an in-depth knowledge an examination of 138 scholarly articles period ranging from 2000 to 2019 threw light on the market and stock indices under consideration. These researches also explored the range of variables employed as inputs in various Machine Learning models. North American and Asian markets are the most studies market. Several Statistical indicators like return rates, Simple Moving Averages and Relative Strength Index are the most popular indicators used in ML techniques. So also Neural Networks and Support Vector Machines are the preferred algorithms with excellent accuracy rates. However it is worth mention that precision and efficacy of these models can vary highlighting the need of further research to refine predictive ability of these ML techniques(Kumbure et al., 2022).

Stock Market trends, being time series data, require systems that can predict profits. The nature of these markets is non-stationary, dynamic and having chaotic attributes, further enhancing difficulty in predictions. A comprehensive literature review for the period 1991 to 2017 of 57 articles identifies state of the art Machine Learning Techniques to predict stock market trends(Henrique et al., 2019). Apart from traditional notions, utilizing advanced technologies like SVM, represent a potential paradigm shift, opening avenues for more refined approaches to understand and predict stock market behavior (Paiva et al., 2019).

Estimating market Betas for capital assets (shares or stock) through ML algorithms has proven to be the most superior approach living behind established benchmark models both economically and statistically. The ML estimators are not only providing lowest forecasting and hedging errors but also contributing to the development of marketneutral anomaly strategies and minimum variance portfolios (Drobetz et al., 2021). Over the past two decades the benefits of computerized techniques to predict stock market trends have been recognized and acknowledged. The study of these techniques have been divided into four main categories (1) Artificial Neural Networks studies, (2) SVM studies, (3) Studies using Genetic Algorithms, (4) Studies using hybrid approaches, giving a deep understanding of strengths and weaknesses inherent in each of these predictive methods (Strader et al., 2020).

2.3 Deep Learning Models

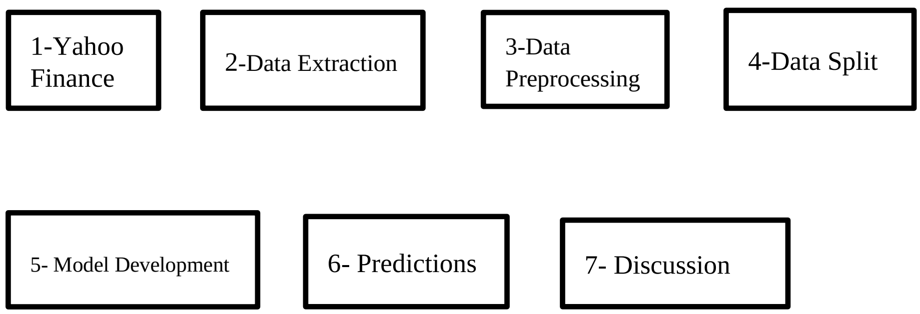
Famous Deep Learning techniques like Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN) are specifically aimed at predicting closing price of stock markets of developing countries like Nepal as such markets are more complex as various factors like Macroeconomics factors etc. can disturb these markets (Pokhrel et al., 2022). Capital markets or stock exchanges can be well predictable with non-linear models. Multi-Layer Neural Networks are well known for their efficacy in estimating complex target functions, having non-linear attributes, multiple variables and extensive parameterization (Fallahgoul et al., 2024).

It is not easy to invest in a single financial asset or portfolio of financial assets due to abnormality of financial markets that do not allow simple models to predict future asset values with required accuracy. To overcome the problem Recurrent Neural Network and especially LSTM model is used to predict future stock market values more accurately then do traditional simple models (Moghar & Hamiche, 2020). Likewise, Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) can be used for predictions on the basis of historical time series data, in such model CNN can be used to extract features from the data and LSTM is used to predict stock prices with highest prediction accuracy(Lu et al., 2020).

Some other Deep Learning Methods like Multilayer Perceptron (MLP) and Attention-Based Neural Network can also be used to predict next day price according to historical data. These methods can maintain their accuracy in any kind of market whether it is most developed, less developed or developing market (Gao et al., 2020). Stock market performance can be measured by the Market Capitalization ratios so also by measuring many other factors. This task can also be done with Artificial Neural Networks (ANN) which is a modern tool. A three-layer Feed Forward Neural Network using back propagation algorithms with 5 nodes in hidden layer and learning rate is set to be 0.01 generates the best model. The estimated errors of ANN are lesser then estimated errors of other old methods (Rubi et al., 2022).

