**Second Review Document**

**STOCK MARKET TREND PREDICTIONS AND ANALYSIS USING DEEP LEARNING MODELS AND VISUAL REPRESENTATION**

**Sanchita Suryavanshi**

**18BCE0506**

**+91 7715992021**

***sanchitasanjay.s2018***

***@vitstudent.ac.in***

**Karen Pinto**

**18BCE0596**

**+91 7022565882**

***karen.pinto2018***

***@vitstudent.ac.in***

**Mr. Ramani S**

**Assistant Professor Sr. Grade 1**

**ramani.s@vit.ac.in**

**B.Tech.**

**in**

**Computer Science and Engineering**

**School of Computer Science &amp; Engineering**



**Abstract**

Most of the daily stock market traders lose their money due lack of expertise and understanding of the movement of the market, so through this project we want to bring awareness to the people who are new and not well versed with trading on the stock market. Since, it is impossible to keep track of the performance of stocks and shares in the market every hour, every minute and every second most traders lose money. Predicting stocks has been an old age problem which is yet to be resolved. Few of the prediction systems were implemented but the problem of stocking up the data and updating data based on time periods became a challenging task. Hence, we are planning to solve this issue by implementing various deep learning models and comparing them to determine which algorithm would suit perfectly according to the situation with the data taken from Yahoo Finance which is one of the most authentic websites with regard to stock data. This project will help a layman to understand the data pattern as complex as that of a stock market. Algorithms such as LSTM, Basic RNN & GRU along with their bi-directional versions will be the ones that would be used for prediction of stocks in the market. Some agents which are stock market trading strategies will also be used to help investors to invest by predicting stock market trends. Monte Carlo and Joint Plot Visualization are used to depict the performances of various stocks. Sentiment Analysis is used to understand financial news and make decisions on stock market investments.

***Keywords:***

* *Recurrent Neural Network,*
* *LongShort Term Memory,*
* *Gated Recurrent Units,*
* *Turtle Agent,*
* *Moving Agent,*
* *Monte Carlo Simulation*
* *Joint plot Simulation*
* *Sentiment Analysis*

1. **Introduction**

**1.1 Theoretical Background**

The stock market prediction is almost a 50 years old problem but the importance of it has only risen through time. The institution involved in stocks business and regulating authority requires it today more than ever, due to high dynamics and increasing liquidity of the market, the predictions have turned out to become difficult and only difficult by each and every passing day. But the data stack has also increased by tenfold and is also taken into care, stored in storage spaces controlled by the regulating authorities of the stock markets. The stock market has two very important objectives. The first is to provide funding. It is a measure of a company's revenue and its ability to meet short-term obligations and the performance of a business fund. The second purpose of the stock market is to give investors - those who buy stocks - the opportunity to participate in the profits of publicly traded companies. The best profit is for companies to be able to use it to fund and expand their businesses.

A large number of people are investing in the stock market, but without accurate knowledge of the styles they end up losing a large portion of their investment quickly.Generally the traders because of lack of experience and understanding lose their cash on daily basis, so through this project we are willing to create exposure for the layman in this field. The stock market determines the future movement of the value of a stock exchange. Accurate forecasts for stock prices will lead to additional profits that investors can make. These days, almost all well-established companies rely on the use of speculative models on a regular basis. Here, the goal is to help investors and small traders to invest wisely to make a good profit in order to maintain the shire monopoly in the market.

We have come up with a solution for ordinary traders who understand market movements and predict positive results for sure. We have developed a systematic approach using existing solutions where people having minimum knowledge in this field can make investments by seeing the graphs and visualizations . This project will use deep learning models to predict the future outcomes of the stock market, by analyzing and finding the pattern in the data set provided.

**1.2 Motivation**

Maintaining stocks in the stock market is equally challenging and it is impossible to have direct and accurate market updates every second as it takes a second for the market to fluctuate. Therefore, to solve this problem it is time to implement a solution using machine learning and in-depth learning algorithms. The reason for choosing this topic is the demand as most of the people wish to generate passive income by investing some parts of their active income in share markets. This project aims to provide such people the platform on how they should approach investing in stocks. Nowadays, multinational corporations and tech giants have well established prediction models that encourage them to compete with their counterparts. The regular traders who are not very serious, keep losing money on a daily basis in the market which discourages them, eventually quitting the trading. Hence, for building the confidence in people we need to have one certain system which can predict the movement of scrip so that everyone using the system can make money out of the markets and when they lose they don’t lose big but when they gain they must gain a big amount of money. Therefore, here we are coming up with a solution for regular traders which understands the market movement and predicts a certainly good outcome.

**1.3 Aim of the proposed Work**

* To identify factors affecting the share market.
* To generate a pattern from a large set of data of the stock market for prediction of future stock market movements.
* To predict whether the stock price is going to increase or decrease or steady in the next 30 days.
* To guide users as to when to buy or sell stocks.
* To help users interpret the prediction trends using visualization techniques.

**1.4 Objective(s) of the proposed work**

* Our project is working on the principles of macroeconomics and also on some parameters of Deep Learning. Some commonly used terminology will be used in our project which is used in the stock market.
* The project will consist of a mix of three models which are :

1. RNN
2. GRU & BI-DIRECTIONAL GRU
3. LSTM & BI-DIRECTIONAL LSTM

* The comparison of accuracy of these models will tell us which is suitable for the given dataset.
* The aim of this project is to use agents which will help users to predict stock trends and guide users by notifying when to sell and buy a particular stock.
* To make it easier for a layman to understand this project also has data visualization models like Monte Carlo Simulation and Joint Plot Visualization.

1. **Literature Survey**

**2.1 Survey of the Existing Models/Work**

Ghosh, A and Bose investigated data analysis as a game changer in the stock market domain. According to the author, ML techniques have the ability to find patterns and details that can be used to make accurate predictions. The LSTM model (Long-Term Memory) and the corporate growth algorithm were used to analyse and predict the company's future growth.[1] Lee and Chang, J. predict stock values using Regression and LSTM based Machine learning concepts. It makes use of factors such as open, close, low, high and volume and accordingly addressed in the paper thereby predicting the future value of the financial stocks of a company using machine learning algorithms such as Regression and LSTM.[2] Parmar and Agarwal aim to implement the efficient deep learning-based prediction models, specifically gated recurrent unit (GRU) and long short-term memory (LSTM) to handle the price volatility of bitcoin and to obtain high accuracy. There are several virtual currencies such as bitcoin, ripple, ethereum, ethereum classic, lite coin, etc. The use of cryptocurrencies impacts on international relations and trade, due to its high price volatility. As bitcoin is the most popular cryptocurrency, the price volatility issue should be handled within a short period of time. Their study involved comparing these two time series deep learning techniques and proved the efficacy in forecasting the price of bitcoin.[3]

The work on predicting stock recovery using the Recurrent Neural Network and Long-Term Memory involved a program consisting of 5 categories namely: 1: Mature Data: historical stock data is collected at www.quandl.com/data/NSE and used in stock price forecasts. 2: Pre-Data Processing: Involves a) Data classification: Part of reducing data by importance, especially numerical data b) Data modification: Orientation. c) Data cleaning: Fill in the missing numbers. d) Data aggregation: Integration of data files. 3: Feature Discharge: Here the attributes to be assigned to the neural network are selected i.e.. Date, turn, up, down, close, and volume. 4: Nerve Network Training: Data is delivered to the neural network and trained to predict to give random bias and weights. Stage 5: Output Production: In this layer, the output value of the RNN output layer is compared to the target value.[4] Awoke introduces a novel and quick way to build stock market dictionaries, based on mathematical steps used in a large collection of labeled messages from StockTwits, a special stock market microblog. The authors compared three statistical variables, such as pointwise mutual information (PMI), two related statistics and the use of emotional points for positive and negative contexts. They also use a dictionary to easily generate emotional impressions on Twitter investors and analyze their relevance to research sentiments. The new microblogging indicators have shown a correlation with the popular indicators of Investors Intelligence (II) and the American Association of Individual Investors (AAII). Therefore, the new microblogging method is one that is better than standard test indicators with benefits such as creating cheap, high-frequency waves.[5]

Roondiwala, Patel & Varma predicted the stock returns using Recurrent Neural Network and Long Short-Term Memory . The system consists of 5 stages which are - 1: Raw Data: the historical stock data is collected from https:// www.quandl. com/data/NSE & used for the prediction of future stock prices. 2: Data Preprocessing: It involves a) Data discretization: Part of data reduction with importance, especially for numerical data b) Data transformation: Normalization. c) Data cleaning: Fill in missing values. d) Data integration: Integration of data files. 3: Feature Extraction: Here the features which are to be fed to the neural network are chosen ie. Date, open, high, low, close, and volume. 4: Training Neural Network: The data is fed to the neural network and trained for prediction assigning random biases and weights. Stage 5: Output Generation: In this layer, the output value generated by the output layer of the RNN is compared with the target value.[6] Moghar & Hamiche works has the daily opening prices of two New York Stock Exchange NYSE (GOOGL and NKE) stocks released on yahoo finance, in the GOOGL data series covering the period from 8/19/2004 to 12/19/2019 and NKE data include time from 1/4/2010 to 12/19/2019. To create a model using LSTM RNN, this model uses 80% of the training data and 20% of the data for testing. To train we use a square error to adjust the model. Different periods of training data (12 times, 25 times, 50 times and 100 times) are used for the model.[7]

In the eighth research paper research was also explored the commodity markets and futures. The number of hidden neural layers and the right time delay were calculated by LSTM using a genetic algorithm. It was found to offer the best performance compared to machine learning models. . It was found to offer the best performance compared to machine learning models. These are then transmitted to the deep neural network for speculation. Improved stock trading performance reported.[8]

In the ninth research paper twelve models were trained each time and in stock. The trained models are used for consideration in test times, which excludes the possibilities of the softmax layer. So, in each season, we have the opportunities we have in each class, depending on the 12 different models. Using these possibilities, produce three different types of combinations. The first two are a collection of equal weight and a combination of performance. To show that predictions come from the existence of a collection of models, we have also produced a collection in which, by predictable perceptions, they look only at the best model (according to the AUC of 390 previous observations). The authors then explored the hypothesis that all the information contained in lasso and ridge classifiers is also present in a work-heavy ensemble. Since both models are simple, we want to explore whether it is possible for a working-class collection to lose some basic information due to advanced complexity. He also concluded that no real value could be added to the lasso and ridge ensemble dividers, and thus all information was already included in the working weight ensemble.[9]

In the tenth research paper it proposes a LSTM unit with several novels to distinguish common and auxiliary objects. Specifically, we design input gates that are controlled by standard and past circuits hidden in standard and auxiliary features. By filtering data from both common and auxiliary features, these input gates generate memory cell inputs that will be integrated before updating cell conditions. With respect to the value gap between these input cells, it is necessary to assign different weights to different cell inputs and combine them into a single memory input for a measured amount. We use the attention method to give different weights. We can use the attention method to calculate the combustion weights of different cell inserts based on cell insertions and previous cell conditions. Attention weights are learned consistently by the training process. The main contributions to this paper are: Identifying potential use of stocks related to using the related stock price to predict the future price of a targeted stock. Also propose a MI-LSTM novel model that allows ordinary people to determine the use of certain features and apply a two-stage attention-grabbing approach to different input of memory cells and hidden instances of different time measures to improve predictive accuracy. Then compare our proposed model with various high-end models to evaluate the effectiveness of MI-LSTM in stock data from Chinese stock markets. MILSTM achieves 9.96% improvement over LSTM over medium error (MSE) error[10].

