iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42239

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Stock Market Analysis and Prediction

Paresh Shrikhande[1], Raghu Ramani[2], Rushi Bhalerao[3]
[1, 2, 3]*Final Year Undergraduate students, Department of Electronics and Telecommunication Engineering, Pune Institute of Computer Technology*

Abstract: ***Stock price Analysis & Prediction is a one of the sought after and popular topic throughout the last decade. By using machine learning and deep learning (RNN & LSTM) methods to make stock price prediction using real time data. While using Deep learning functions to predict and analyze stock prices are becoming more prevalent in these recent days. Its observed and assumed that machine learning as well as deep learning methods with RNN and LSTM could produce accurate results in stock price prediction. That is why we would like to try our own methods for this project. We have used a number of stocks from the S&P 500 namely as inputs and target. The data will contain open, high, and low, close, adjacent close and volume as its 6 variables. The adjacent close of each stock will become the target, and rest of the variable for a particular stock as well as all the other variables of other respective stocks will become as inputs.***

Index Terms: RNN: Recurrent Neural Network, S&P 500: Standard and Poor's 500, LSTM Long short-term memory, DSP: Digital Signal Processing.

I. INTRODUCTION

The prediction of stock prices in a stock market is an interesting research area for both researchers and investors. A lot many forecasting models have been proposed and are under research to build a prediction model which could forecast stock prices precisely to better understand the market behaviour, make profitable investments and trade. When a client needs to study and assess the potential gains before buying or selling the shares. to solve this problem, we come up with software that could analyse the data and provide insightful information to the client. Stock price prediction is a challenging task owing to the complex patterns behind time series. Autoregressive integrated moving average model and backpropagation neural network model are popular linear or nonlinear models for time series forecasting the integration of two models can effectively capture the linear and nonlinear patterns hidden in time series. Financial Indices Modelling and Trading utilizing deep learning techniques, Marios, Spiridon Likothanassis. Prediction and modelling of the financial indices are challenging and equally demanding problems because of their dynamic nature. Modern approaches have also been challenging the fact that they are dependencies between different global financial indices. All this complexity in combination with the large volume of historical financial data raised the need for advanced machine learning solutions to this challenge.

II. THEORETICAL BACKGROUND

There are 505 companies in the S&P 500. The data set that we had chosen includes all stock parameters values like Open, High, Low, Close, Adj Close price and trading Volume. The data ranges between 31/07/2001 and 22/12/2015. Our data set contains 3030 columns while values in 564 columns are missing. This might have been a case of company listings after the year 2001. We decide to drop this column containing missing values since neural networks cannot be fed with such missing values as inputs.

Outliers can cause obstruction in modelling the data. Many of these outliers may be present due to volatile price series. No outliers are found in the target series as a boxplot is used to check on the outliers. Visualization in the target series shows an upward trend which is common for most of the stock prices in the S&P 500. The financial crises of the year 2008 can be seen by a huge drop in the series. If statistical analysis is used then a lot of statistical tests and data processing is required since the dependent and independent variables that we have used must be stationary or need to be co-integrated. The better approach to this is by using neural networks where we can use raw time series.

The initial step is the feature selection data processing methodology. While data visualizing the first 200 variables had higher correlation coefficient than target variables. Our target variable is the variable having its correlation coefficient as one. The second step is to shift the data set by one day ahead. Setting up End and start times for data grab. This is done to turn the corresponding time feature into lagged features and split the data for training and testing purposes. The prices for different stocks are recognised on different scales. The prices of a stock and volumes are also on different scales. Scaling of the inputs is done to make the neural network converge faster. We have used a Min-Max scaler for scaling inputs so that their range comes under the range of 0-1 while training the dataset. The scaling process is split to avoid the information leakage that could have leaked testing period data into training period data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

III. TECHNICAL SPECIFICATIONS

The project includes the use of neural network models using Keras module and TensorFlow as backend. The neural network uses the LSTM approach to predict the closing prices of four stocks considered here. The steps included in the implementation of the project were to calculate - change in price of stock overtime, moving average of various stocks, daily return of the stocks on average, correlation between different stocks' closing prices, risk in investing in a particular stock and finally predicting closing prices of different stocks based on previous data.

The LSTM network used has 2 layers having 100 and 50 dimensions of hidden units respectively. We use dropout layers after each LSTM. The final layer is the output dense layer. The final layer consists of 1 neuron with a linear activation function. The project uses Adam Optimizer with Mean Squared Error as loss function. The batch size used here is 2175 which is used for training the data under 1 epoch. The validation set is set from the training set randomly. Overfitting can be prevented by keeping the track of validation loss.