3- Proposed Methodology

The steps of methodology will follow following steps;



.

Figure 1: Proposed method

3.1 Data Extraction and Preprocessing

The price data of stocks of world-famous companies is readily available on Yahoo Finance web site. We have selected Technology Industry, selected companies are Apple, Google, Microsoft and Amazon. Total 8 attributes (Date, Open Price, High Price, Low Price, Close Price, Adj. Close Price, Volume and Company Name) are selected for these Tech stocks and then data is extracted for each attribute, period selected is 21.02.2023 till 16.02.2024. For this purpose, Python Language is selected as this programming language is easy and provides a large list of libraries for data extraction, data preprocessing, data visualization, model creation and predictions etc. A historical view of closing prices of these stocks is as under;

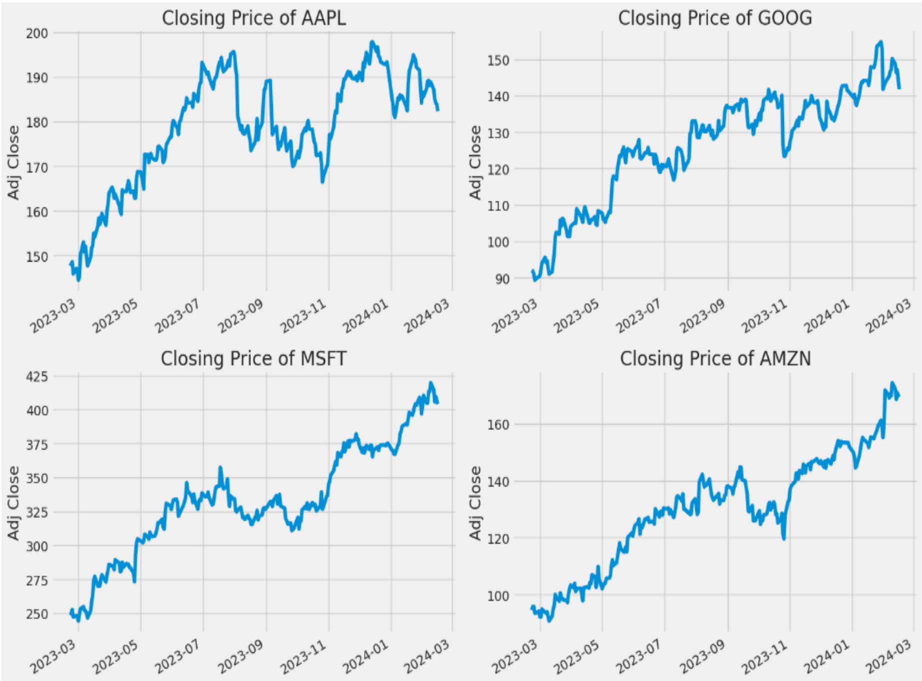


Figure 2. Graph of closing prices of selected Tech Stocks (Apple, Google, Microsoft and Amazon)

For same extracted, Moving Averages are calculated based on 10, 20 and 50 days. The graph so visualized showed that Moving Averages calculated on the basis of 50 days started to move away from adjusted close price.