In the eleventh paper, they have developed an in-depth study model to predict the magnitude of the valence-arousal intensity of stock issues such as trend and trade dynamics. The HKAN model proposed in this paper incorporates Luong-based attention-based approach to optimizing HAN to extract the hidden element of the text and predict the valence arousal magnitude of stock issues as a message in the stock market. Test results show that our proposed HKAN model can effectively extract logical semantics (from word to phrase and keyword-based attention) between stock issues and both sides of the emotion. In addition, CNN can improve the predictive performance of stock valence-arousal intensity[11].

In the twelfth paper they have proposed the use of a stacked LSTM network model for predicting stock market behaviour, using data from NASDAQ Composite (ˆIXIC) . The model was trained and the results obtained show that the model was able to predict stock market behaviour with some accuracy. An open issue remains in that the volatility of the stock market cannot be mitigated using only historic data, but factors of the present also need to be analysed including current news in the world of politics and economics that could affect the behaviour of investors and ipso facto the behaviour of stock markets[12].

In the thirteenth paper they developed a new model using an in-depth study method to predict stock markets and apply in Bahrain all share index data (BAX). They explored the benefits of using an LSTM autoencoder and two technical indicators on multi-term audio output scales and data slides. The study investigated a BAX data set to predict many future horizons through the integration of in-depth learning methods based on LSTM design. The results show that the proposed model produced higher performance and produced better predictable values ​​compared to the simpler LSTM type and the shallow MLP network[13]. In the fourteenth paper they have introduced an in-depth study model that combines a convolutional layer with a repetitive layer to predict daily stock price movements and uses as an inclusion a combination of technical indicators and news headlines. RCNN properties can model local sequences of story headings and their interim features. The release of this model is used by a trading agent for the purpose of performing trading activities. It is important to note that in-depth learning models require a large amount of data, but when dealing with stock market prices to predict future events has nothing to do with future behaviour[14]. In this paper, the authors perform in-depth integrated reading and reinforced reading to represent stock signal and online trading. In this framework, an LSTM-based agent can automatically detect stock market volatility and reduce the difficulty of designing indicators from big data. In addition, they have added financial indicators to the LSTM network to reduce the impact of market noise. The reinforcement learning algorithm is used to train the agent to learn trading policy. Test results confirm the effectiveness of our approach. There are research guidelines to focus on in the future. Although they have added some technical indicators to reduce the impact of noise on the stock market, there is a lot of uncertainty in the stock data. How to reduce uncertainty remains a challenge[15].

In the paper by Aishwarya Singh, historical data about company stock prices listed publicly is used. He has used a combination of machine learning algorithms to predict the company's future price target, starting with simple algorithms such as average reversal and line, and moving on to advanced strategies like Auto ARIMA and LSTM. The main idea of ​​this article is to show how these algorithms are used.[16] In this paper, Ishita Parmar aims to predict the future value of the company's financial stock. The latest trend in stock market forecasting technology is the use of a study machine that makes predictions based on current stock market indicators by training in their previous values. Machine learning itself uses different models to make predictions easier and more realistic. The paper focuses on the use of retrofit and LSTM based machine learning to predict stock prices. Contemplated items are open, closed, low, high and volume. Predicting how the stock market will perform is one of the most difficult things to do. There are many factors involved in prediction - physical factors against mental, sensible and irrational behaviours, etc. All of these factors combine to make price fluctuations and make it extremely difficult to predict with high accuracy.[17]

In the paper by Sotirios, he contributes to the ongoing discussion about the nature and characteristics of the broadcast channels of crash events in the international stock markets. In particular, it investigates transfer mechanisms in all stock markets and outcomes from bond and financial markets. This approach incorporates a strong predictor of the stock market crash event. The advanced approach incorporates various learning algorithms that come with daily stock data, bond and currency from 39 countries covering a large economic spectrum. In particular, it utilizes the suitability of a range of strategies including Dividing Trees, Vector Support Machines, Random Forests, Neural Networks, Extreme Gradient Boosting, and Deep Neural Networks. With our best knowledge, it is the first time that comprehensive reading and development strategies have been considered in the literature as a way to predict episodes of the stock market crisis. The independent variables included in this data contain information related to the two basic funding links that can be implemented: compensation and flexibility. They use a series of machine learning algorithms to select alternatives related to the larger set of proposed ones. Finally, they use bootstrap samples to fix conflicts of available input data. Test results provide strong evidence that stock market problems often show persistence. Important evidence of trust and the effects of infection between the stock markets, bonds and funds are also available. Finally, it shows that the use of Deep Neural Networks significantly increases the accuracy of the sections, while providing a robust way to create a more efficient and sensitive global system warning tool than those currently established. Therefore, major banks can use these tools to quickly adjust their monetary policy, to ensure financial stability.[18]

Advanced neural networks (DNNs) have transformed the field of natural language processing (NLP). The Convolutional Neural Network (CNN) and the Recurrent Neural Network (RNN), the two main types of DNN structures, are extensively explored by Wenpeng Yin in [19] to manage various NLP activities. CNN should be good at outsourcing positions on different features and RNN on successive modeling units. The current state of many NLP operations is often changing due to the CNN and RNN war. This work is the first systematic comparison of CNN and RNN in the various functions representing NLP, which are intended to provide a basic guide to DNN selection. Will Koehrsen, through his work, tries to show the most effective way to learn the science of data in deep learning. This article is about how to build and use a repetitive neural network in Keras to write copyright summaries. The result is that you can build a useful system and discover how the in-depth learning process of natural language processing works. He uses a groundbreaking approach which is to learn how to use the technique before going back and integrating theory.[20]

Madhav Bhusal, with his paper, tries to use the Markov chain model to predict the behaviour of the Nepal Stock Exchange (NEPSE) index. Stock market behaviour forecasts are very important for investors looking for financial recognition. The use of the Markov series model to predict future trends is based on a powerful random feature of the NEPSE index. This paper aims to assess the long-term behaviour of the NEPSE index, the expected number of visits to a particular country and to determine the expected return time to the various provinces. In the study, the NEPSE index of 2741 trading days from 15 August 2007 to June 18, 2017 was taken as secondary data from the NEPSE office. The first regional vector and matrix for transformation opportunities, used to predict index behaviour, were obtained from a close examination of transformation numbers from one region to another.[21]

Stock market data is a good choice for LSTM because it is common and widely available to everyone. Derrick Mwiti demonstrates how to start a timeline for predicting LSTM models. He develops a comprehensive Python learning model that will predict future stock price trends. Although real estate forecasting is a steep climb, we can create a model that will predict whether the price will rise or fall.[22]

One of Nagesh Singh Chauhan articles as presented in this is associated with the stock market or stock market which focuses on buying and selling but fails to address the size and movements of the new investor. The general practice of looking at the stock market in public is that it is very risky for investment or not suitable for trading so most people are not interested. Seasonal variability and the continuous flow of any indicator will help both existing and inexperienced investors to understand and decide to invest in the stock / stock market. To solve these kinds of problems, time series analysis will be the best tool for predicting a trend or future. Trend charts can provide sufficient guidance for the investor. Explains the concepts expressed in machine learning and timetable chart[23].

Live streaming data services, such as in demand video, are becoming increasingly popular. In this paper, the importance of having live streaming service providers deviate from service requests due to problems or changing end user behaviour to ensure that end users receive the highest quality of service provided is stated. Therefore, in this Martin Boldt study, he investigates whether Markov-based models can be used for random detection using end-user control sequences in video streaming, i.e., event sequences such as playback, pause, start and stop. This confusing discovery method is also investigated by three different data solutions, in particular: 1 hour, 1 day and 3 days. The proposed anonymous detection method supports continuous live streaming times as it recalculates the probability of a particular session being different from each recently broadcast live broadcast control event. Two tests are used to measure the strength of the method, which gives promising results with accuracy[24] .

Saud A S et al demonstrates by his works that GRU and LSM models can surpass VRNN in stock price predictions. From the analysis made in the paper, it can be proved that GRU performed slightly better than LSTM (GRU gave a forecast of 4.74 and 4.71 while LSTM gave a forecast of 5.58 and 5.06) in two stocks. GRU has fewer parameters than LSTM so it is relatively quick to train and predict results. The reason behind the better performance of GRU over LSTM is the limited amount of data available for training model training. The approach taken introduced a novel review period and successfully identified its values ​​best suited for LSM and GRU. It is also found that the optimal turnaround time used with LSTM should be less than 5 and the optimal turnaround time used with GRU networks should be between 5 and 10. It is also noteworthy that, using large amounts of the retrospective period may provide poorly priced stock price performance.[25]

**2.2 Summary/Gaps identified in the Survey**

Summary:

A good amount of papers were found in the field of stock prediction and realized that there are plenty of ways to predict stocks. Through the papers we got an exposure to feature engineering and how different authors made use of it to benefit the accuracy of their model.