IV. INPUT DATA

The real-time data for 25 stocks is used as input to the model, which is retrieved from Yahoo Finance. This real-time data is fed to seaborn (python library) and the data changes for these stocks are visualised over the past year. This data is visualised over the date exactly one year back till now.

V. ANALYSIS OF DATASET

A. Closing Price and Sales

The dataset used is the real-time data fetched from the Yahoo finance server. It gives the data for number of stocks namely constrained over a time period of 1 year. The data is then analysed with the help of matplotlib, pyplot library of python. The stock prices of the companies viz. Tesla, PepsiCo, Cisco and Marriott International are also visualised over the total volume of stock being traded each day, i.e., sales volume of the company stocks.

Fig 1 Closing price

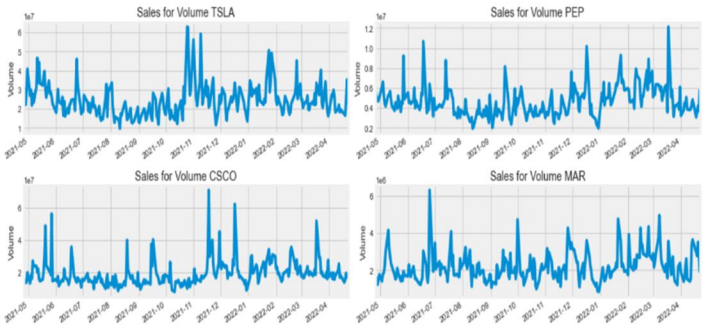


Fig 2 Sales

B. Calculating Moving Average

Now that we saw the visualizations for the closing price and the volume traded each day, we went ahead and calculated the moving average for the 4 stocks. The moving averages for the stocks viz. Mondelez International, Netflix, NVIDIA and Starbucks are visualised over three sets of time periods of 10 days, 20 days and 50 days using the matplotlib.pyplot library of python.

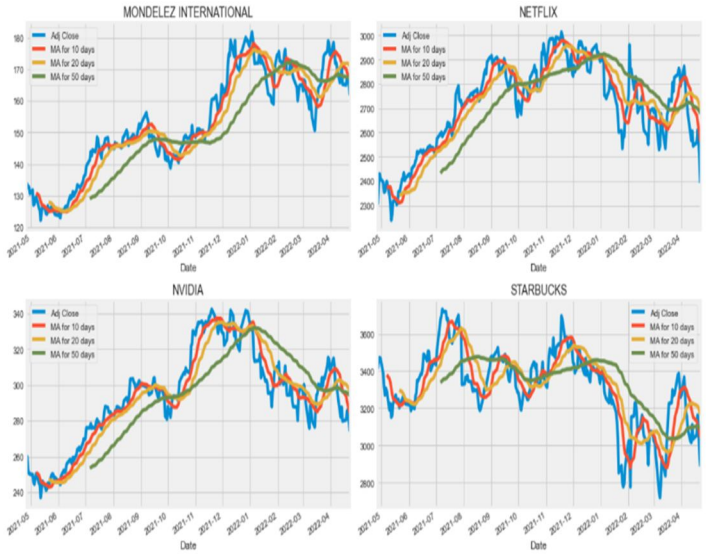


Fig 3 Moving Average

C. Retrieve Daily Returns

After doing some baseline analysis, we went ahead and dived a little deeper. We now analyzed the risk of each stock. In order to do so we needed to take a closer look at the daily changes of the stock here taken Texas Instruments, Electronic Arts, Intel and Johnson & Johnson, and not just its absolute value. We use the pandas framework to retrieve the daily returns of a single stock. We use the pct_change() function to calculate the percentage change for each day. Henceforth, plot() function is used to obtain the graphs of the daily returns of each stock. A look at the average daily return is visualised through histogram of the 4 stocks namely, Ford Motor Company, Bank of America, Walmart and Procter & Gamble Company.