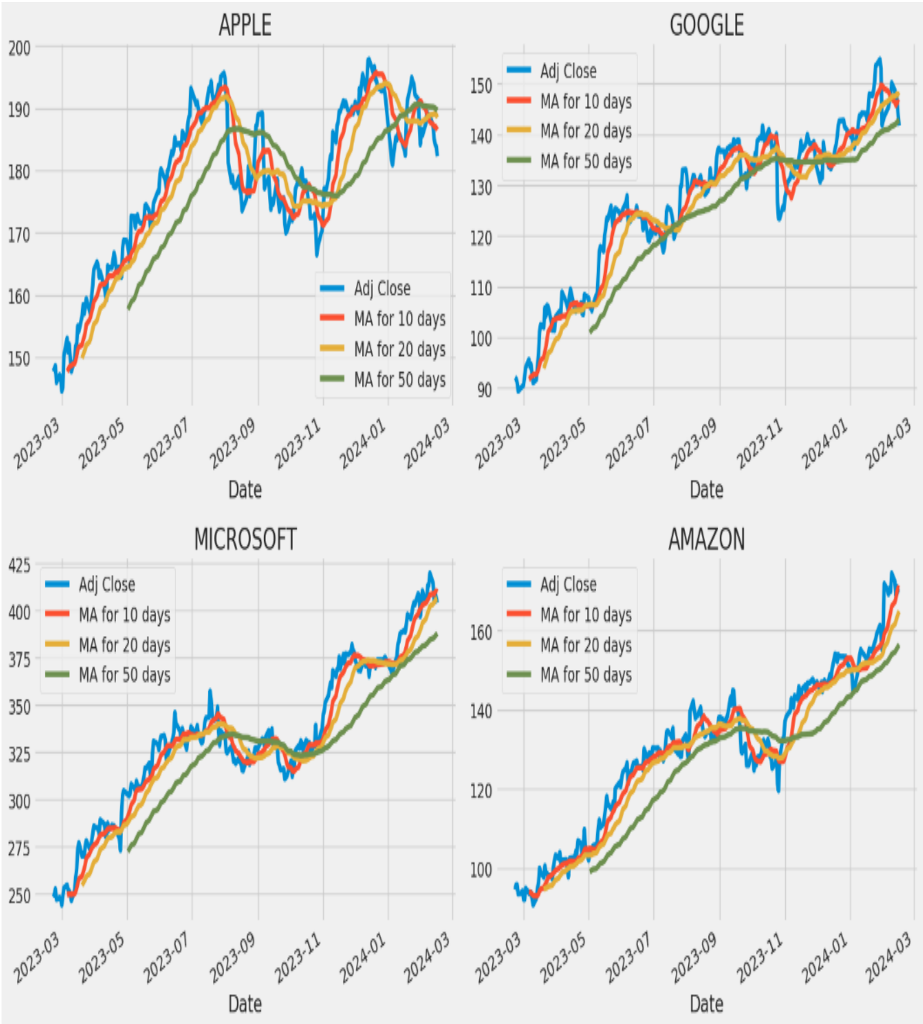


Figure 3. Graph of Moving Averages based 10, 20 and 50 Days.

Daily return of these stocks can also be extracted from Yahoo Finance and can also be visualize as under;