Gaps:

* The authors had collectively used various factors affecting the stocks, based on our opinion, it would be better suited if the inputs were categorized and used accordingly as they affect the stock trends differently.
* Certain models are stale and unreachable or unable to be interpreted by the public.
* Certains prediction models could have been tuned more to increase the accuracy.

1. **Overview of the Proposed System**

**3.1 Introduction and Related Concepts**

Stock market predictions have been an ongoing problem for many years and things have really gotten worse in this competitive world where only a limited number of people can own IPO shares.Whenever an investor decides to buy stock with the intention of investing. Your goal should be to acquire a good price, especially if you have been buying stocks for a long time. But before you invest in stocks you should do a thorough research, analyse the basics of stock and see if that stock fits into your portfolio before buying. You not only buy stock but also become a shareholder in that company, so as an investor you should be doing the right analysis.

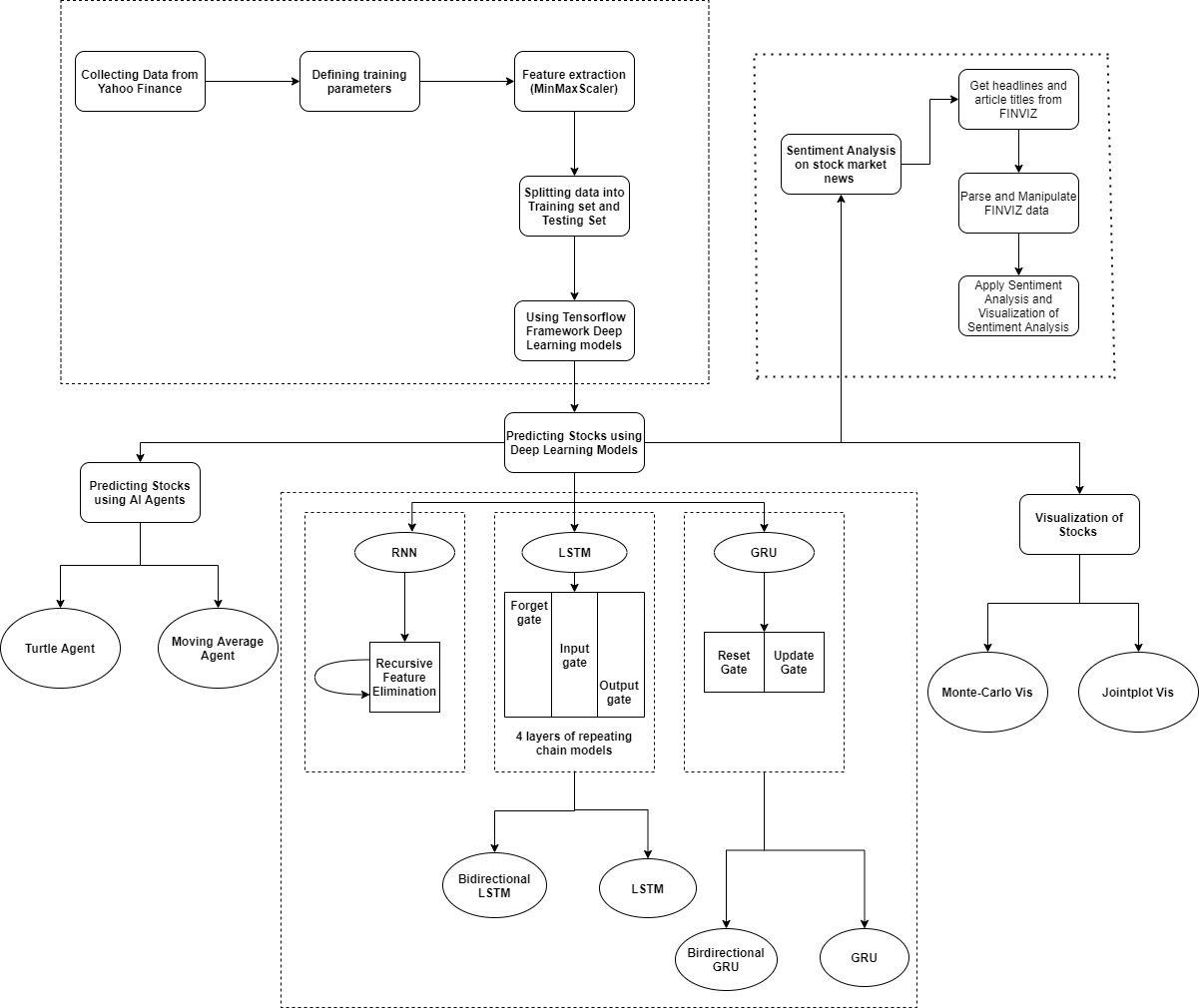
Here are some of the things one needs to learn before buying stocks and investing-

First, you need to determine the time period before buying the stock as it plays an important role in deciding whether to buy that stock or not. Your investment period can be short-term, medium-term or long-term, depending on your financial goals. It is important to learn a variety of investment strategies and choose one that fits your investment style Value Investing, Growth Investing and Income Investing. Investors should check the basics before buying a stock. Some of the most important criteria to consider before buying a stock are: Price-to-Earnings Ratio (P/E Ratio), Debt to Equity Ratio and Price-to-Book-Value Ratio (P/B Ratio). Investors should also consider how stocks compare with their peers, websites like StockEdge and Google Finance help companies compare with their peers. Investors should check the stock exchange before buying stocks and invest in those companies that have a high profile, High Domestic Domestic Institutional Investor holding and High Foreign Institutional Investor holding. When a stock is held by multiple joint ventures, it is generally considered to be a safe stock compared to other stocks that can be held by any joint venture. The size of the company you are considering investing in plays an important role in the amount of risk you want to take in order to buy stock. It is therefore important to consider the size of the company compared to your risk tolerance and horizontal before buying stock. Dividend shares are known for giving a portion of their profits to their investors in the form of dividends.

Investors who follow an investment strategy should try to invest in these shares. If an investor's goal is to make money by investing in it, they should look at the company's share capital before buying its stock. Leading investors looking for a higher level of revenue compared to the stock price should look at company shares. yield expressed as a percentage. Before buying stocks, investors should look at those growing companies. This can be determined by looking at both the income and the profit. Shares with high volatility levels will rise sharply on bullish days, and fall like bricks on bearish days. If you invest in a volatile stock that is slowing down and the latest uptrend is starting to decline, then you can go into your profits before they disappear. On the other hand, stocks that are moving faster do not give you more time to invest and if the trend slows down it could lead to a loss.

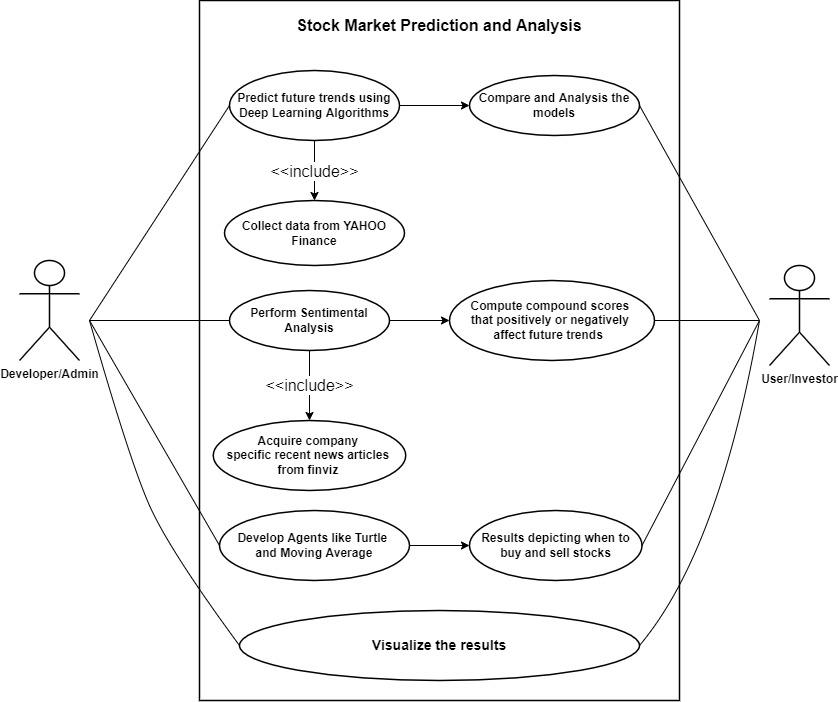
This project will use deep learning models to predict the future outcomes of the stock market, by analysing and finding the pattern in the data set provided. The model will be composed of four functions which will be evaluating several different tasks and will have separate complexity according to their use cases. The best speculation models are used by large hedge fund companies to understand the market and do their best in all trading strategies but they do not release their models so it is clear that no open source model can help traders market predictions. Therefore, we come up with a solution for ordinary traders who understand market movements and predict positive results for sure. In this case we will develop a systematic approach using existing solutions where people have used linear regression that works well for static data points but which is data failure that causes users to switch to a different algorithm.

**3.2 Framework, Architecture or Module for the Proposed System(with explanation)**



**3.3 Proposed System Model(ER Diagram/UML Diagram/Mathematical Modeling)**

UML Diagram:



**3.4 Methodology Adapted**

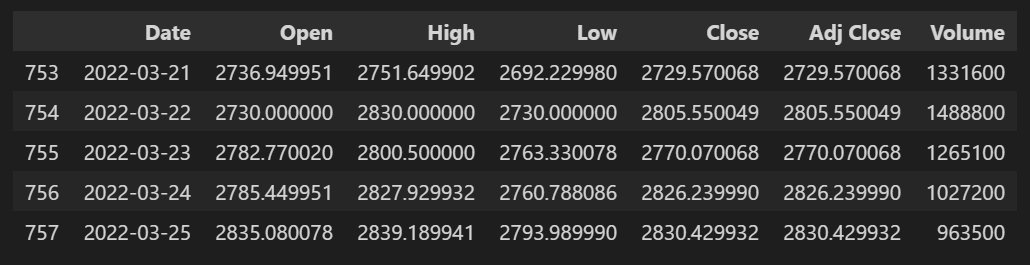
**3.4.1. Deep Learning Models**

**3.4.1.1 Dataset**

Google Historical Stock Prices in USD

Time Period: 2years (3years for LSTM)

Imported from yfinance library



**3.4.1..2. Model Creation**

The deep learning models were imported from *keras* and *keras.layers.*

Each Deep Learning Model included an Input Layer and an Output Layer. Additional Hidden layers were used to help model the complex data. We gave different Sequential models a different number of hidden layers ranging from 0 to 2 based on its learning and accurate predictions.