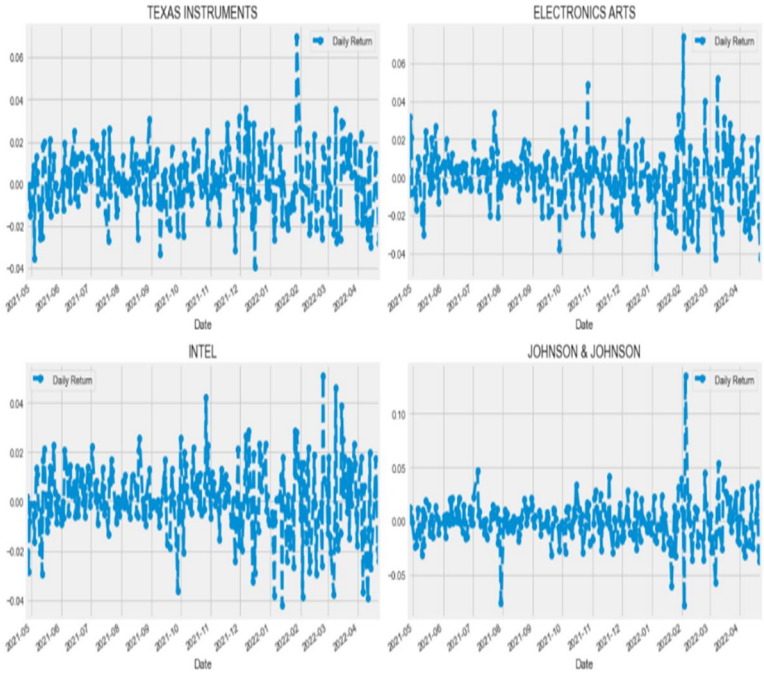


Fig 4 Daily Returns

To create the histogram and kde plot of the same stock, seaborn library of python is used, along with the help of dropna() function as seaborn is unable to read NaN values.

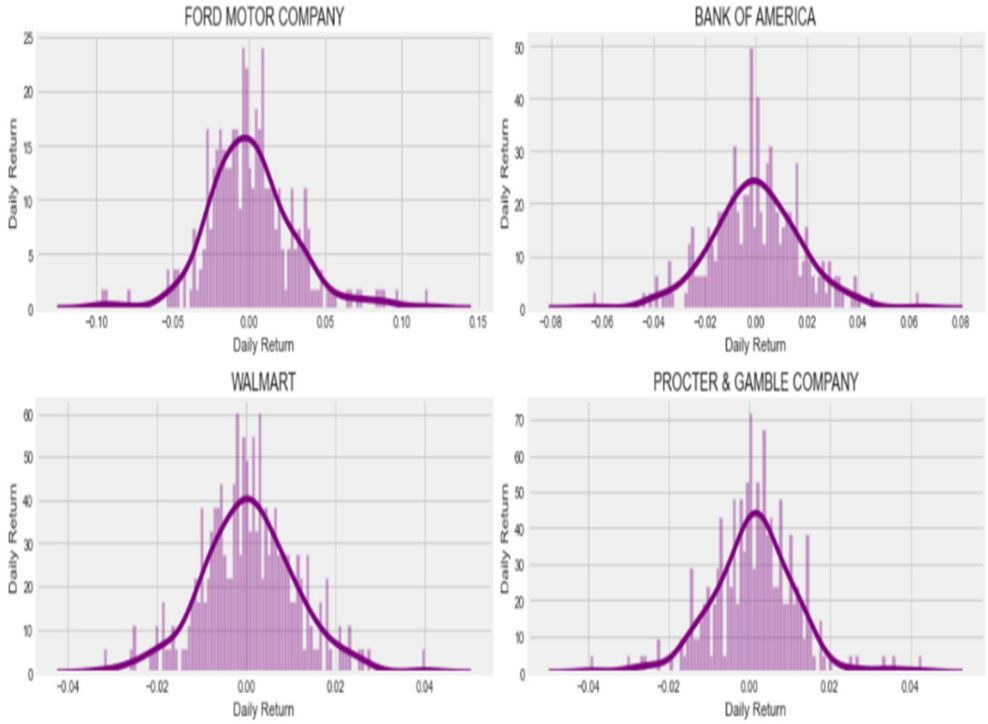


Fig 5 Daily Returns (kde plot)

D. Correlation Between Stocks

To analyse the returns of all the stocks in our list, we built a dataframe with all the [‘Close’] columns for each of the stocks dataframes. Similarly, another fresh dataframe is created having daily returns of all the stocks. Firstly comparing Google stock to itself, we obtain a perfectly linear relationship.

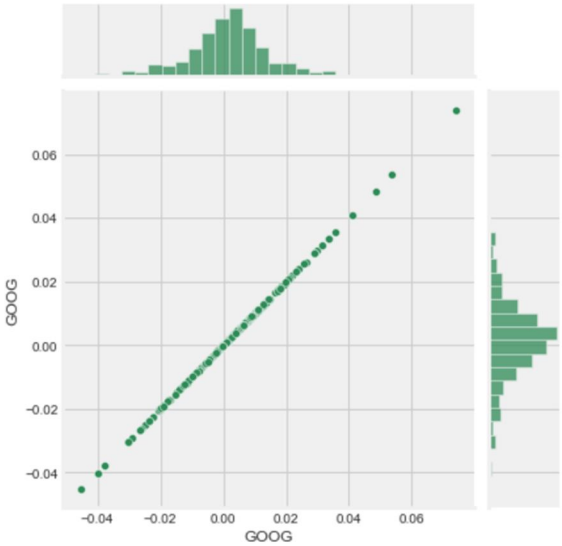


Fig 6 Correlation of Stocks

Now, we use joinplot() function to compare to compare the daily returns of Google and Microsoft. So now we can see that if two stocks are perfectly (and positively) correlated with each other a linear relationship between its daily return values should occur.