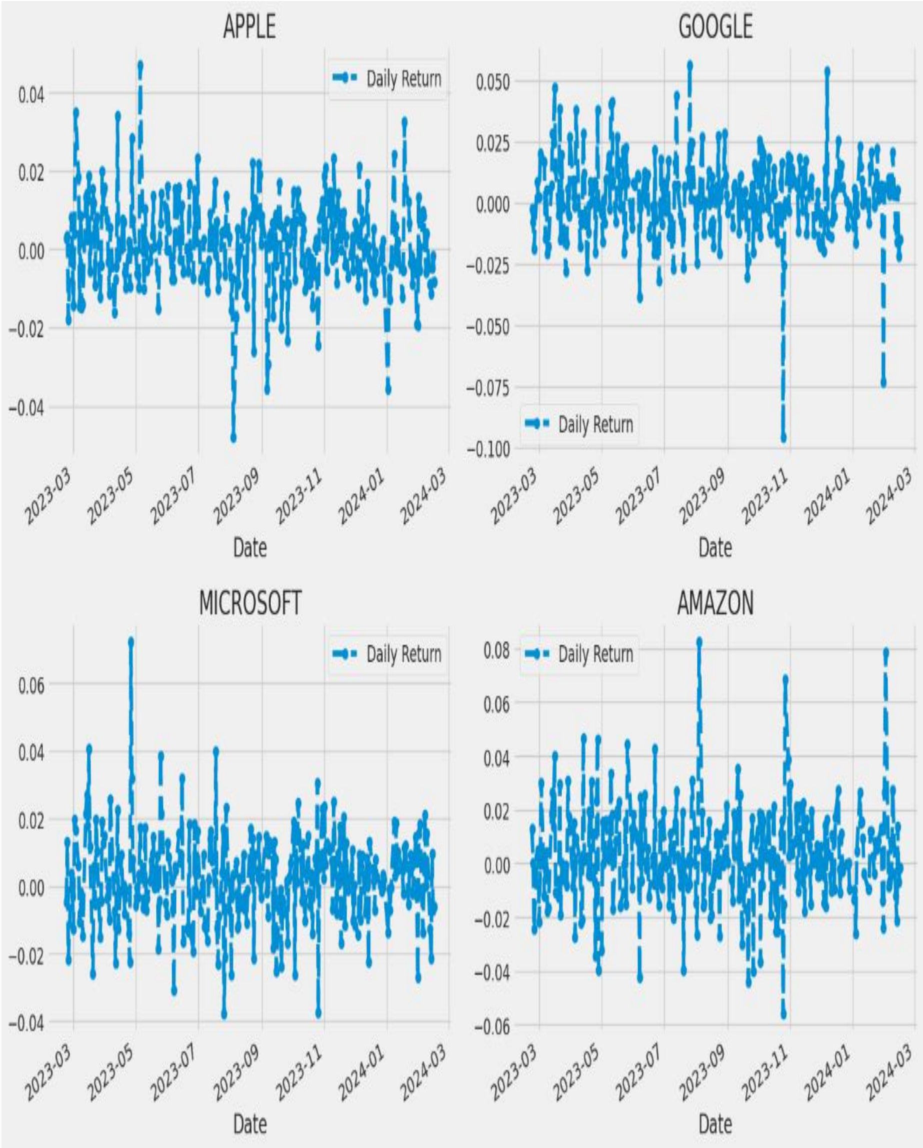


Figure 4. Graph of Daily Return of selected Tech Stocks

For better understanding, Histogram for Daily Return can also be plotted.

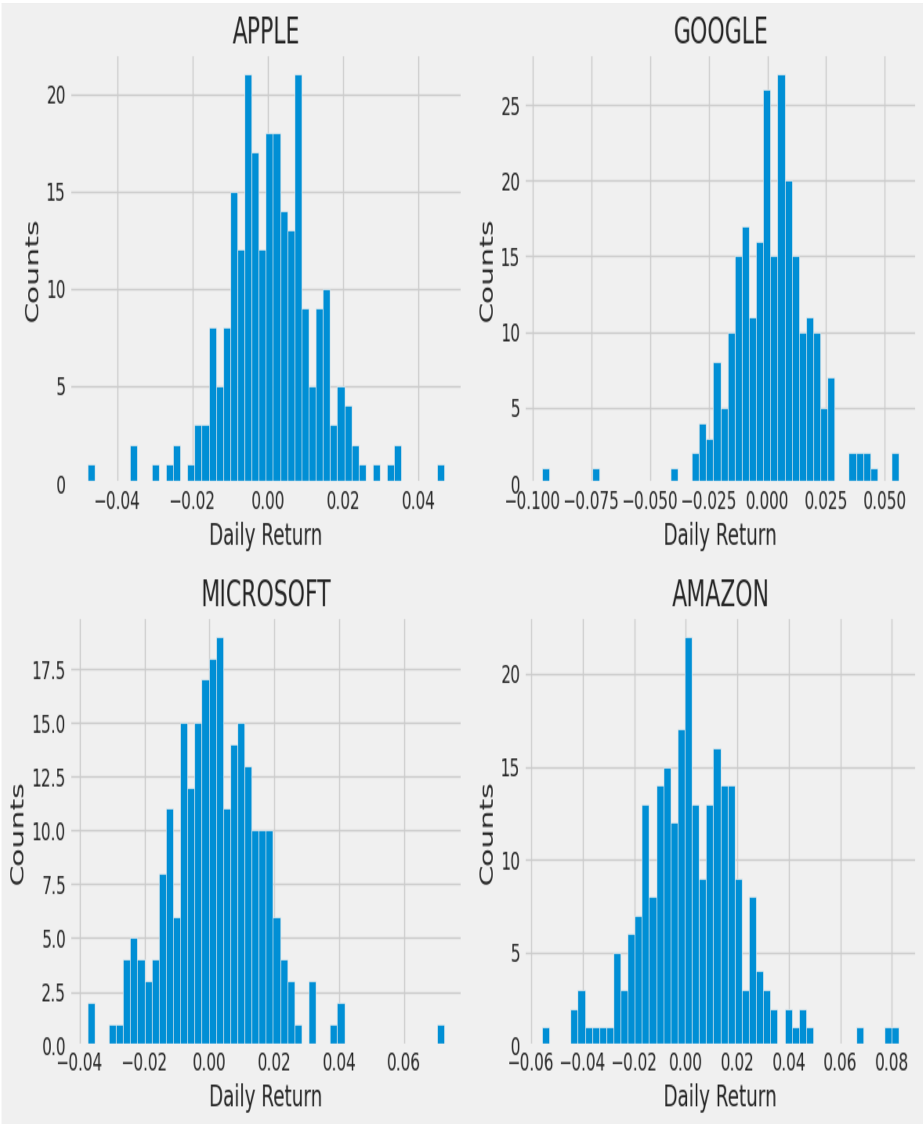


Figure 5. Daily Return Histogram

These stocks belong to same industry, so there are high chances that prices and returns of these stocks are highly correlated.