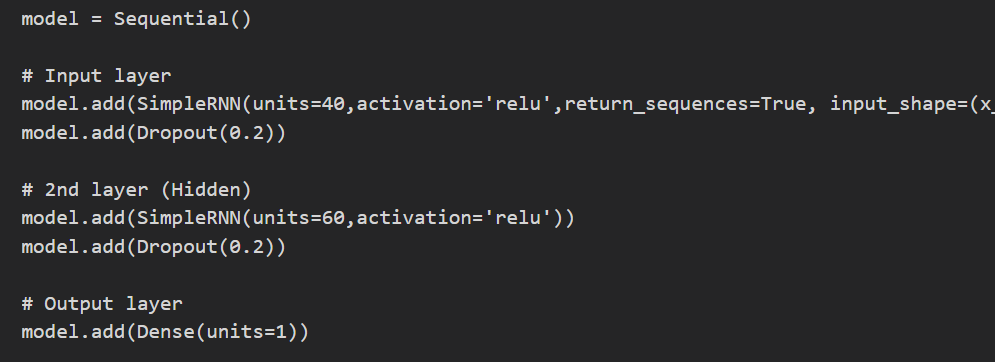
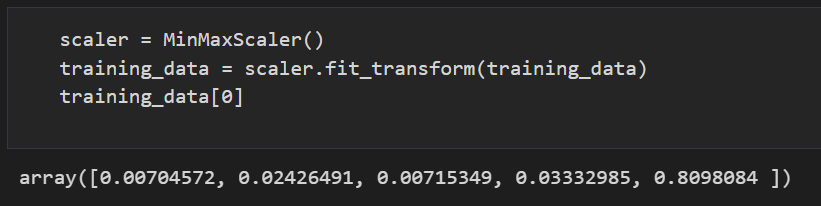


Fig. Sample model building of SimpleRNN with 3 layers

**3.4.1.3. Scaling the data**

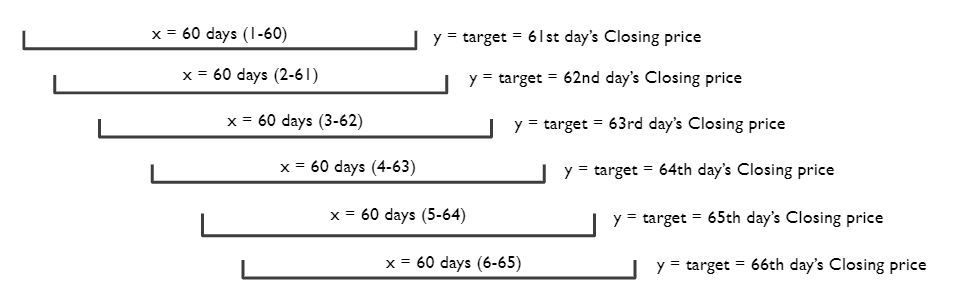
Library used: MinMaxScaler() from sklearn.preprocessing

MinMaxScaler normalizes the data in the range [0,1]

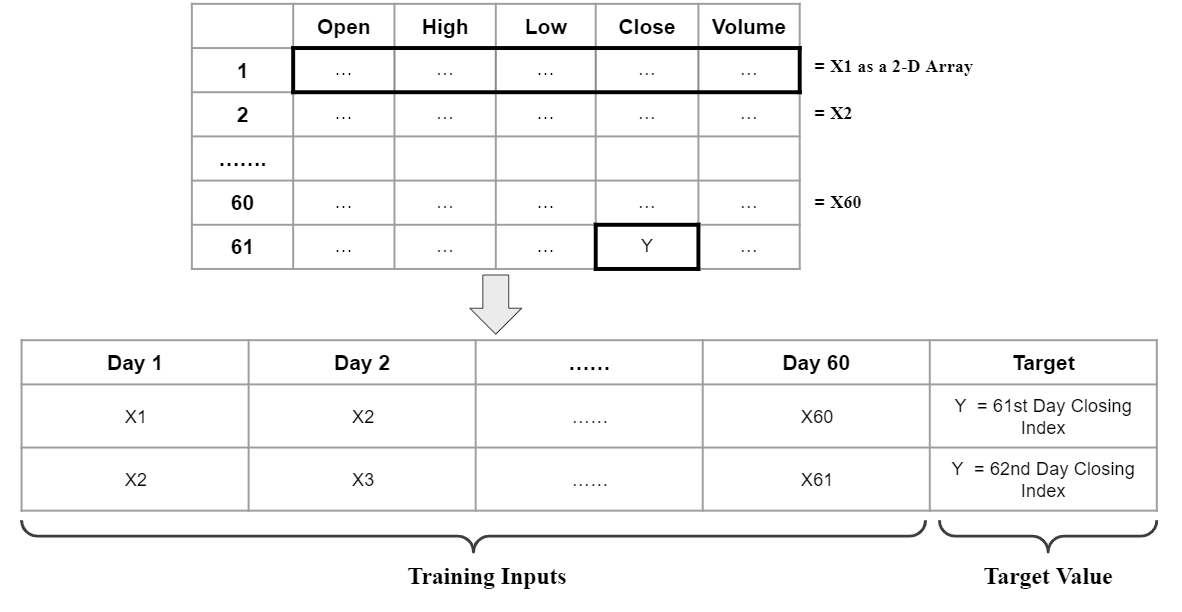


**3.4.1.4. Training Dataset**

Here, we are dividing and creating a dataset of 60 columns each. Basically our input values will include the Open, High, Low, Close and Volume prices from the dataset of 60 days in order to **predict the target closing index of the 61st day.**

****

The dimensions of our inputs will be **[60, 60, 5]** as shown below -

****

**3.4.1.5. Accuracy**

Two types of accuracy/error computational formulas were used. Both are solely based on the mathematical distance of the predicted value from its real value.

* **Mean Absolute Percentage Error (MAPE)**

accuracy = (1 - np.mean(np.abs((real-predict)/real))) \* 100

* **Root Mean Squared Error (RMSE)**

accuracy = (1 - np.sqrt(np.mean(np.square((real-predict)/real)))) \* 100

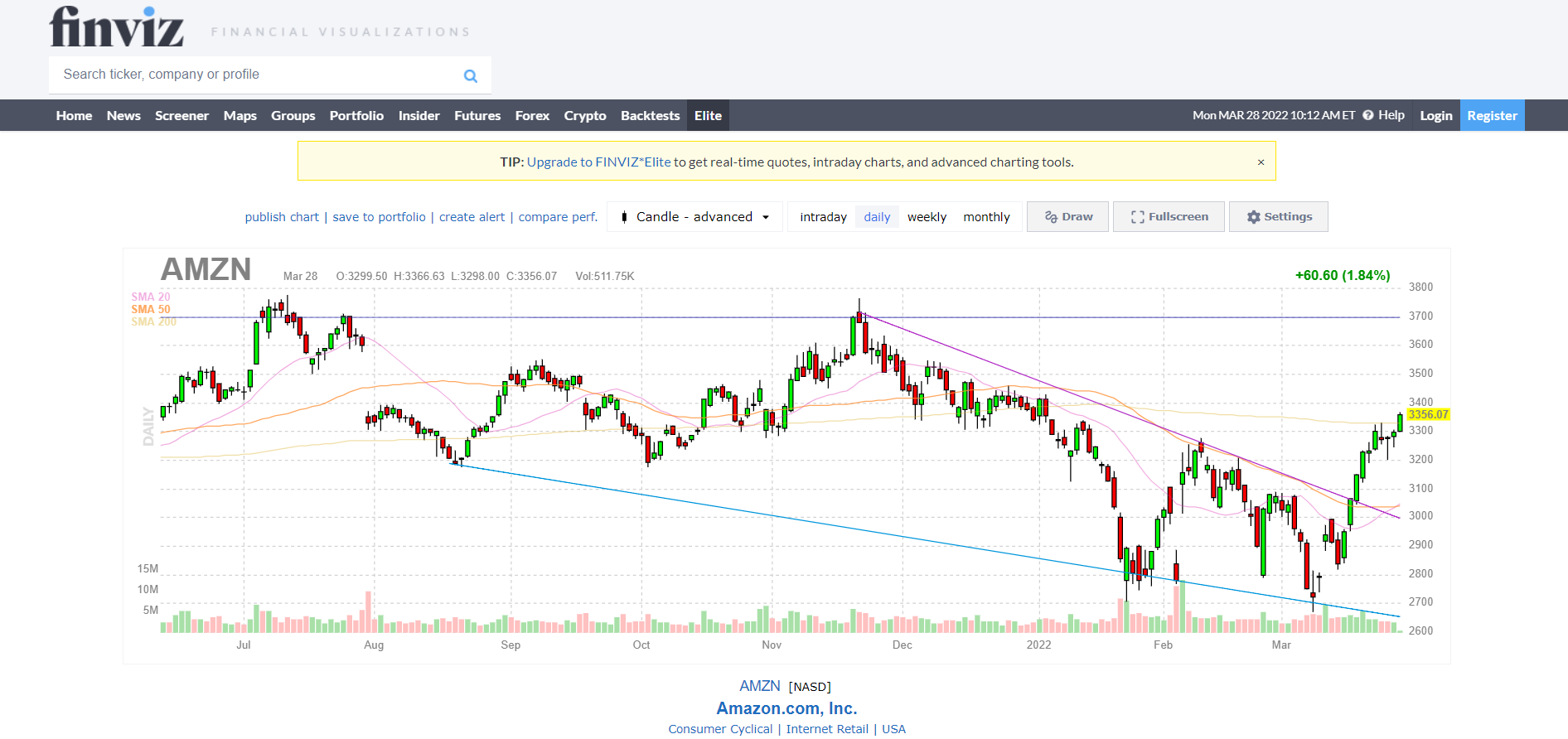
**3.4.1.6. Visualizing the Prediction Trend**