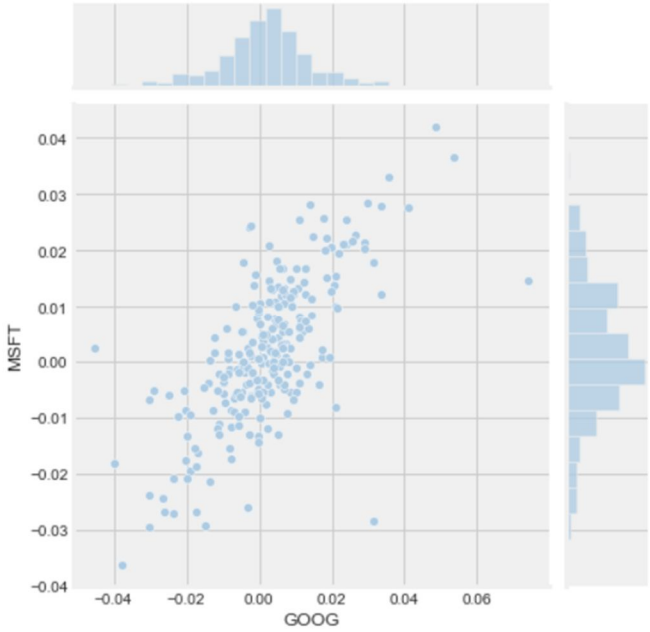


Fig 7 Daily returns of Google and Microsoft

E. Calculating Risk

There are many ways we can quantify risk, one of the most basic ways using the information we've gathered on daily percentage returns is by comparing the expected return with the standard deviation of the daily returns. So, a new dataframe is defined as a cleaned version of the original dataframe (i.e. without the NaN values). A scatter plot (risk v/s expected values) of values of the stocks is generated using the mean and the standard deviation of the values of the stock prices.

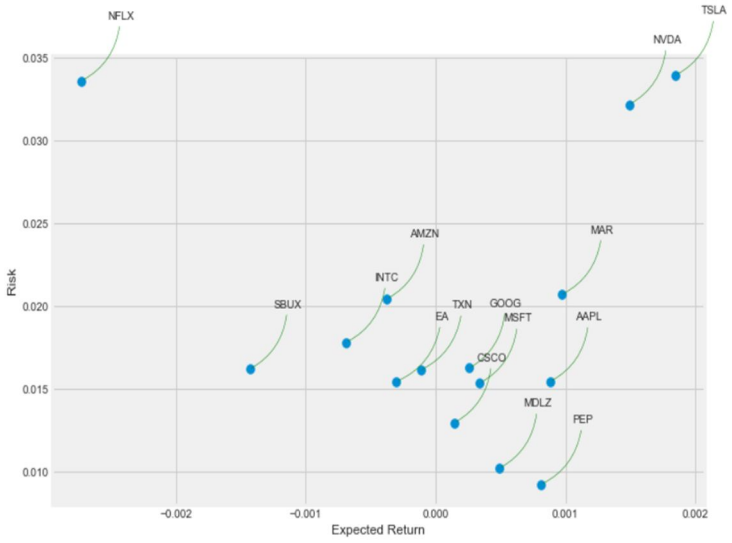


Fig 8 Calculating Risk

F. RNN

Also known as RNNs, Recurrent Neural Networks come under the class of neural networks which contain hidden states allowing previous data to be used as a feed to the model (i.e. for inputs). This gives us the advantage of processing input strings of any length. Along with their computation being slow, the major disadvantage of using RNNs is that there is a difficulty of accessing information which has been input a long time ago.

G. LSTM

To solve the short memory problem of RNNs, Long Short Term Memory (LSTM) architecture is used. LSTMs are the special kind of RNNs which are capable of learning long-term dependencies. A set of recurrently connected blocks known as memory blocks are present in a LSTM layer. One or more recurrently connected memory cells along with three multiplicative units- input, output and forget gates.

H. Predictions

This is the other main subpart of the project- Predicting the closing prices of stocks, taking into account the past values of closing prices for the time period from 2012 to 2021.

Firstly, we analyse and visualize the previous closing prices of the stock, here taking Netflix for reference. The below graph shows the pictorial representation of the changing trends in the closing prices of Netflix stock over the years.