For example, graph between the returns of Google and Microsoft is plotted to see whether their returns are correlated or nor.

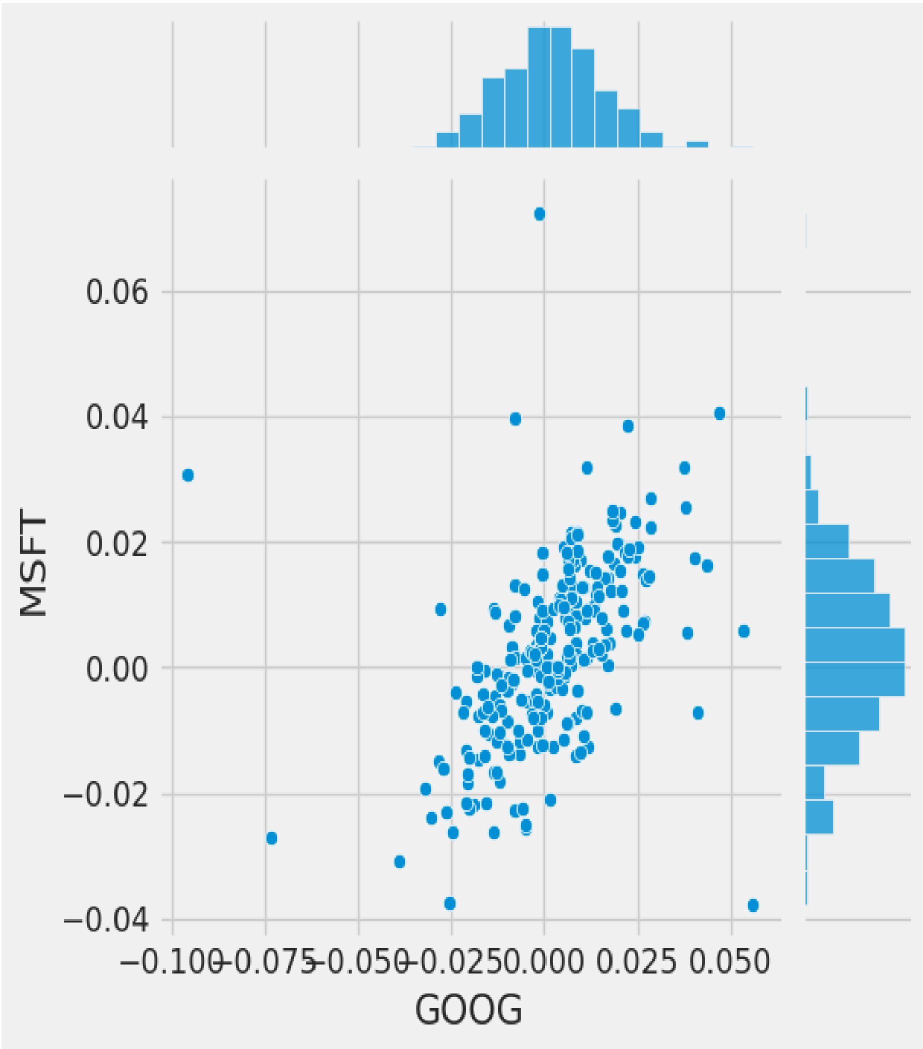


Figure 6. Graph showing positive correlation between returns of Google and Microsoft