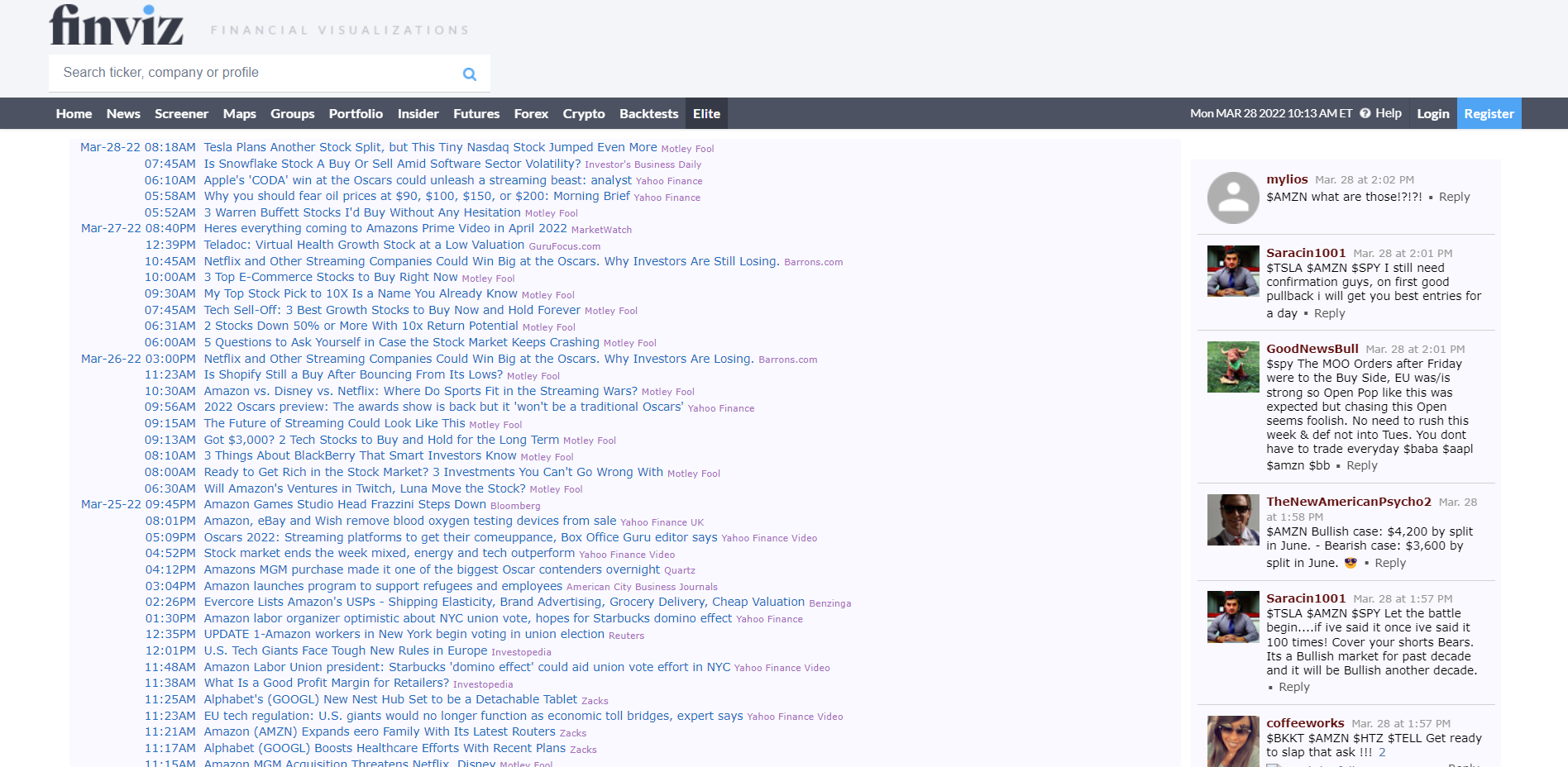
* Output: Closing prices of Google for the last 60 days
* Library used to plot the graph: matplotlib
* Forecasts: 3 simulations (Blue, Orange and Green)
* True Trend: Prediction line in black

**3.4.2. Sentiment Analysis-**

We are using Sentiment Analysis to understand financial news and make decisions on stock.

In this we used Python to parse through finviz.com, gather all the news article titles and then we do sentiment analysis to understand if everyday averaging news is positive , negative or neutral.





To perform Sentiment Analysis-

1. We'll use **Beautifulsoup in Python** to scrape article headlines from FinViz
2. Then, we'll use **Pandas** (Python Data Analysis Library) to analyze and run **sentiment analysis** on the article headlines
3. Finally, we'll use **Matplotlib for visualization** of our results

We will be doing this in 3 steps -

## Gathering and Parsing FinViz Data

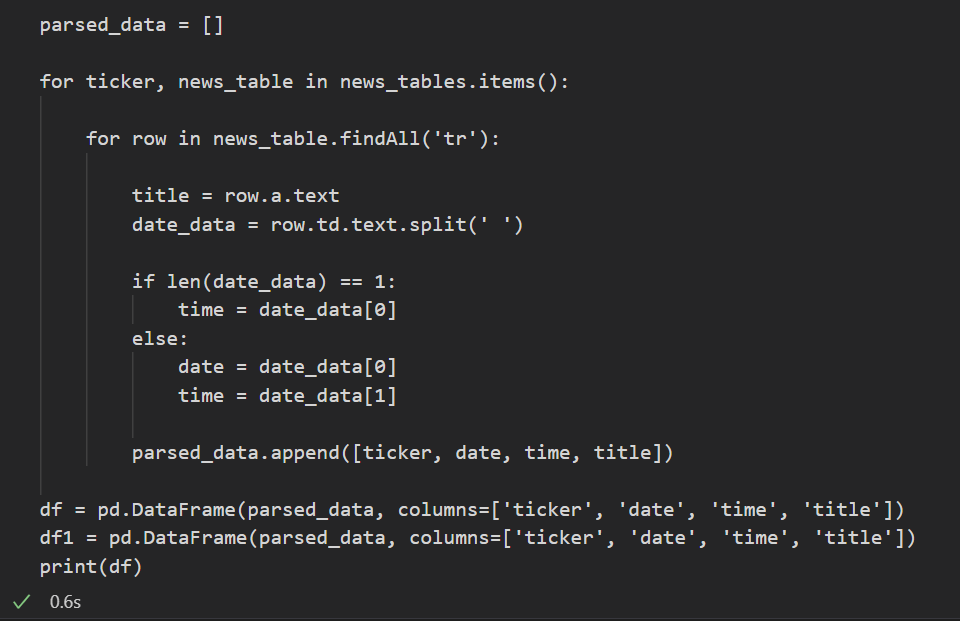
FinViz is a free website that makes stock data easily accessible to traders and investors. We'll gather the stock data from FinViz for a specific stock ticker.



Here we are getting the article title / headline from finviz then we are storing the article title along with the timestamp in a dictionary .

Then we parse through the table contents which are made from the dictionary and select article titles and we split the timestamp into date and time components.

At the end of this step we have scraped the components which are needed for Sentiment Analysis ie. tickers , date , time and article title.

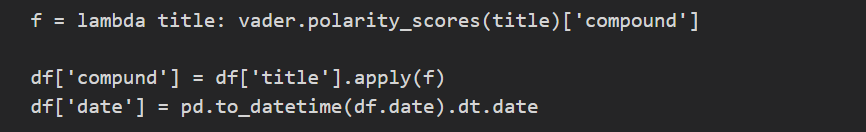


## Applying Sentiment Analysis

Applying sentiment analysis on the titles is done with NLTK (Natural Language Toolkit) **vader** that allows us to pass in a string into its function.

We initialize the SentimentIntensityAnalyzer, and then create a lambda function that takes in a title string, applies the vader.polarity\_scores() function on it to get the results in the above image and then only returns back the compound score. Using the apply function in Pandas, we can create a new 'compound' column in the data frame with all the compound scores from each title.





## Visualizing the Results in MatPlotLib

We visualize the data frame in MatPlotLib to see how our Stocks fared every day from public perception in news articles. For that we group our dataset based on the ticker and dates of each row, and then visualize the average compound score of each day.

1. **Proposed System Analysis and Design**

**4.1. Introduction**

This project uses deep learning models to predict the future outcomes of the stock market, by analyzing and finding the pattern in the data set provided. The model will be composed of functions which will be evaluating several different tasks and will have separate complexity according to their use cases. The model will be formed on RNN architecture using BasicRNN cells, LSTM cells and GRU cells which will be executed on another multi-layered RNN model.Thebi-directional version of LSTM and GRU have also been implemented. The outcome of these models will be used to compare which of the above models yields accurate results. The dataset values used for training were taken from 15th March 2021 to 15th March 2022.

Apart from deep learning modules we have implemented sentiment analysis for stock trading headlines and we will also be implementing visual representations of various simulations available such as Monte Carlo Simulations and Jointplot regression simulation which will help customers of the stock market make buying and selling calls on stocks more conveniently.

**4.2. Requirement Analysis**

**4.2.1. Functional Requirements**

**4.2.1.1 Product Perspective**

This project is based on prediction of the Stock Market using Deep learning models. Along with the various deep learning models we have also used 2 stock market trading strategies like turtle trading and moving average strategy and based on these strategies we have developed a model which will help us to predict the stock trend. We have also incorporated sentiment analysis for stock trading based on news and headlines which are available on finviz.com which is an online financial company. In the we have included 2 data visualization techniques which will help any user be it an expert or new comer in this field of stock trading to invest by seeing graphs and diagrams. Thus we will be using Monte Carlo and Joint plot regression simulation as these are simple and easy to understand.