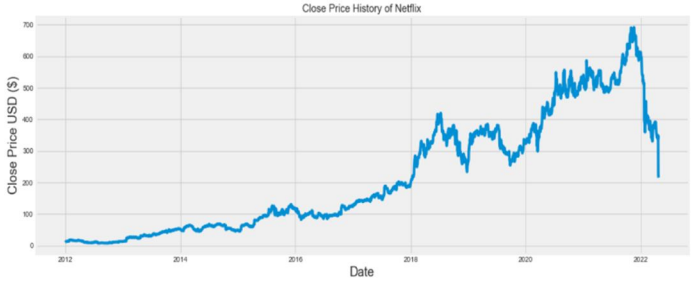


Fig 9 Close price history

Now, to predict the values, we use the LSTM model. We use dropout layers after each LSTM. The final layer is the output dense layer. The final layer consists of 1 neuron with a linear activation function. The project uses Adam Optimizer with Mean Squared Error as loss function. The batch size used here is 2175 which is used for training the data under 1 epoch. The validation set is set from the training set randomly. Overfitting can be prevented by keeping the track of validation loss.

Seaborn and pandas make it very easy to repeat this comparison analysis for every possible combination of stocks in our technology stock ticker list. We can use sns.pairplot() to automatically create this plot.

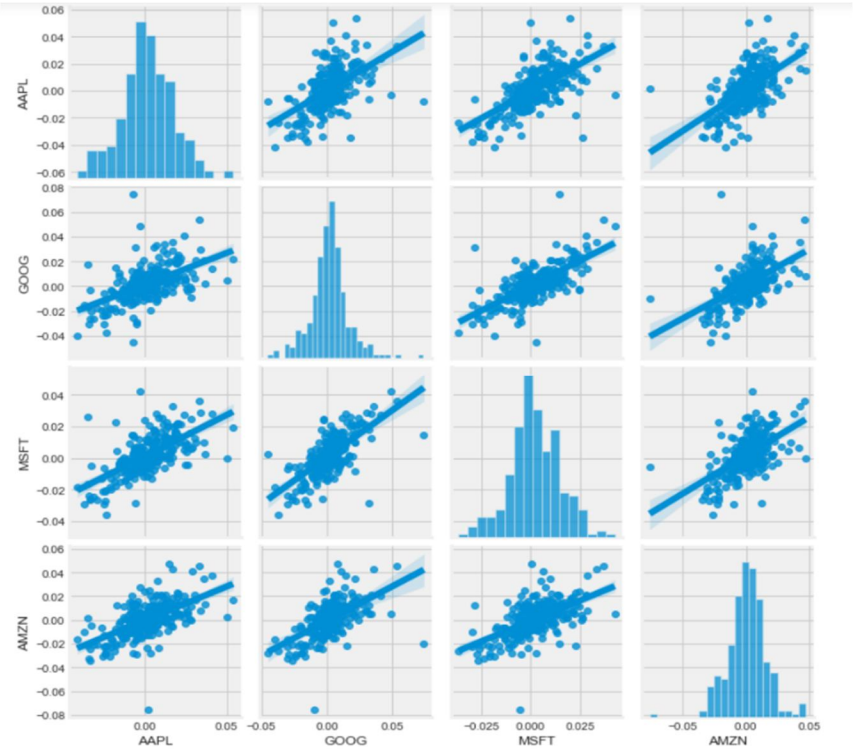


Fig 10 Relationship of Daily Returns of all stocks (sns.pairplot)

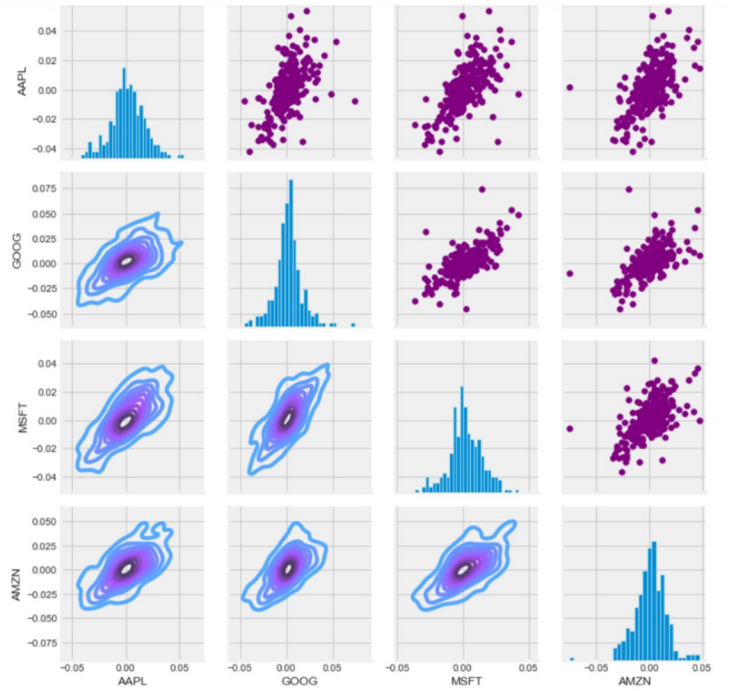


Fig 11 Relationship of Daily Returns of all stocks (sns pairgrid)