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

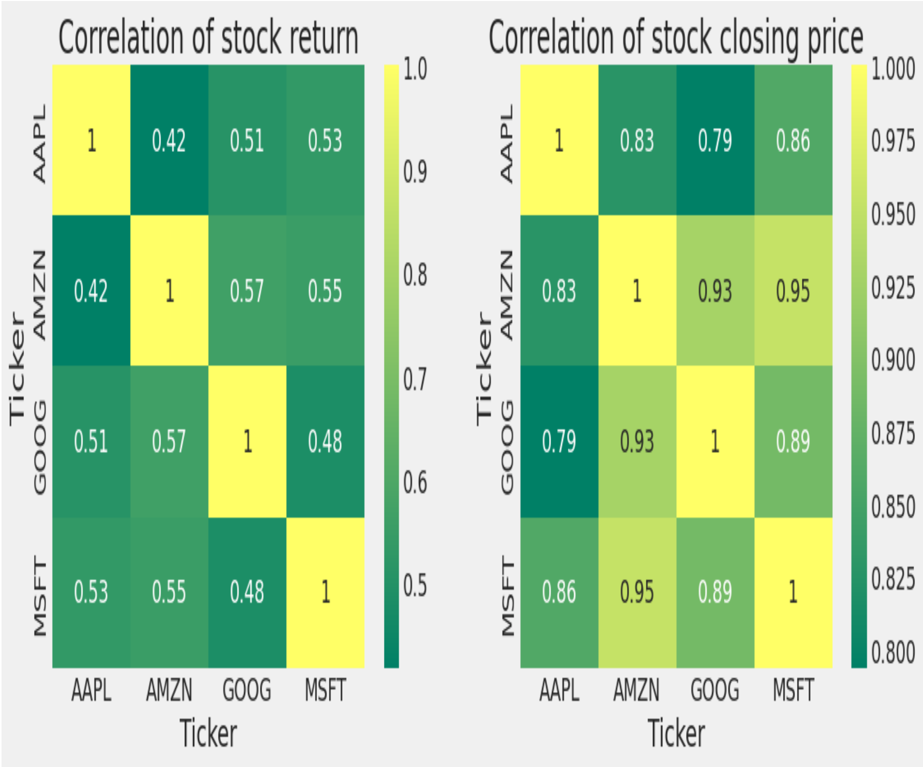


Figure 7. Correlation plot of numerical values of selected stocks

From graph it is clear that if there is positive correlation between prices of two stocks then there is also positive correlation between returns of said stocks. For example, correlation between returns of Microsoft and Amazon is 0.55 likewise correlation between their prices is 0.95. Now such kind of portfolio is riskier as drop in values of one stock can cause drop in other stock. This drop might happen due to various reasons (Economic, Political etc.).

Risk Assessment is also conducted, the graph between expected returns and associated risk of each stock is as under;

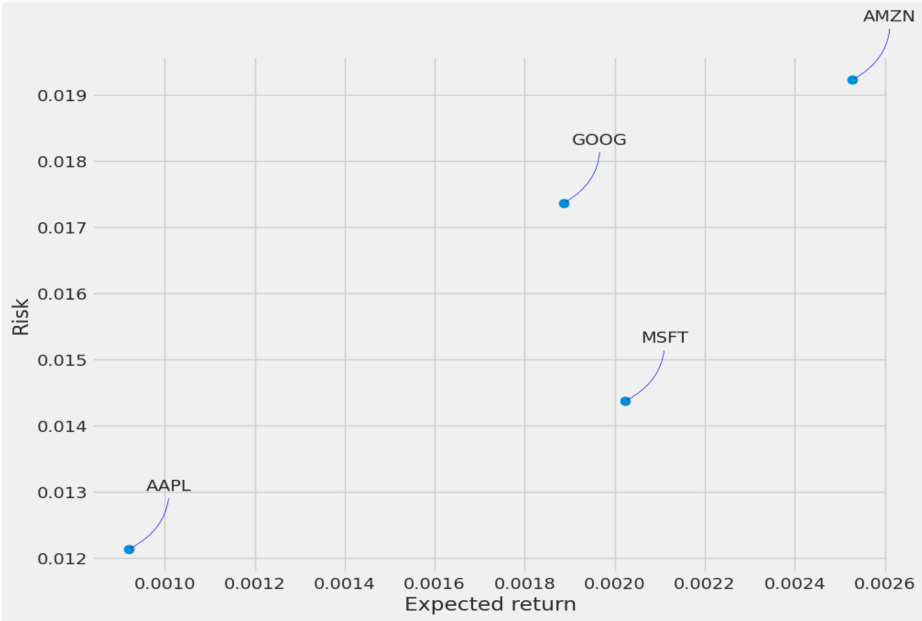


Fig. 1. Risk Assessment: Graph between Risk and Expected Returns

From graph it is visible that Apple is less risky and so its returns are also on lower side. However, Amazon is the most riskier but at the same time its returns are higher than all other selected stocks. Clearly endorsing the concept higher the returns, higher the risks.

3.2 Data Split and Model Creation

For this purpose, stock quotes for Apple company for the period 03.01.2012 till 16-02-2024 is extracted, this dataset consists of 3051 rows and 6 columns (Open Price, High Price, Low Price, Close Price, Adj. Close Price, Volume), the extracted data can be visualized as under;

Fig. 2. Closing Price History showing rising trend in Apple Prices

Split the extracted dataset into training and testing data, for this purpose 2899 rows out of 3051 rows (95% data) is selected for training data and remaining 152 rows are set aside for testing purpose. Data is scaled using MinMax scaler technique into 0s and 1s. This scaled data is used to create training dataset.