**4.2.1.2 Product features**

The main idea of this project was to make a prediction model which could help people trade in the stock market despite being proficient in the stock market. For this we have used 3 different techniques to predict the stock trend - Deep Learning Models, Stock Trading Strategies as agents and Data visualization models. Each technique is unique in a certain way and along with this we have incorporated sentiment analysis for sudden outbreaks in the stock market due to unpredictable situations like world crisis or other factors (chances of such factors are pretty less but we are including such outliers as well).

We will be comparing the outputs and accuracy of 3 most used deep learning models Basic RNN, LSTM , GRU and their Bi directional versions ie. Bidirectional LSTM and GRU. This will predict the stock trend and give traders an idea which stock one should sell/buy and the accuracy will tell us which model for the particular data is best suited. The agents which we are using are based on most popular stock trading strategies, these give a detailed trading analysis which will give a full schedule when to buy or sell how many units of stock and the output will be a visualization graph which makes it easy to understand the prediction done. Monte Carlo Simulation and Joint plot simulation help customers of the stock market make buying and selling calls on stocks more conveniently. Sentiment analysis used here will understand financial news and make decisions on stock based on current headlines and news.

**4.2.1.3. User characteristics**

Users must be legally adults and should satisfy the requirements mentioned below. Users should carefully study the graph outcomes and visualization outputs and accordingly invest in the specified company stocks.

**4.2.1.4. Assumption & Dependencies**

**4.2.1.5 Domain Requirements**

\*\*\*\*\*

**4.2.1.6. User Requirements**

To invest in stocks publicly listed on the market, the user needs to fulfill the following requirements:

Personal documents -

* PAN Card
* Aadhaar Card
* Name on a canceled cheque from your active bank account
* Proof of residence based on a list of documents that have been accepted by your stock broker, depository participant, or bank
* Account statements
* Passport-size photographs
* Demat Account
* Trading Account
* Linked Bank Account.

**4.2.2. Non Functional Requirements**

**4.2.2.1. Product Requirements**

**4.2.2.1.1. Efficiency (in terms of Time and Space)**

The efficiency is dependent on the dataset. Bigger the dataset, the time needed, more factors or values (closing value, opening values, highest lowest etc) more the time and similarly the space also depends on the dataset and factors. The deep learning models chosen work in different situations like LSTM work well for datasets with gaps or missing data, GRU works better in case of smaller datasets. RNN needs a complete dataset.

**4.2.2.1.2. Reliability**

The reliability of the project outcome will be dependent on the accuracy of the dataset. And each deep learning model used for prediction of stock trend has its respective accuracy which is displayed in the outcome along with the stock trend.

**4.2.2.1.3. Portability**

The project is completely portable and the recommendations and predictions are completely trustworthy as the data is dynamically updated.

**4.2.2.1.4. Usability**

The project is extremely user friendly as the outputs produced by the models, agents, visualizations and sentiment analysis part are easy to interpret and understand and even a new user or expert can fully trust on these predictions and accordingly invest.

**4.2.2.2. Organizational Requirements**

**4.2.2.2.1. Implementation Requirements**

The only requirements a system needs is a python/ jupyter notebook and some libraries like nltk, matplotlib, pandas, seaborn etc. This project can also be implemented on google collab. We just need updated dataset latest data so get accurate results.

**4.2.2.2.2. Engineering Standard Requirements**

\*\*\*\*

**4.2.2.3. Operational Requirements (Explain the applicability for your work w.r.to the following operational requirement(s))**

* **Economic:**

The analysis done in this project can help interpret the future movements of the stock value in terms of financial exchange. An accurate prediction of the share prices could further lead to more profits investors can make. By giving people a chance to invest and become part-owners of an enterprise, it could help large income inequalities.

* **Environmental:**

Certain prediction trends could help reduce environmental degradation by enforcing strong regulations and actions onto listed companies/enterprises, so as to use greener technologies. Green investing will seek out investment opportunities that will benefit the natural environment.

* **Social:**

Stocks affect social causes in three critical ways: They allow small investors to invest in the economy. They help savers beat inflation. They help businesses fund growth.

* **Political:**

Political events usually have a great impact on the stock market. In many cases, the stock market fluctuates because of political announcements such as regulation promulgation, law amendments and national elections. Keeping this in mind, the sentimental analysis on news articles helps compound their effects on stock trades.

* **Ethical:**

If stock returns are essentially random, the best prediction for tomorrow's market price is simply today's price, plus a very small increase. Our model is based on this principle. For this reason we have used short term time series algorithms such as LSTM and GRU that simply takes inputs from the previous day's output.

* **Health and Safety:**

After the outbreak of the COVID‐19, the stock market came under fear as the BSE Sensex and NSE Nifty fell by 38%. Such critical incidents affecting the population as a whole can drastically affect the stocks as people would not want to take risks with even the slightest inaccuracy in the prediction trends.

* **Sustainability:**

The pressure on organizations to meet environmental, social and governance (ESG) criteria is more widespread than most finance leaders might realize, 85% of investors considered ESG factors in their investments in 2020. Sustainable investing enables individuals to select investments based on values and personal priorities. Thereby leading them towards a model that guides them when and how much to invest.

* **Legality**:

Indian Capital Markets are regulated and monitored by the Ministry of Finance, The Securities and Exchange Board of India and The Reserve Bank of India. providing an efficient legislative framework for securities markets. This ensures investors safe trading in terms of legality and therefore encourages them to invest in the same.

* **Inspectability:**

\*\*\*\*

**4.2.3 System Requirements**

**4.2.3.1. H/W Requirements(details about Application Specific Hardware)**

* Working Desktop
* Wifi Routers/Net connectivity for API calls

**4.2.3.2. S/W Requirements(details about Application Specific Software)**

* Python/Jupyter Notebook
* Packages such as nltk, tensorflow, pandas, seaborn, matplotlib
* Valid Datasets from Yahoo Finance which includes the date and required trading factors needed for the Deep Learning models.
* Valid Url requests made to finviz.com to acquire company specific news articles used for Sentimental Analysis.

1. **Results and Discussion**

* **Predictions and Analysis**
* **Simple RNN:**

Here, the model used 3 layers, with a batch size of 64 at 300 epochs. In terms of error functions, MAPE gave a mean accuracy of 91.91% and RMSE gave an accuracy of 86.01%

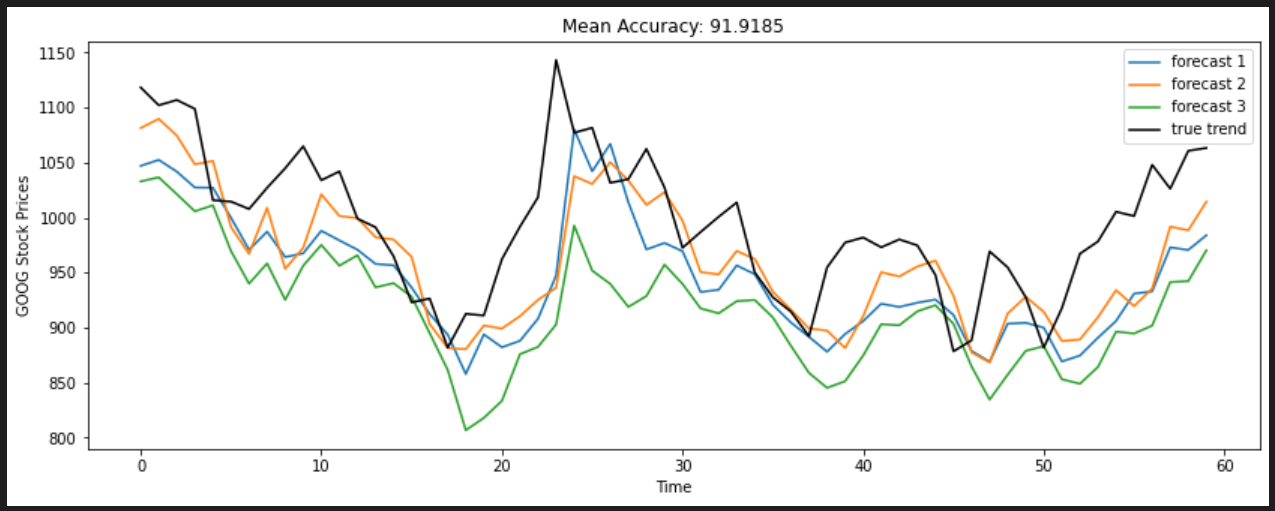


Fig 5.1. Simple RNN Model Predictions

* **LSTM:**

Here, the model used 4 layers, with a batch size of 32 at 300 epochs. In terms of error functions, MAPE gave a mean accuracy of 92.90% and RMSE gave an accuracy of 89.73%.

For LSTM, a dataset with a time period of 3 years was used, as it works well with a larger dataset helping the model learn better.