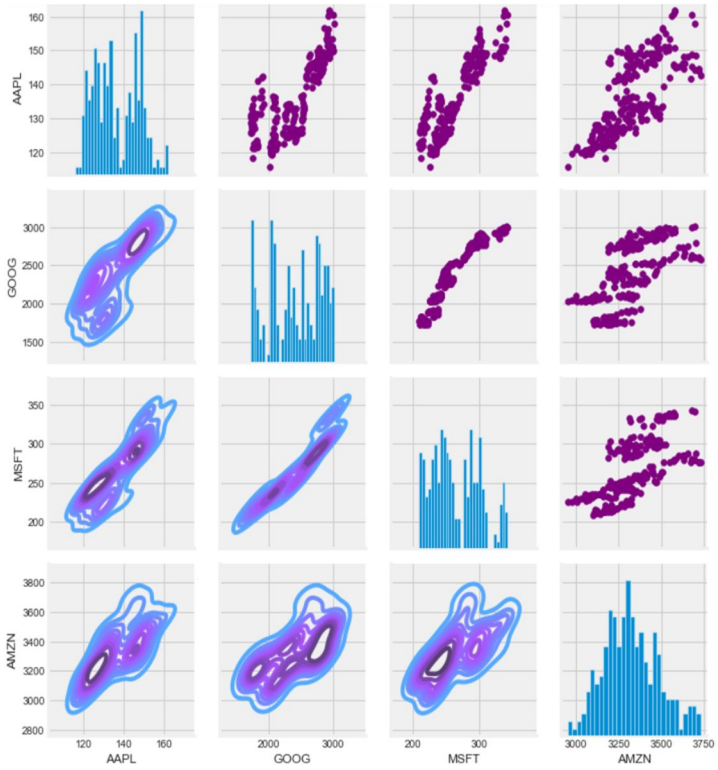


Fig 12 Relationship of Daily Returns of all stocks

Now, could also do a correlation plot, to get actual numerical values for the correlation between the stocks' daily return values. We use heatmap to compare these values. By comparing the closing prices, we see an interesting relationship between Microsoft and Apple.

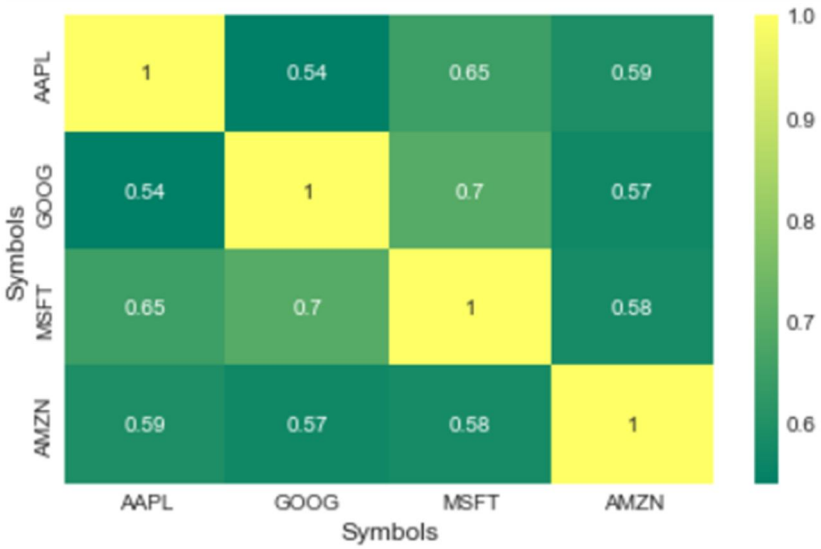


Fig 13 Relationship of Daily Returns of all stocks (Heatmap)

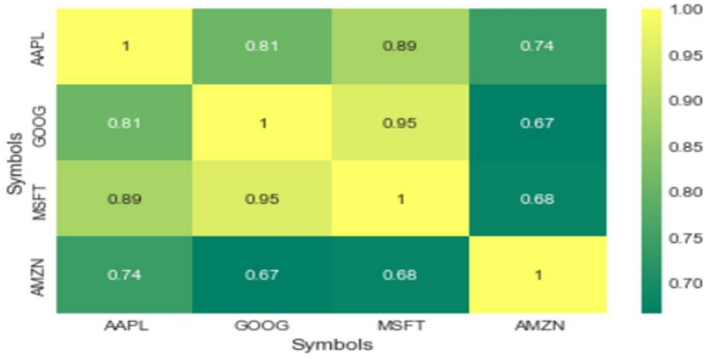


Fig 14 Closing prices of all stocks(Heatmap)

Finally, to predict the closing prices, we use Netflix stock as an example. We firstly print the table containing all the information of Netflix Stock over the years form 2012 to 2021.