Now LSTM model (the basic requirement of this research) is developed using libraries like Keras.Models and Keras.Layers available in Python. Two layers of LSTM with 128 memory cell and 64 memory cells are added in the model. These layers are responsible for learning long-term dependencies in the input sequence.

Similarly, two Dense (fully connected) layers with 25 and 1 neurons respectively added. These layer are used for learning non-linear transformations of output from the proceeding layers. Model is now complied with Adam optimizer, loss is set as Mean Square Error and Epochs are set to 5. After training the model, it is provided with testing data to predict 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 is calculated. These errors are Regression Metrics, most suitable for Time Series or Sequential data type. The quality of RMSE is that it gives the error in units (Hodson, 2022). So, it has increased understandability. For example, RMSE is 5 and dataset in Grams then it estimated or predicted error is 5 Grams.

MAPE stands for Mean Absolute Percentage Error. It is used to measure accuracy of forecasting model, LSTM in this research paper. Both RMSE and MAPE are important for error estimation of a predicting model. Both are equally important and their use depends upon type of dataset. The predicted values pf prices by LSTM model are as under;

Table 1. Comparison of Close and Predicted Prices

	es	Prices
3	190.54	186.7
14.07.2023	190.7	186.83
17.07.2023	193.99	186.99

18.07.2023	193.73	190.45
…		
13.02.2024	185.04	183.59
14.02.2024	184.15	182.13
15.02.2024	183.86	181.59
16.02.2024	182.31	181.26

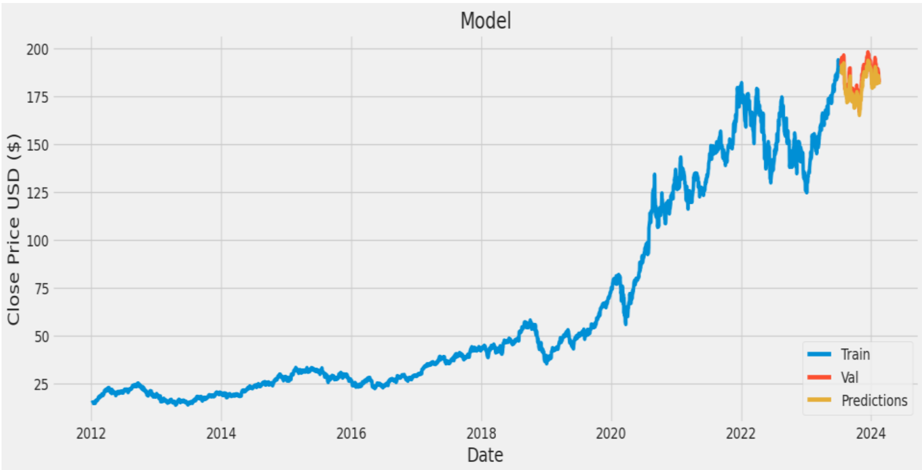


Fig. 3. Visualization of Close and Predicted Prices

The calculated value of RMSE and MAPE (using Python Programming) for our LSTM model are 3.867 and 98.19 % respectively, means that there is a difference of 3.867 Dollars in predicted values and model accuracy is 98.19%.

4.2 Discussion

LSTM model has proven to be one the best model to predict time series dataset. To validate this result let us compare performance of LSTM model with two other famous models (1) Feed Forward Neural Network (FNN) and (2) Recurrent Neural Network. The calculated values of RMSE, MAPE and Weighted Average (giving 50% weights to each error) are available in below table.

Table 2. Comparison of RMSE, MAPE and Weighted Averages

Model	RMSE	MAPE	hted Aver
LSTM	3.87	98.19	0.0171

FNN	183.67	99.49	0.5
RNN	183.63	99.5	0.5

According to above table RMSE of LSTM is far better than both FNN and RNN, the RMSE for LSTM is 3.87 whereas for FNN and RNN it is 183.67 and 183.63 respectively. Hence, keeping in view results of RMSE, LSTM is better model for Time Series Predictions.

If MAPE(accuracy) is taken into account, results of FNN and RNN are same that is 99.50% and hence well performed then LSTM model whose accuracy is 98.19%.

In this situation where one cannot make a decision about which model is best predictor, technique of Weighted Averages, giving equal weights to each metrics, can be used. The model with lessor value of Weighted Average will be declared as the best model. Doing so, it is discovered that combine score or Weighted Average of LSTM is less then both FNN and RNN that is 0.0171.