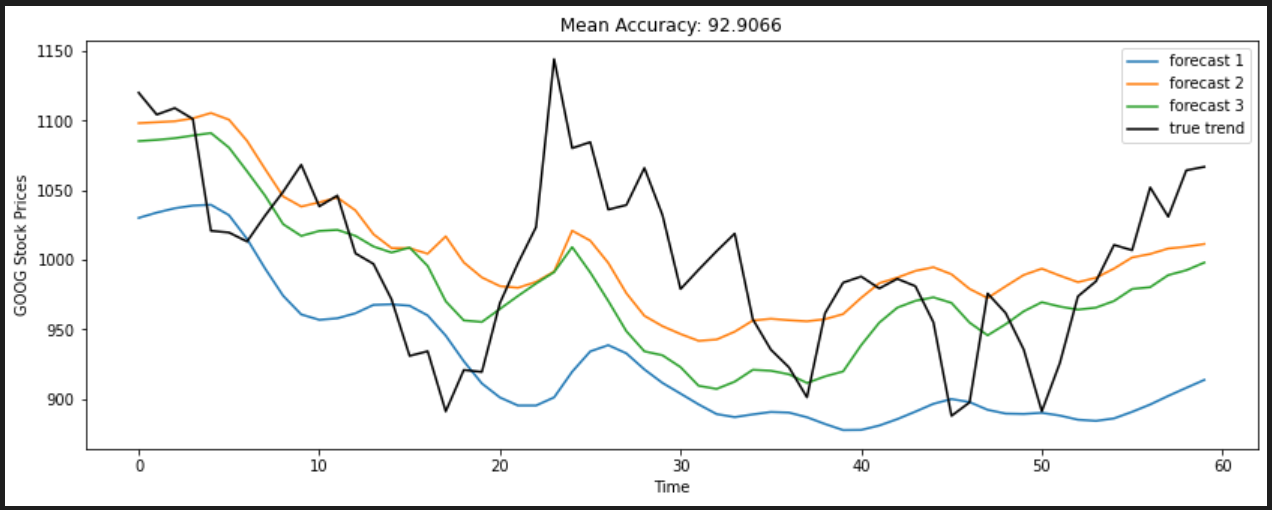


Fig 5.2. LSTM Model Predictions

* **GRU:**

Here, the model used 2 layers, with a batch size of 32 at 300 epochs. In terms of error functions, MAPE gave a mean accuracy of 93.09% and RMSE gave an accuracy of 91.39%

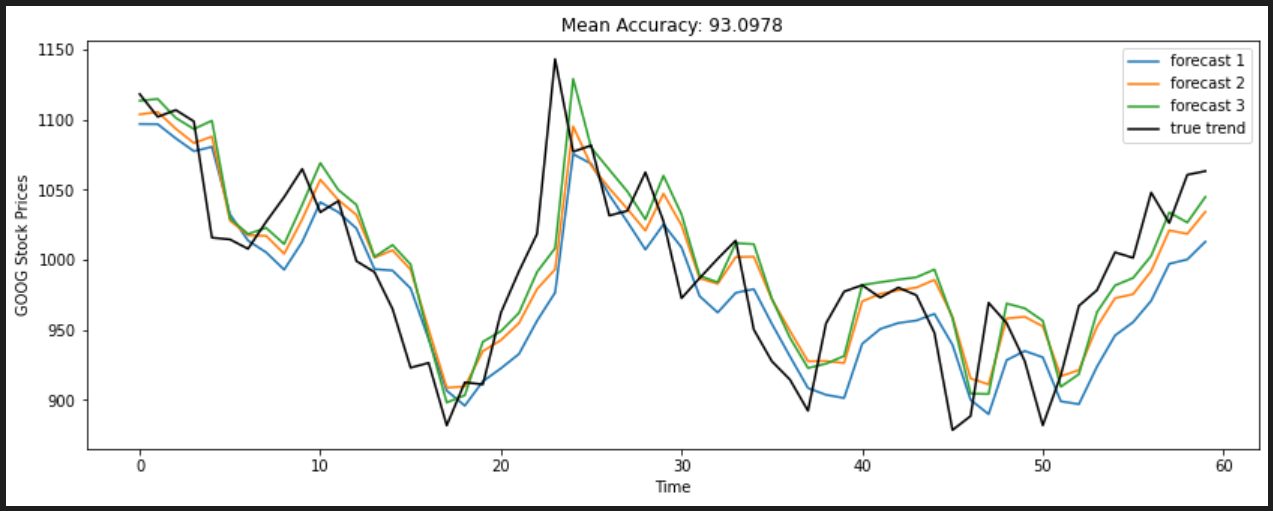


Fig 5.3. GRU Model Predictions

* **Bidirectional LSTM:**

Here, the model used 4 layers, with a batch size of 32 at 300 epochs. In terms of error functions, MAPE gave a mean accuracy of 94.53% and RMSE gave an accuracy of 94.06%.

Here, we used a timestamp of 30 days, as it gave a higher accurate prediction trend.

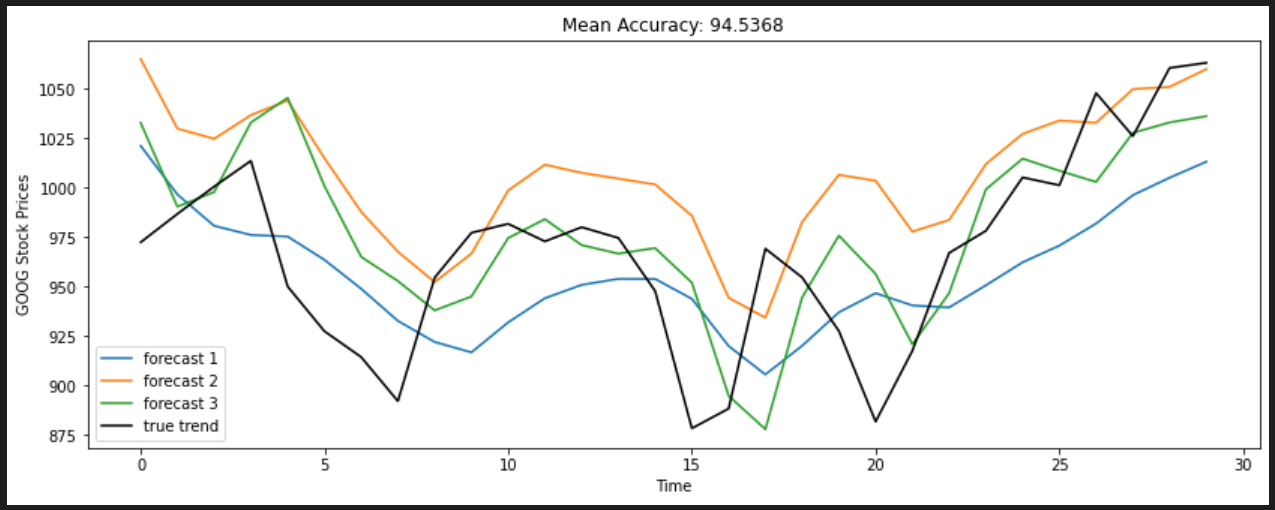


Fig 5.4. Bidirectional-LSTM Model Predictions

* **Bidirectional GRU:**

Here, the model used 3 layers, with a batch size of 32 at 300 epochs. In terms of error functions, MAPE gave a mean accuracy of 92.98% and RMSE gave an accuracy of 90.95%

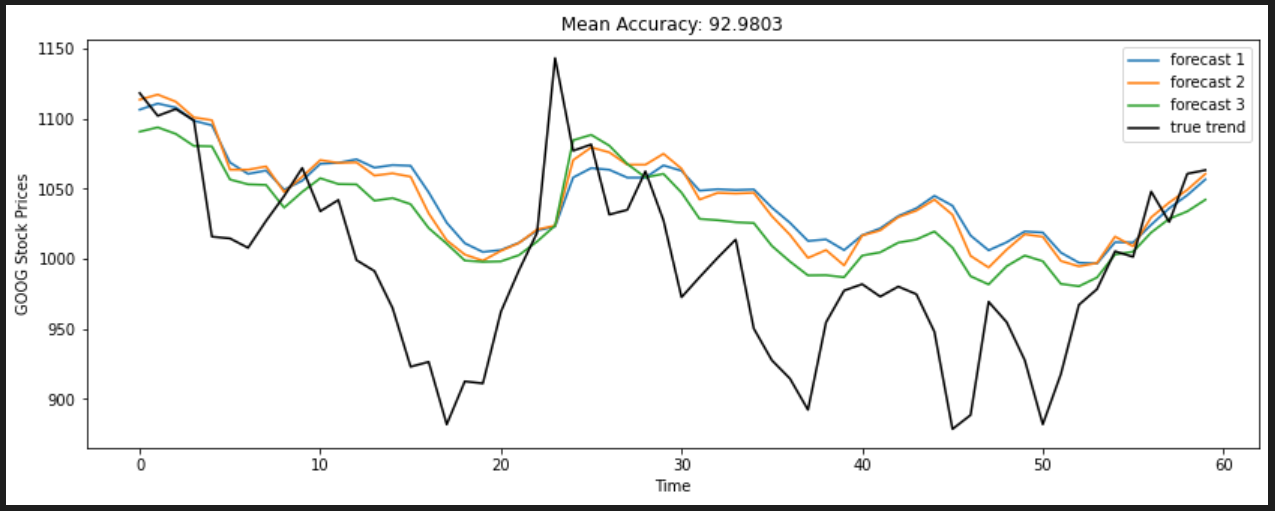
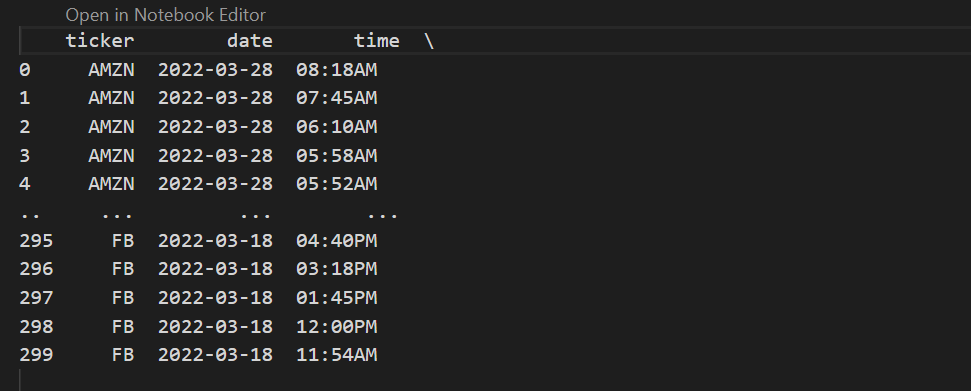


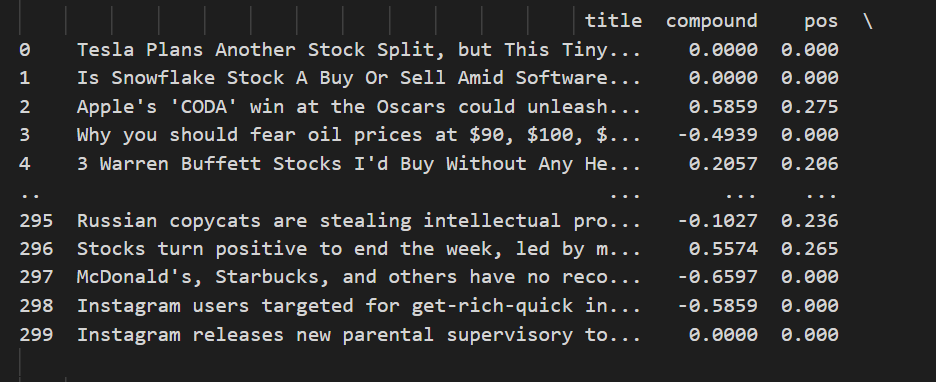
Fig 5.5. Bidirectional-GRU Model Predictions