Now, the predicted graph for the Netflix stock is visualized as below. The graph below shows the trends of the Netflix stock with respect to the training prices (in blue), validation prices (in red) and the predicted prices (in yellow) over the time period.

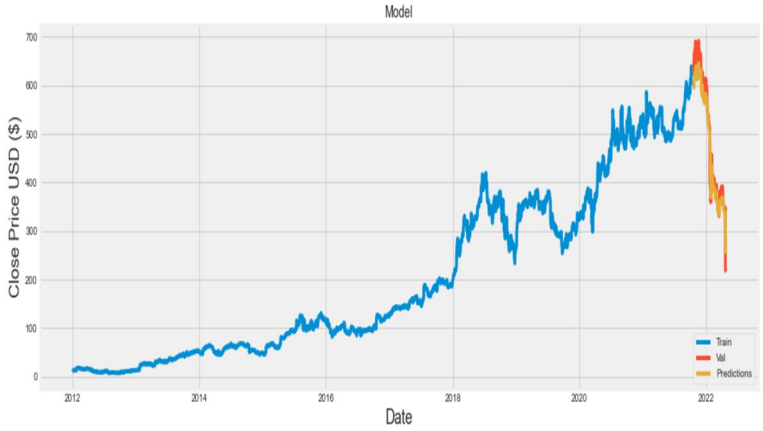


Fig 15 Prediction for Netflix

Here, we see the tabular form of the actual closing prices and the predicted closing prices for the Netflix stock over the referred time period.

Date	Close	Predictions
2021-10-19	639.000000	598.838135
2021-10-20	625.140015	601.520386
2021-10-21	653.159973	595.754700
2021-10-22	664.780029	603.720520
2021-10-25	671.659973	616.736084
...	...	...
2022-04-18	337.859985	339.360199
2022-04-19	348.609985	336.197113
2022-04-20	226.190002	337.711548
2022-04-21	218.220001	293.223541
2022-04-22	215.520004	252.840500

Table 1 Closing Prices and Prediction of closing prices for Netflix

VI. CONCLUSION

The neural network has an interesting section namely convergence of loss function. This function or convergence gives us an idea if the model is successfully run, showing overfitting, underfitting, or normal status. From the graphs given below, it can be observed that epoch number is inversely proportional to the loss of training and validation set. The increase in an epoch indicates the successful run of the model.

REFERENCES

[1] International Research Journal of Engineering and Technology (IRJET) e-ISSN: Stock Price Prediction Using Long Short Term Memory 2395-0056 Volume: 05 Issue: 03 | Mar-2018
[2] European Journal of Molecular & Clinical Medicine ISSN 2515-8260: Volume 07, Issue 02, 202: Sentimental Analysis of stock market
[3] Hegazy, O., Soliman, O.S., Salam, M.A. Machine learning model for stock market prediction: (2014).
[4] Khan, Z.H., Alin, T.S., Hussain, M.A.International Journal of Computer Applications. Price prediction of share market using artificial neural network (ANN). (2011).
[5] Shah, V.H. Foundations of Machine Learning.(2007)
[6] Choudhry, R., Garg, K. World Academy of Science, Engineering and Technology, machine learning system for stock market forecasting. (2008)
[7] Xingjian, S.H.I., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., and Woo, W.C. (2015).Convolutional LSTM network: A machine learning approach for precipitation neural information processing systems:
[8] Zhu, Maohua, et al.Training Long Short-Term Memory With Sparsified Stochastic Gradient Descent. (2016)
[9] Understanding LSTM networks: A blog by colah
[10] Soulas, Eleftherios, and Dennis Shasha. machine learning algorithms for stocks and currency exchange prediction
[11] Mittal, A., Goel, A. Stock prediction using Twitter sentiment analysis. Standford University, CS229 (2012).
[12] Shen, S., Jiang, H., Zhang, T. Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford. (2012).
[13] Nunno, L. Stock market price prediction using linear and polynomial regression models. The University of New Mexico. (2014)

Leading journals to complete their grades. In addition, the published research work also provides a big weight-age to get admissions in reputed varsity. Now, here we enlist the proven steps to publish the research paper in a journal.

Identify the constructs of a Journal – Essentially a journal consists of five major sections. The number of pages may vary depending upon the topic of research work but generally comprises up to 5 to 7 pages. These are:

1) Abstract
2) Introduction
3) Research Elaborations
4) Results or Finding
5) Conclusions

In Introduction you can mention the introduction about your research.