Hence, LSTM is one of the best option to predict Time Series based stock market data (prices and returns)

5-Conclusion

The nature of stock market data is volatile and dynamic. Investors are always afraid of sudden price fluctuation in negative direction. As such reductions can cause them severe losses. Although there are many traditional, Accounting and Financial Management based techniques available, but they are not free from errors. With the advent of latest computer based techniques like AI Models, Machine Learning and Deep Learning Models, the volatility and fluctuations in stock markets can be predicted more precisely and accurately. Models like LSTM, FNN and RNN can be really helpful in predicting stock prices. However, there is always room for improvement as above results revealed that these models are also not 100% accurate. One can use these models for predictions keeping in mind that the chances of deviation are also present. Endeavors should be made to develop models that can minimize chances of error.

References

[1] Drobetz, W., Hollstein, F., Otto, T., & Prokopczuk, M. (2021). Estimating Stock Market Betas via Machine Learning. *Available at SSRN 3933048*.

[2] Fallahgoul, H., Franstianto, V., & Lin, X. (2024). Asset pricing with neural networks: Significance tests. *Journal of Econometrics*, *238*(1), 105574.
[3] Gao, P., Zhang, R., & Yang, X. (2020). The application of stock index price prediction with neural network. *Mathematical and Computational Applications*, *25*(3), 53.
[4] Gitman, L. J., Juchau, R., & Flanagan, J. (2015). *Principles of managerial finance*. Pearson Higher Education AU.
[5] Gonzales, R. M. D., & Hargreaves, C. A. (2022). 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*(2), 100130.
[6] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251.
[7] Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1–10.
[8] Kevin, S. (2022). *Security analysis and portfolio management*. PHI Learning Pvt. Ltd.
[9] Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659.
[10] Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, *2020*, 1–10.
[11] Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, *170*, 1168–1173.
[12] Mokhtari, S., Yen, K. K., & Liu, J. (2021). Effectiveness of artificial intelligence in stock market prediction based on machine learning. *arXiv Preprint arXiv:2107.01031*.
[13] Najem, R., Amr, M. F., Bahnasse, A., & Talea, M. (2022). Artificial Intelligence for Digital Finance, Axes and Techniques. *Procedia Computer Science*, *203*, 633–638.
[14] Ndikum, P. (2020). Machine learning algorithms for financial asset price forecasting. *arXiv Preprint arXiv:2004.01504*.
[15] Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655.
[16] Pan, S., Long, S. C., Wang, Y., & Xie, Y. (2023). Nonlinear asset pricing in Chinese stock market: A deep learning approach. *International Review of Financial Analysis*, *87*, 102627.
[17] Peykani, P., Sargolzaei, M., Sanadgol, N., Takaloo, A., & Kamyabfar, H. (2023). The application of structural and machine learning models to predict the default risk of listed companies in the Iranian capital market. *Plos One*, *18*(11), e0292081.

[18] Pokhrel, N. R., Dahal, K. R., Rimal, R., Bhandari, H. N., Khatri, R. K., Rimal, B., & Hahn, W. E. (2022). Predicting nepse index price using deep learning models. *Machine Learning with Applications*, *9*, 100385.

[19] Rubi, M. A., Chowdhury, S., Rahman, A. A. A., Meero, A., Zayed, N. M., & Islam, K. A. (2022). Fitting multi-layer feed forward neural network and autoregressive integrated moving average for Dhaka Stock Exchange price predicting. *Emerging Science Journal*, *6*(5), 1046–1061.

[20] Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y.-H. J. (2020). Machine learning stock market prediction studies: Review and research directions. *Journal of International Technology and Information Management*, *28*(4), 63–83.

[21] Iqbal, Muhammad Javaid, Usama Ijaz Bajwa, Ghulam Gilanie, Muhammad Aksam Iftikhar, And Muhammad Waqas Anwar. "Automatic Brain Tumor Segmentation From Magnetic Resonance Images Using Superpixel-Based Approach." Multimedia Tools And Applications 81, No. 27 (2022): 38409-38427.

[22] Iqbal, Muhammad Javaid, Muhammad Waseem Iqbal, Muhammad Anwar, Muhammad Murad Khan, Abd Jabar Nazimi, And Mohammad Nazir Ahmad. "Brain Tumor Segmentation In Multimodal MRI Using U-Net Layered Structure." Comput Mater Contin 74, No. 3 (2022): 5267-5281.