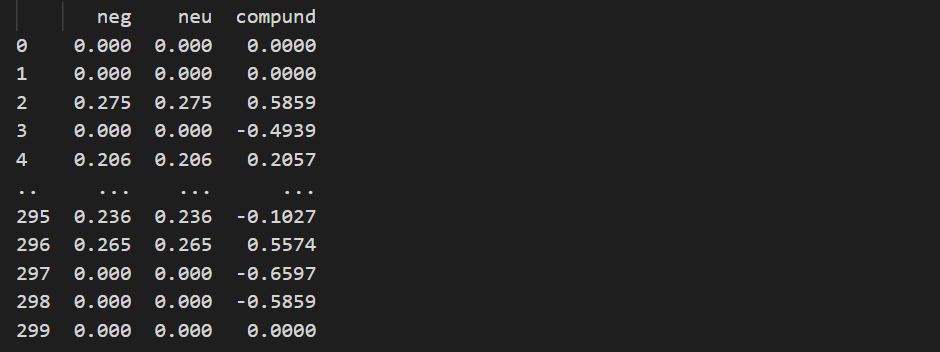
**5.2 Sentimental Analysis**

For Sentiment Analysis we have used python for scraping the finviz.com to gather the article titles then we used pandas to analyze and run nltk module for sentiment analysis and finally in the end we use matplotlib for visualization. In the output we get graphs where we have specified 3 companies - Amazon, Google and Facebook. These three companies are our tickers, we have gathered news articles of these companies and performed sentiment analysis on them. We later on visualized the output and we can see the correlation and average daily score (compound score) of how that stock of a company has performed based on the news.

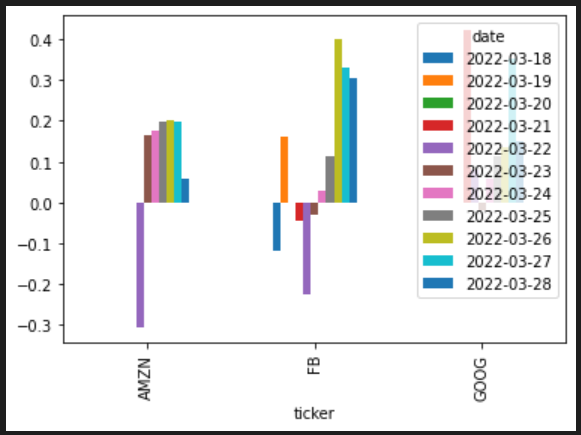
We get the data frames as such which will have tickers, date, time, article titles related to stocks, compound scores, positive, negative and neutral scores. -



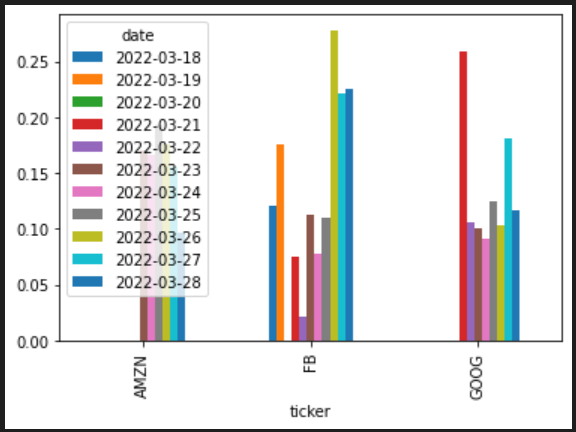




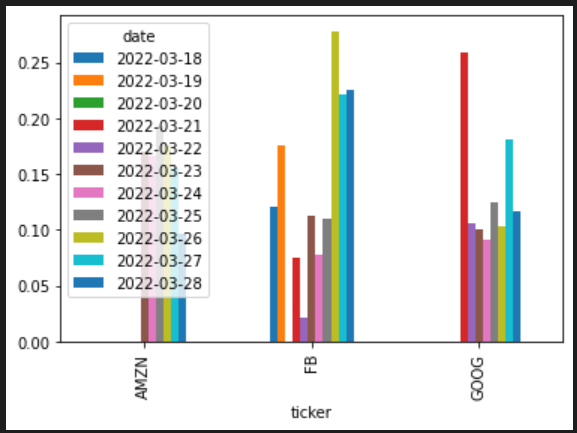
In the first graph we calculated the average compound scores of all articles to see whether its positive, negative or neutral.We have visualized it in bar chart format -



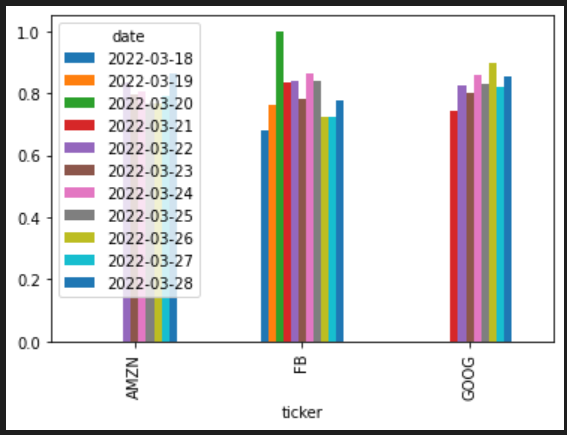
Here we have considered the positive score -



Here we took the negative score -



In this graph we took the neutral score -



These graphs give us an idea about whether the daily news of a particular company is positive or negative and whether one should invest in it or not. Each graph has bars of different days and we have done that for three companies.

This bar chart tells us what the average score of each company is for the past 9 days.This will make it easier to compare between different companies. Here we took the cross section of the compound score row, flipped the data frame so that we have the dates as the x-axis, and then plotted it as a Bar chart -



1. **References**

[1] Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019, September). Stock price prediction using LSTM on the Indian share market. In Proceedings of 32nd international conference on (Vol. 63, pp. 101-110).

[2] Lee, M. C., Chang, J. W., Hung, J. C., & Chen, B. L. (2021). Exploring the effectiveness of deep neural networks with technical analysis applied to stock market prediction. ComputerScience and Information Systems, (00), 2-2.

[3] Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018, December). Stock market prediction using Machine Learning. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)

(pp. 574-576). IEEE.

[5] Awoke, T., Rout, M., Mohanty, L., & Satapathy, S. C. (2021). Bitcoin price prediction and analysis using deep learning models. In Communication Software and Networks (pp. 631-640). Springer, Singapore.

[6] Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting stock prices using LSTM. International Journal of Science and Research (IJSR), 6(4), 1754-1756.

[7] Moghar, A., & Hamiche, M. (2020). Stock market prediction using lstm recurrent neural network. Procedia Computer Science, 170, 1168-1173.

[8] Yadav, A., Jha, C. K., & Sharan, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. Procedia Computer Science, 167, 2091-2100.

[9] Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high‐ frequency stock market classification. Journal of Forecasting, 38(6), 600-619.

[10] Li, H., Shen, Y., & Zhu, Y. (2018, November). Stock price prediction using attention-based multiinput LSTM. In Asian Conference on Machine Learning (pp. 454-469). PMLR.

[11] Wu, J. L., Yang, C. S., Liu, K. H., & Huang, M. T. (2019, October). A Deep Learning Model for Dimensional ValenceArousal Intensity Prediction in Stock Market. In 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), IEEE.

[12] Al-Thelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2019, April). Forecasting of bahrain stock market with deep learning: Methodology and case study. In 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO) (pp. 1-5). IEEE.

[13] Vargas, M. R., dos Anjos, C. E., Bichara, G. L., & Evsukoff, A. G. (2018, July). Deep leaming for stock market prediction using technical indicators and financial news articles. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[14] Jia, W. U., Chen, W. A. N. G., Xiong, L., & Hongyong, S. U. N. (2019, July). Quantitative trading on stock market based on deep reinforcement learning. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[15] Ojo, S. O., Owolawi, P. A., Mphahlele, M., & Adisa, J. A. (2019, November). Stock market behaviour prediction using stacked LSTM networks. In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), IEEE.

[16] Aishwarya Singh(2018), Stock Prices Prediction Using Machine Learning and Deep Learning Techniques, Analytics Vidhya.

[17] Chouhan, Lokesh & Agarwal, Navanshu & Parmar, Ishita & Saxena, Sheirsh & Arora, Ridam & Gupta, Shikhin & Dhiman, Himanshu. (2018). Stock Market Prediction Using Machine Learning. 10.1109/ICSCCC.2018.8703332.

[18] Sotirios P. Chatzis, Vassilis Siakoulis, Anastasios Petropoulos, Evangelos Stavroulakis, Nikos Vlachogiannakis, Forecasting stock market crisis events using deep and statistical machine learning techniques, Expert Systems with Applications, Volume 112, 2018, Pages 353-371, ISSN 0957-4174.

[19] Wenpeng Yin, Katharina Kann, Mo Yu and Hinrich Schutze(2017), Comparative Study of CNN and RNN for Natural Language Processing, Arxiv:1702.01923v1.

[20] Will Koehrsen, Recurrent Neural Networks by Example in Python, 2018, towardsdatascience.com.

[21] Madhav Kumar Bhusal, Application of Markov Chain Model in the Stock Market Trend Analysis of Nepal, 2017, International Journal of Scientific & Engineering Research, Volume 8, ISSN 2229-5518.

[22] Derrick Mwiti(2018), Using Keras Long Short-Term Memory (LSTM) Model to predict Stock Prices, KDnuggets, 18:n45.

[23] Nagesh Singh Chauhan(2020), Stock Market Forecasting Using Time Series Analysis, 20:n20.

[24] Boldt, M., Borg, A., Ickin, S. et al. Anomaly detection of event sequences using multiple temporal resolutions and Markov chains. Knowl Inf Syst 62, 669–686 (2020).

[25] Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. Procedia Computer Science, 167, 788-798.