IDENTIFY, RESEARCH AND COLLECT IDEA

It's the foremost preliminary step for proceeding with any research work writing. While doing this go through a complete thought process of your Journal subject and research for it's viability by following means:

1) Read already published work in the same field.
2) Goggling on the topic of your research work.
3) Attend conferences, workshops and symposiums on the same fields or on related counterparts.
4) Understand the scientific terms and jargon related to your research work.

CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

APPENDIX

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

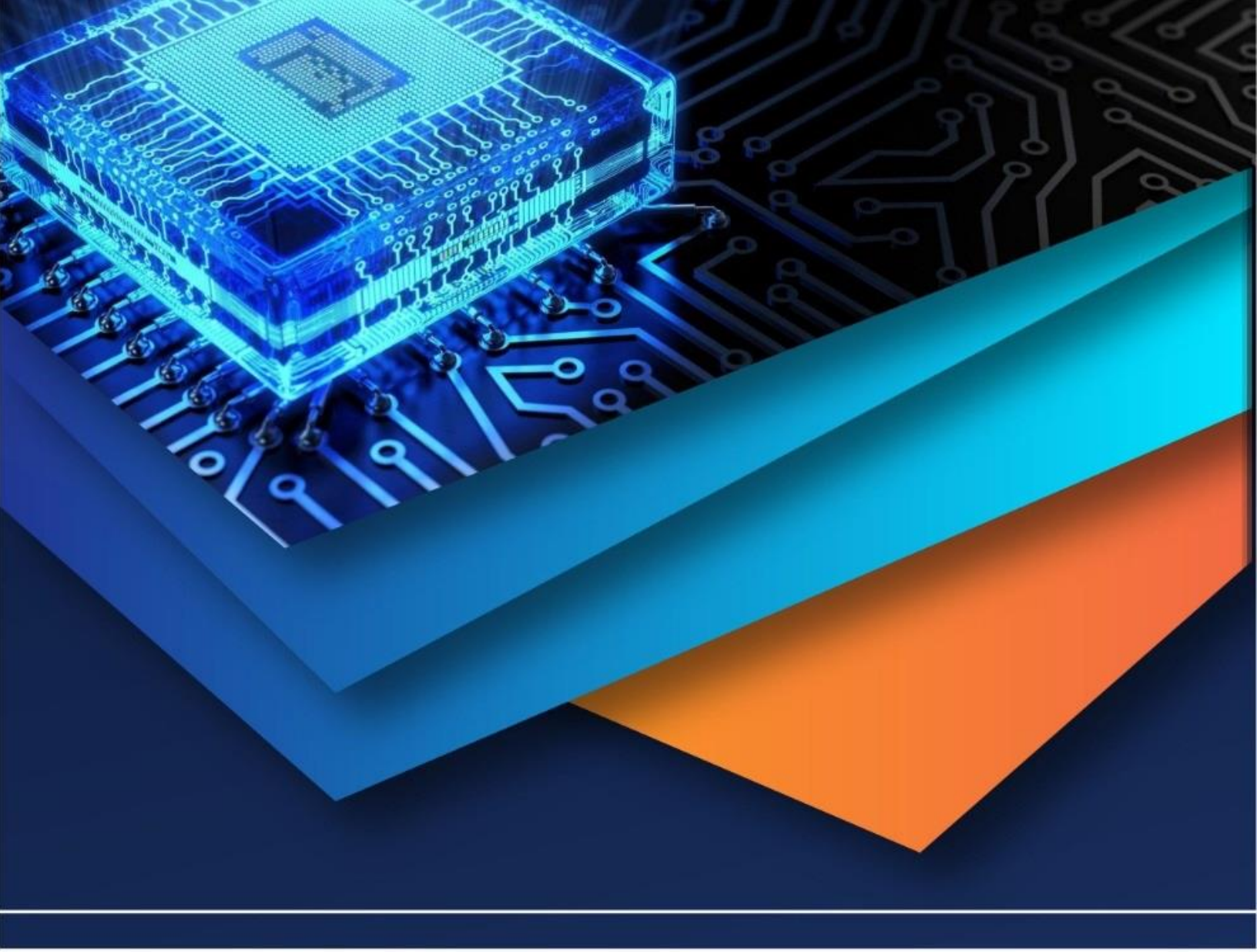
The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

REFERENCES

[1] G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
[3] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
[4] B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
[5] E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” *IEEE Trans. Antennas Propagat.*, to be published.
[6] J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” *IEEE J. Quantum Electron.*, submitted for publication.

AUTHORS

First Author – Author name, qualifications, associated institute (if any) and email address.
Second Author – Author name, qualifications, associated institute (if any) and email address.
Third Author – Author name, qualifications, associated institute (if any) and email address.
Correspondence Author – Author name, email address, alternate email address (if any), contact number.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY