**A PROJECT REPORT**

**On**

**“STOCK PREDICTION USING ML TECHNIQUES”**

**Submitted to**

**KIIT Deemed to be University**

**In Partial Fulfillment of the Requirement for the Award of**

**BACHELOR’S DEGREE IN COMPUTER SCIENCE & ENGINEERING**

**BY**

**NAMAN GERA 1729036**

**PARITOSH SINGH 1729044**

**PRABHAV SINGH 1729045**

**SHIVANGI MISHRA 1729060**

**SHUBHAM SINGH 1729073**

**UNDER THE GUIDANCE OF**

**PROF: DR.DEBAJYOTY BANIK**

SCHOOL OF COMPUTER ENGINEERING

# **KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

BHUBANESWAR, ODISHA – 751024 MAY 2020

**KIIT Deemed to be University**

## School of Computer Engineering Bhubaneswar, ODISHA 751024

**CERTIFICATE**

This is to certify that the project entitled

“**STOCK PREDICTION USING ML TECHNIQUES** ”

Submitted by:

**NAMAN GERA 1729036**

**PARITOSH SINGH 1729044**

**PRABHAV SINGH 1729045**

**SHIVANGI MISHRA 1729060**

**SHUBHAM SINGH 1729073**

is a record of Bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Computer Science and Communication Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during the year 2019-2020, under our guidance.

Date: / /

**DR.DEBAJYOTY BANIK**

Project Guide

**Acknowledgements**

We would like to express our deep sense of gratitude to our project guide Prof. DEBAJYOTY BANIK for the useful comments, remarks and invaluable support throughout the completion of this project. Furthermore we would like to thank the department of Computer Science for this great learning opportunity.We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in the journey of this project.

**NAMAN GERA**

**PARITOSH SINGH**

**PRABHAV PANDEY**

**SHIVANGI MISHRA**

**SHUBHAM SINGH**

**ABSTRACT**

*Predicting stock prices is one of the best ways to take a step closer to predict one’s future. The shareholder’s gain depends on the successful forecast of the stock’s future. The proposed model compares different ML algorithms like LR, RNN, GRU, LSTM, Random Forest, etc. To figure out the best-suited algorithm, i.e., algorithms which return the fewest errors based on an imported benchmark data set. We compare various algorithms upon the graph of their predicted data, after training the data. In the second part, a twitter semantic analyzer is created using a twitter API on a developer ID. This twitter API is used to access consumer Keys, consumer secrets, access tokens, and access token secrets. A class Tweepy is used to import tweets from twitter. Class TextBlock is used to analyze the tweets and return avg. semantic value (polarity) of all tweets. In the third part, after predicting three values by using LSTM, we used an algorithm based on semantic analysis, that will use the predicted values and will help decide based on current tweet trends for the corporation whether investing will be Loss or Profit. Our principal goal is to analyze distinct types of research done already in this field and existing stock prediction models to predict the stock prices with the least error. Then implement the semantic analyzer to predict the nature of the*

*tweets and according to the response. Build a scale for Investment suggestions. [1][2][10]*

*Index terms- Linear Regression(LR), Recurrent Neural Network(RNN),Gated Recurrent Unit(GRU), Long Short-Term Memory(LSTM).*

Contents

**[1.INTRODUCTION](#_5d2nb27807zx) 6**

[1.1 Stock market](#_w40twlth5f03) 6

[1.2 Neural Networks](#_ejshpa22ie3i) 6

[1.3 Semantic Analysis](#_r51xwve5pccp) 6

**[2 .LITERATURE SURVEY](#_xs22p2297ohq) 7**

**[3.Software Requirements](#_z43fa5339s60) 8**

**[4.Requirement Analysis](#_okzokungtvl) 8**

[4.1 Related Work](#_imir83bd5p6p) 8

**[5.System Architecture and Design](#_ld0rvkj4lyhj) 11**

[5.1 Recurrent Neural Network](#_55qtwtnuinhu) 11

[5.2 GRU](#_57qyzvi2xj4j) 11

[5.3 LSTM Networks](#_y4m46fpxjrv) 12

[5.4 Twitter Sentiment Analysis](#_hk5e2ex0ltzl) 12

**[6.Project Planning](#_noxr27w89iyh) 13**

[6.1PROBLEM STATEMENT:](#_h7y04dt8o0kp) 13

[6.2METHODOLOGY USED TO ANALYZE THE ABOVE PROBLEM:](#_p10ngcy8mvcl) 13

**[7.Implementation](#_gx596c6czgo8) 13**

[7.1 Part 1](#_cozo1uq2tvw2) 13

[7.1.1 Results :](#_jbc1brq03m65) 14

[7.2 For Part 2:](#_8ljui7f99bdp) 14

[7.2.1Access Live Data and Processing](#_as7m089ltsjl) 14

[7.3 For part 3](#_jrp5c22u4fq4) 15

**[8.ScreenShots of project:](#_au9sm1tne9a6) 17**

**[10.Refrences](#_47cgve8tgy8s) 19**

# 

# **1.INTRODUCTION**

### 1.1 Stock market

An effective accumulation of capital can improve the organic composition of enterprise capital and promote the development of the economy. [1] The stock market plays a significant role in the GDP of a country. Investing in the stock market can be risky. Predicting and mitigating the imperiling of the stock trend is highly essential if a venture capitalist wishes to invest in the market. This seems practically futile because of the distinct factors influencing stocks like inflation rates, economic environment, political issues, and so on. It is profoundly noteworthy to devise some mechanisms that safely predict stock prices. Scholars have suggested many prediction models that have been quite effective and challenging. Neural Network has been the popular means for stock prediction. However, there are many ways such as Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU)[3][4][5].Sentimental Analysis refers to the use of Artificial Intelligence to interpret, classify, and study a range of emotions within a text using text analysis. Sentiment Analysis came into action as the demands of people using the internet changed overtime. As the opinion of people on the internet became important. Data analysts saw this as an opportunity to gather opinions, search through them, and extract the important information and finally act on them by taking a range of decisions. We have taken twitter reviews on Live Dataset. We have trained the training data using a Naïve based classifier.[1][2][7][8]

### 1.2 Neural Networks

Neural Network has been the popular means for stock prediction. However, there are many ways such as Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

The idea of RNN is straightforward, which is to help us process information in stages. When referring to a standard Neural Network all inputs and outputs are not dependent on each other. Only if the researcher uses this concept, what happens is that the tasks on the neural network are very numerous and stacked up. Whereas if researchers want to build a forecasting system, we need a prediction from the best input so that the model can learn well from historical

data.

### 1.3 Semantic Analysis

Sentimental Analysis refers to the use of Artificial Intelligence to interpret, classify and study a range emotions within a text using text analysis.Sentiment Analysis came into play as the needs of people using internet changed overtime.As the opinion of people on the internet became important.Data analyst saw this as opportunity to collect the opinions search through them and extract the important information and finally act on them by taking a range of decisions.We have taken twitter reviews on Live Dataset.We have trained the training data using a Naïve based classifier.

# 

# **2 .LITERATURE SURVEY**

### 

In this section we will discuss various algorithms and prediction models established in prior research.

1. Pratap and Ambika [2011] proposed trigonometric functional link artificial neural network (FLANN) model employs standard least mean square (LMS) algorithm with search-then-converge scheduling. The network could effectively calculate a learning rate parameter that changes with time and may require less experiments to train the model. Here FLANN is used for long term as well as short term stock market prediction .
2. In 2010, Liu wrote another book chapter about „Sentiment Analysis and Subjectivity‟ [Liu 2010]. Although the new book chapter covers some recent articles, the content in general is very similar to the previous book chapter. The focus of the new book chapter is purely sentiment classification techniques, not covering some of the state-of-the-art opinion summarization methods. As there are already multiple surveys touching the sentiment classification task, in our survey, we focus purely on the recent techniques used in opinion summarization that goes beyond sentiment classification or uses sentiment classification as one of the components in summarization.
3. Miao.Q. et al., (2010) calculated the ranking by analyzing the quality of a sentiment with a study on the sentiment mining and retrieval system and classified the sentiment as positive or negative using the naive bayes algorithm. Sharma.A. et al., (2012) also used artificial neural networks and sentiment lexicon for document level sentiment classification. Morae.R. et al., (2013) attempted a back propagation neural network based document level sentiment prediction for movie reviews.
4. Hwang.J.et al.,(2008) developed part of the sentiment word manually and by forming it into a feature vector they classified the document as positive or negative with a supervised learning algorithm. Denecke. K (2009) performed a rule based classification and a machine learning classification by using an average score and a frequency of word by class as well as a number of parts of speech of a word as a feature.
5. Prabowo.R. et al., (2009) described an extension by combining rule based classification, supervised learning and machine learning into a new combined method. For each sample set, they carried out 10-fold cross validation. For each fold, the associated samples were divided into training and a test set. For each test sample, a hybrid classification is carried out, i.e., if one classifier fails to classify a document, the classifier passes the document onto the next 24 classifiers, until the document is classified or no other classifier exists. Neural networks have seen a rapid growth over the years, and are being applied successfully in various application domains for the classification problems. But the state of the art techniques for neural network based text sentiment classification are found to be rare from the literature (Morae.R. et al (2013), Sharma.A. et al., (2012), Zhu. J. et al.,(2010)) .
6. Martin Wallance[2008] presented in his study how Neural networks provide forecasts of market prices and actions. These can then form the basis for trading the market in an automated system. A pre-trained network is the natural choice for real-time trading. The implementation of forecasts requires a strategy for dealing with adverse market moves; the question of when to enter or exit the market is also largely determined by forecasts, hence neural networks always have a role in finance.

# **3.Software Requirements**

We use GPU support version CuDNN version of LSTM, GRU and RNN. (CuDNN is way faster than a normal algorithm) training parameters for all these algorithms are the same. We used acer predator Helios 300(i5-8gen,16gb ram, gtx 1060 ,6gb vram) as our hardware platform, we chose Tensorflow, keras as our deep learning platform. Duration of each epoch in RNN,LSTM and GRU are 19sec,28sec and 32sec respectively. In each model 50 epochs are called.

# **4.Requirement Analysis**

Our problem is to improvise the accuracy of prediction of open prices of Google stocks.We mainly aim to predict the nature of stock related tweets of live 100 stock using # Google stock and which can further recommend to invest in Google stock in the current scenario using LSTM.

For that we studied the difference between different machine learning algorithms and models.Predicting stock prices is one of the best ways to take a step closer to predict one’s future. The shareholder’s gain depends on the successful forecast of the stock’s future. The proposed model compares different ML algorithms like LR, RNN, GRU, LSTM, Random Forest, etc. To figure out the best suited algorithm, i.e., algorithms which return fewest errors based on an imported benchmark dataset. We compare various algorithms upon the graph of their predicted data, after training the data. In the second part twitter semantic analyzer is created using a twitter API on a developer ID. This twitter API is used to access consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy is used to import tweets from twitter. Class TextBlock is used to analyse the tweets and return avg. semantic value (polarity) of all tweets. In the third part, after predicting three values by using LSTM, we used an algorithm based on semantic analysis, that will use the predicted values and will help decide based on current tweet trends for the corporation whether investing will be Loss or Profit. Our principal goal is to analyse distinct types of research done already in this field and existing stock prediction models to predict the stock prices with least error. Then implement the semantic analyser to predict the nature of the tweets and according to response. Build a scale for Investment suggestions..

***Index terms****- Linear Regression(LR), Recurrent Neural Network(RNN),Gated Recurrent Unit(GRU), Long Short-Term Memory(LSTM).*

## 4.1 Related Work

This section will cover some of the related and prior work done in the are of stock prediction

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Citation | Feature,Extraction methods,  Classifier and Algorithm used | Accuracy Percentage |
| 1 | Y. Guo, S. Han, C. Shen, Y. Li, X. Yin and Y. Bai, "An Adaptive SVR for High-Frequency Stock Price Forecasting," in IEEE Access, vol. 6, pp. 11397-11404, 2018. | * Support machine regression * Adaptive SVR * Particle swarm optimization |  |
| 2 | Oyewola, David & Dada, Emmanuel & Ezekiel, Omole. (2019). Predicting Nigerian Stock Returns using Technical Analysis and Machine Learning. European Journal of Electrical Engineering and Computer Science. 3. 1-8. 10.24018/ejece.2019.3.2.65. | * Classification problem with two class values * Random Forest * Logistic Regression * SVR * Neural Network |  |
| 3 | Hegazy, Osman & Soliman, Omar S. & Abdul Salam, Mustafa. (2013). A Machine Learning Model for Stock Market Prediction. International Journal of Computer Science and Telecommunications. 4. 17-23. | * Least Square Support Vector Machine * Particle Swarm Optimization * Levenberg-Marquardt (LM) algorithm |  |
| 4 | A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, 2017, pp. 506-509. | * Linear Regression |  |
| 5 | R. Iacomin, "Stock market prediction," *2015 19th International Conference on System Theory, Control and Computing (ICSTCC)*, Cheile Gradistei, 2015, pp. 200-205. | * ANN * Support Vector Machines * Principal component analysis |  |
| 6 | Wanjawa, Barack Wamkaya(2012), “A Neural Network Model for Predicting Stock Market Prices at the Nairobi Securities Exchange”, University of Nairobi research archive, Volume – 73404, Page- 56. | Feed forward multilayer perceptron with error bark propogation |  |
| 7 | Khan, W., Malik, U., Ghazanfar, M.A. *et al.* Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Comput* (2019). https://doi.org/10.1007/s00500-019-04347-y | * Sentimental analysis * Sequential minimal optimization * ANN | 20% |
| 8 | Dongdong, Lv & Yuan, Shuhan & Li, Meizi & Xiang, Yang. (2019). An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy. Mathematical Problems in Engineering. 2019. 1-30. 10.1155/2019/7816154. | * Deep neural network (DNN) * Long Short-Term Memory |  |
| 9 | J. S. Park, H. Sung Cho, J. Sung Lee, K. I. Chung, J. M. Kim and D. J. Kim, "Forecasting Daily Stock Trends Using Random Forest Optimization," *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea (South), 2019, pp. 1152-1155. | * Random forest technique |  |
| 10 | Qiu J, Wang B, Zhou C (2020) Forecasting stock prices with long-short term memory neural networks based on attention mechanism. PLoS ONE 15(1): e0227222. | * LSTM * GRU * RNN |  |
| 11 | Zhuge, Qun, Lingyu Xu, and Gaowei Zhang. "LSTM Neural Network with Emotional Analysis for Prediction of Stock Price." *Engineering letters* 25.2 (2017). | * LSTM * RNN * MLP * Emotional analysis |  |
| 12 | Nelson, David MQ, Adriano CM Pereira, and Renato A. de Oliveira. "Stock market's price movement prediction with LSTM neural networks." *2017 International joint conference on neural networks (IJCNN)*. IEEE, 2017. | * LSTM * ANN |  |

# **5.System Architecture and Design**

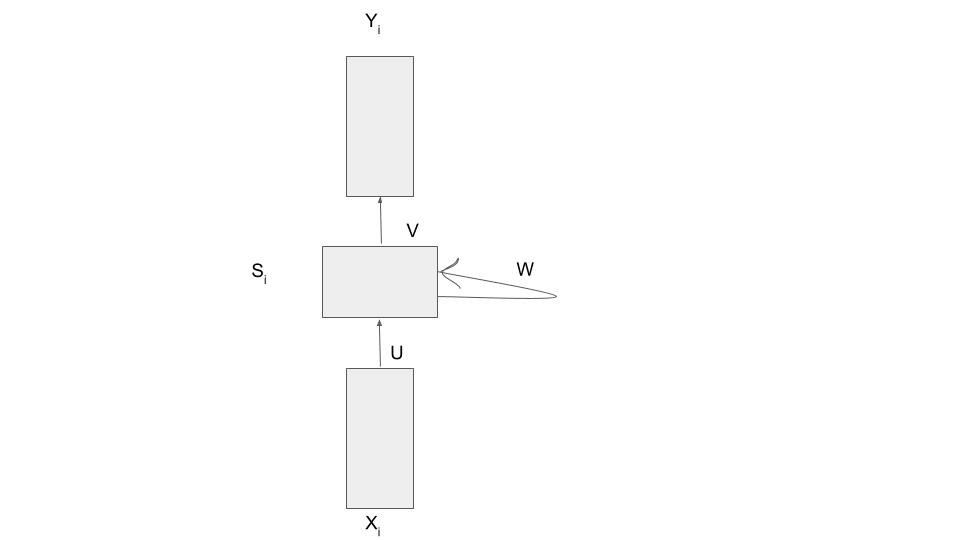
### 5.1 Recurrent Neural Network

**Recurrent Neural Network**

Stock Prices for every day is one of the nice examples of sequence data, this is a real-world example where we can apply RNN, we can’t use traditional models like feed-forward neural network for modelling the sequence data efficiently as different sequences have different lengths and it is hard to model with feed-forward neural networks. Also, sequence dependency is not taken into account when we use something like a feed-forward neural network to model these sequences.

Therefore we use RNN that is used for modeling sequential data.

In RNN we use to model and we share the weights across all the time steps In the sequence the model weights are shared and then we apply the model on the input coming at each time step and then we get the output and that output is fed to the next time step and that is used to calculate the output. So, this output is combined with a dot product of input with its parameter vector. Also, the dot product of this output is done with its parameter vector, and we add a bias term and perform tanh activation on that. Then we proceed in this manner till the end of the sequence.[3].



Si=σ(Ux+Wsi-1+b)

yi=σ(Vsi+c)

Or

yi=f(xi,si,W,U,V)

Si is the state of the network at timestep i.

The parameters are W,U,V,b,c which are shared across timestep .

The same network(& parameters) can be used to compute y1,y2, ….or y100.

### 5.2 GRU

GRU is the non- trivial recurrent system.GRUs are improved versions of standard recurrent neural networks. To solve the vanishing gradient problem of a standard RNN, GRU uses, so-called, update gate and reset gate. Basically, these are two vectors that decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or removing information which is irrelevant to the prediction. [4]

Gates:

Ot=σ(Wo St-1+Uo xt +bo)

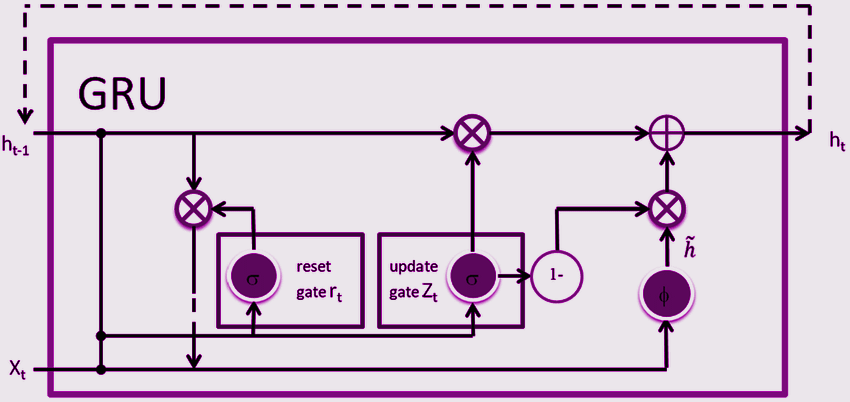
it=σ(WiSt-1 + Ui xt +bi)

States:

št = σ(W(Ot ⊙ St-1) + Uxt +b)

St = (1-it)⊙St-1+it ⊙št

The gates depend directly on St-1 & not the intermediate ht-1 as in case of LSTM’s.

****

### 5.3 *LSTM Networks*

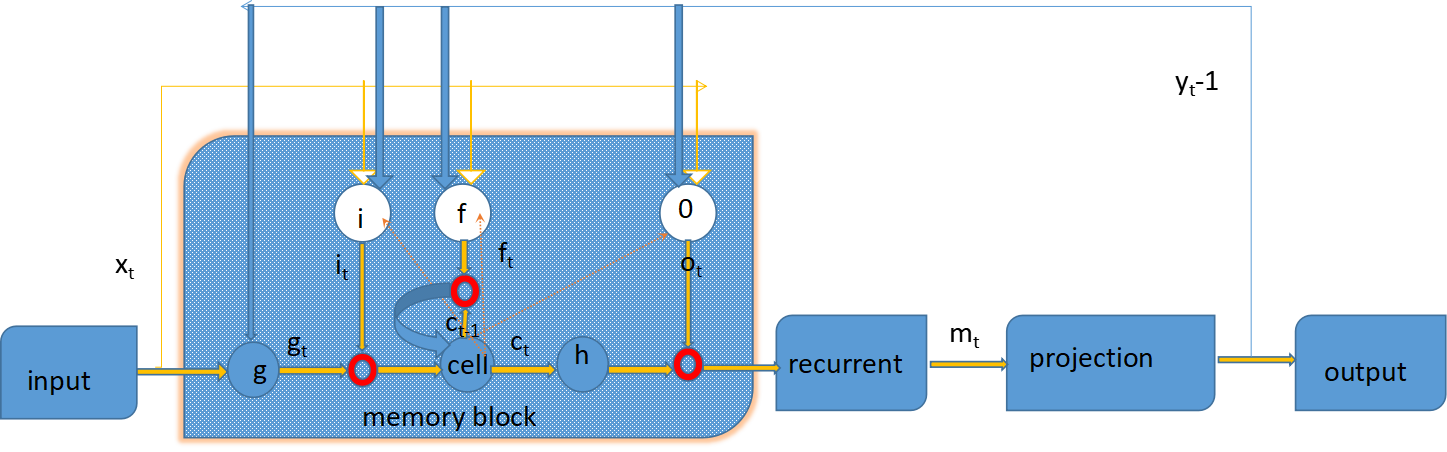
Long Short Term Memory networks (LSTM)

They can also be used as special RNN.

The algorithm's existence came in 1997

designed by Hochreiter & Schmidhuber.They were designed to mitigate the vanishing and exploding gradient problems or we can say to avoid long-term dependency problems.

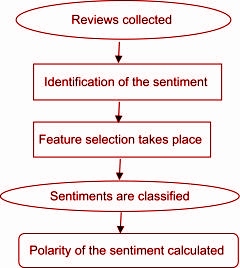
Apart from the hidden state vector each LSTM maintains a cell state vector and at each time step the next LSTM can choose to read from it or write to it or reset the cell using a gating mechanism.Each unit has three gates of the same shape. Each of them is like a binary gate.The input gate controls whether the memory cell is updated.The forget gate controls if the memory cell is reset to zero.Output gate controls whether the current state is made visible.They all have a sigmoid activation.It's used to constitute smooth curves in the range 0 to 1. And also so that the model remains differentiable. Apart from these gates we have another vector called c bar that modifies the cell state and it has tanh activation. Tanh distributes gradients. Hence prevents vanishing/exploding.Each gate takes the hidden state and the current input X as its inputs and concatenates the vectors and applies a sigmoid. C bar represents a new candidate value that can be applied to the cell state.

****

### 5.4 Twitter Sentiment Analysis

The information available from social networks is beneficial for business analysts for mining the user opinion about their products and considers these opinions as feedback to improve their policies, planning and process for product development. Sentiment analysis is used to extract such opinions and remarks of users by classifying them as positive, negative and natural sentiment . Although there are a number of definitions about sentiment analysis in the literature, in simple terms sentiment analysis is a technique used to extract intelligent information based on the person’s opinion from raw data available on the internet. In this definition, the term opinion means a person’s perspective about an object or issue.

There are some challenges related to sentiment analysis, the first challenge is a word that is used to express an opinion; it can be positive as well as negative depending upon the type of sentiment.[6][9][10]



As the live Twitter Data is going to be accessed, the user needs to import the Tweepy (Python Client for official Twitter API) and TextBlob. [6][9]][10]

# **6.Project Planning**

### 6.1PROBLEM STATEMENT:

Our problem is to improvise the accuracy

of prediction of open prices of Google stocks

We mainly aim to predict the nature of stock related

tweets of live 100 stock using # Google stock

and which can further recommend to invest in

Google stock in the current scenario using LSTM.

For that we studied the difference between different

machine learning algorithms and models.

### 6.2METHODOLOGY USED TO ANALYZE THE ABOVE PROBLEM:

We have divided the above problem in three parts

● The first part deals with the different prediction models - RNN,GRU and LSTM.We find the best model that gives the result with the least error. So that the prediction can be accurate. [2][3][4][5]

● The second part deals with the sentimental analysis.It is used to review the polarity of the tweets and according to that give us a fundamental mathematical integer from 1 to -1

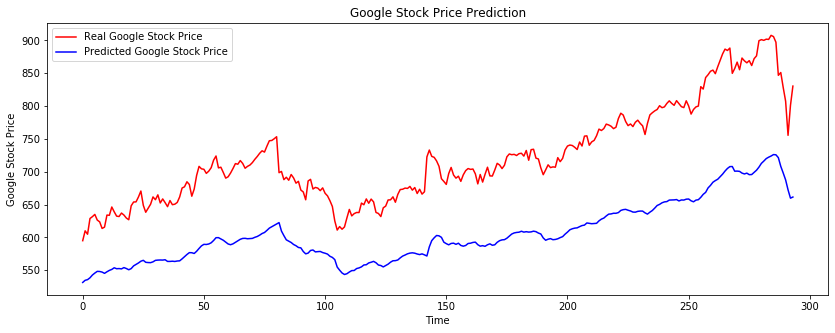
● The third part deals with how we devise if it is useful to invest in the stocks or not. By building an Investment suggestion meter.

# **7.Implementation**

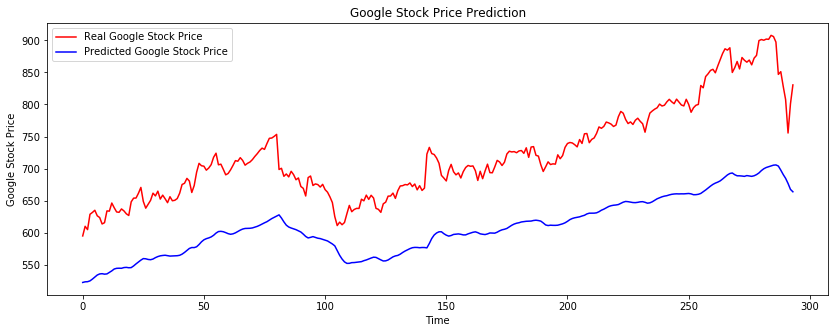
## 7.1 Part 1

LSTM and GRU were the most successful ones out of all deep learning-based models. One of the important factors in the formation of this result is less effect of vanishing gradients problem in these models. The GRU models work much faster than LSTM models and have achieved very similar results throughout the runs.GRU model does not show sufficient performance even if it has close error values according to the LSTM model, especially where the fluctuations are high in data-set.With wider time intervals negatively affected the RNN model with interrupted gradient flow. LSTM models contain a more sophisticated memory than GRU models, so it has the ability to predict more temporal related data. GRUs will perform better than LSTMs when data is limited or the risk of overfitting is high. In this part we are going to determine which of these neural network models is going to work best for stock prediction. Here we are going to use google stock data set to train our models and compare them which one of them did the best prediction.

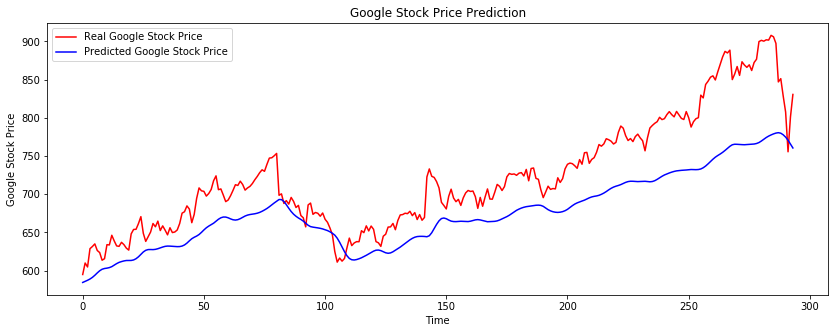
### 7.1.1 Results :



Output of RNN model



Output of GRU model



Output of LSTM model

## 7.2 ***For Part 2:***

Here we use sentiment analysis on tweets. We use tweepy to access data from twitter and then use textblob to find polarity of those tweets data.

### 7.2.1Access Live Data and Processing

I. Tweepy: - The users have to register their app on the Twitter website and then get the tweets. This is done at the Application Management of the Developers Section of Twitter. This is a very important step for the OAuth Authentication of Tweepy tool .

After creation of the application, the user needs to generate private keys. These are required for “OAuthHandlertakes” as the parameters. This is shown in dev.twitter.com. After successful generation of all the keys, we will copy the keys which will be further used in our Algorithm.

II. TextBlob: - The data in the text format is usually processed using the Textblob Library. TextBlob objects are treated as if they were like Python strings that learned to do the natural language processing

Algorithm Description

The algorithm proposed in our work has mainly 3 major steps.

• Authenticate twitter account.

• GET Request is made to twitter for getting the tweets.

• Select the tweets. Segregate each tweet positive, negative.

The user has to make a twitter api-client. The given class contains the function which allows us to access the tweets. **SentimentAnalysis** function is used for authentication purposes. The clean tweets is used for cleaning the dataset acquired by the twitter api. For this the user has to import the Regular Expression library which is available in Python .

Function the user uses the following piece of code to call the API to get the tweets is: - self.tweets = tweepy.Cursor(api.search, q=searchTerm, lang = "en").items(NoOfTerms)

Function the user uses the TextBlob module is:-

analysis = TextBlob(self.clean\_tweet(tweet))

A classifier function divides the tweets as positive and negative polarity in the range of -1.0 to 1.0. Creation of Sentiment Classifier :-

•TextBlob contains a Movie data-set where positive and negative reviews have already been labelled.

•From each positive and negative review, positive and negative features are extracted respectively.

•This data with positive and negative features is now trained on Naive Bayes Classifier. Then, the user can see a TextBlob class, where sentiment.polarity method can be used to get the polarity of tweets between -1 to 1.

Then, we calculate the average polarity of those tweets and we return it.

## **7.3 For part 3**

This part is a combination of part 1 and part 2. In this part we use the best algorithm from part 1 (LSTM) and also used sentiment analysis part to calculate polarity of tweets and after that we design an algorithm based on sentiment analysis and stock prediction outputs. In this part we predicted exactly 3 future stock prices of GOOGLE stock. For this we use the same method as part 1 but this time training Data is used from 2004-08-19 to 2020-04-01. Let's consider those predicted values y\_pred1 , y\_pred2 and y\_pred3 .

For 2 nd step we called sentiment analysis as a parameter(#googlestock , 100) here we pass the search term for which we want to polarize and for how many no of tweets we want to run. This code return avg polarity on #googlestock of 100 tweets.

Then we design an algorithm for which provides us with a scale to how profitable would be if anyone invested in provided stock. If a suggestion meter is positive then there is profit in future and suggestion meter is negative then there is loss in future.

***ALGORITHM DESIGN***

*if(y\_pred1<y\_pred2<y\_pred3):*

*x=1*

*x=(x+polarity)/2\*100*

*elif(y\_pred1<y\_pred2>y\_pred3):*

*x=0*

*x=(x+polarity)/2\*100*

*elif(y\_pred1>y\_pred2<y\_pred3):*

*x=.50*

*x=(x+polarity)/2\*100*

*elif(y\_pred1>y\_pred2>y\_pred3):*

*x=-.50*

*x=(x+polarity)/2\*100*

Here the range of x is in between -100 to +100. Higher the value of x(+ve) then higher the Profit is going to be in Future. Lower the value of x(-ve) then higher the Loss in Future. We also get the percentage of profit or loss value, for the stock we are trying to predict, if the value of x is greater than zero then output will be percentage profit, i.e. percentage recommendation to invest in that stock.

Result-

We are getting values for the search “#googlestock”, y\_pred1 as 1136.5042, y\_pred2 as 1162.5419 and y\_pred3 as 1117.313.By using above values and above formulae we are getting polarity of 0.11392121212121215. So, as polarity is positive therefore our system is recommending to invest and stated it as “Weakly Positive” with 5.6960606060% .

# 

# **8.ScreenShots of project:**

Figure 1 & 2 are part of twitter semantic analyzer which will be used in the final code.

Figure 3,4&5 are the final code which is used to predict the stock prices and also checks the polarity and recommends whether to invest or not by telling the future percentage of profit or loss of that stock.

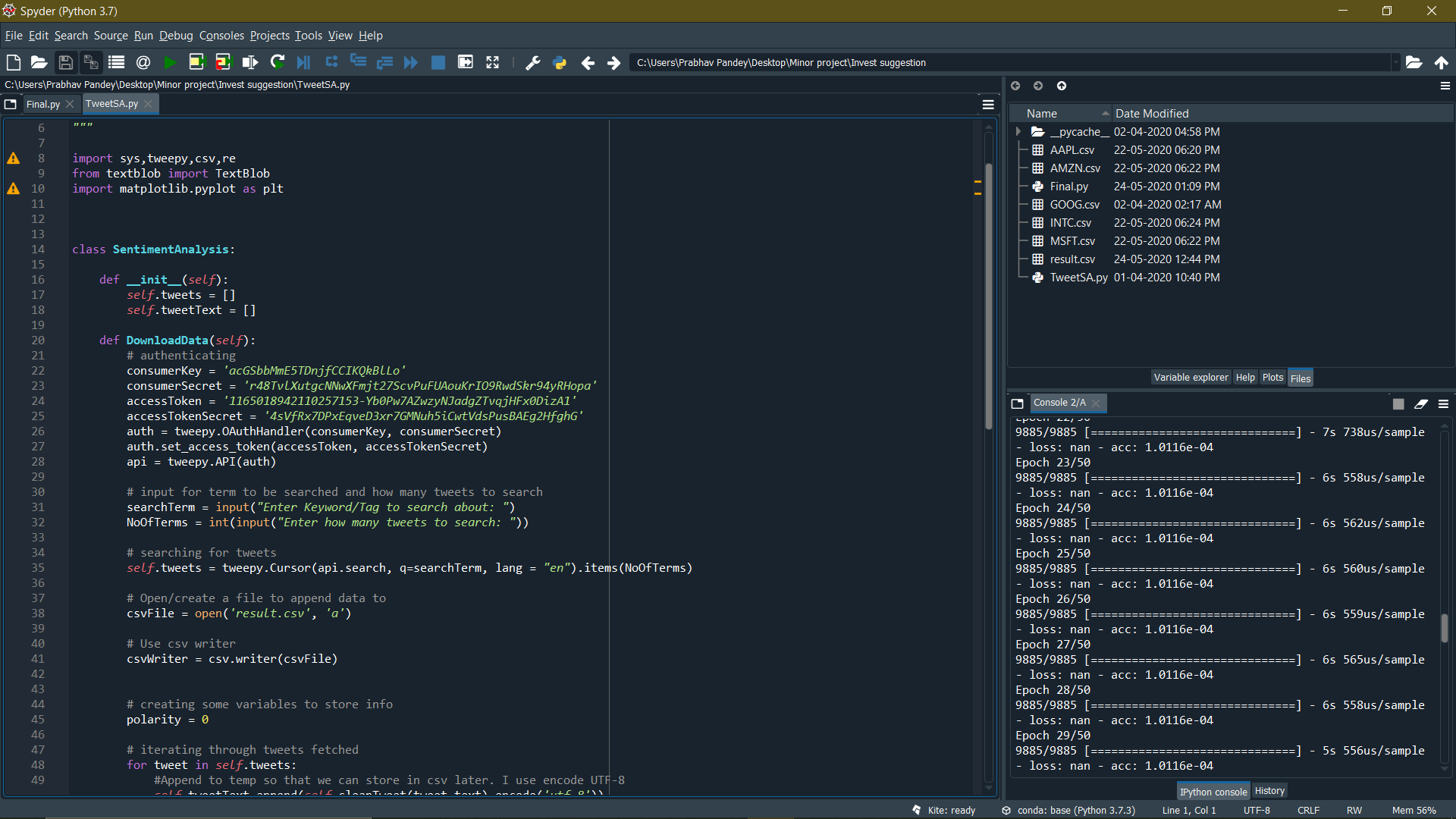


Fig-1

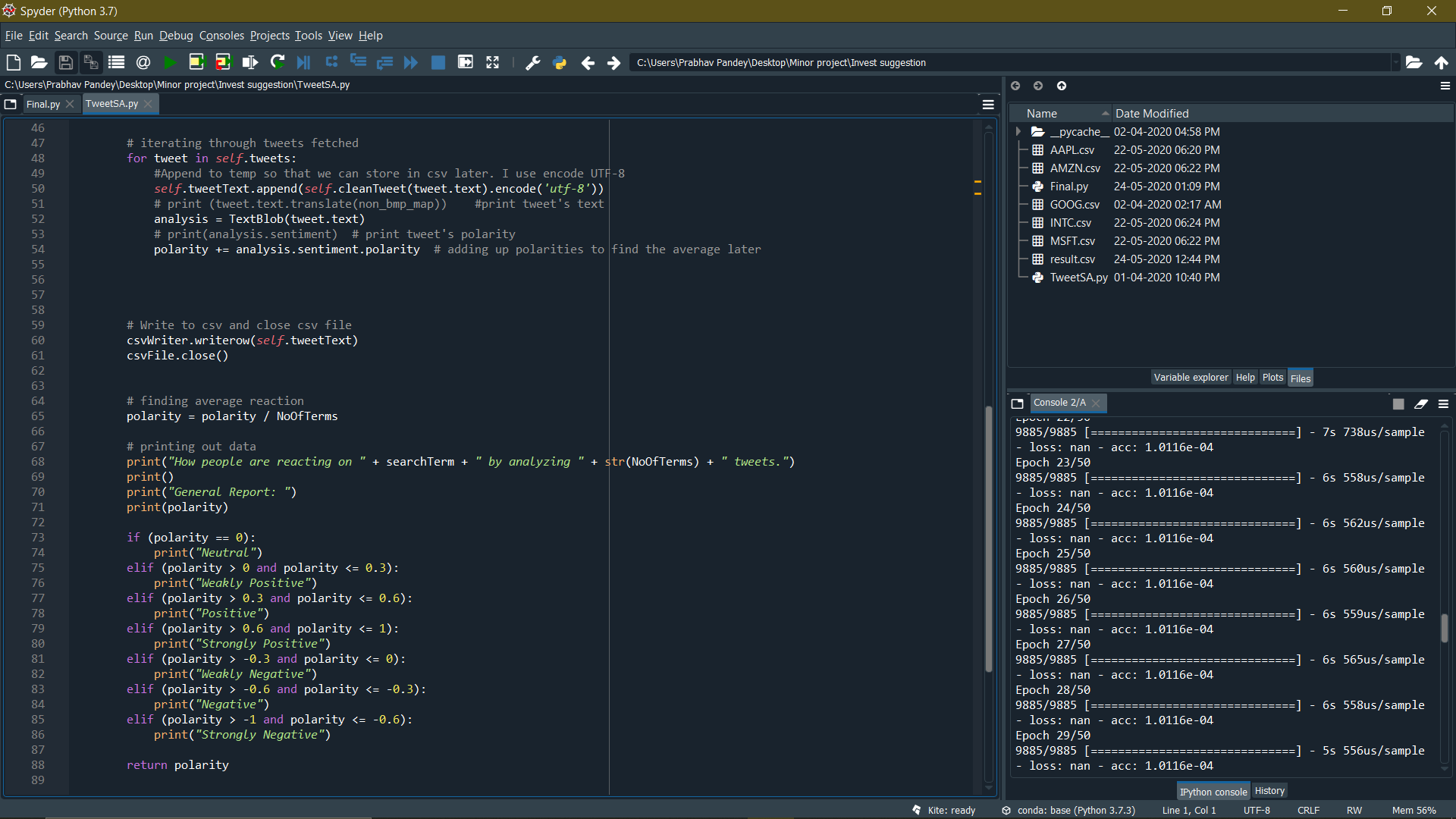


Fig-2

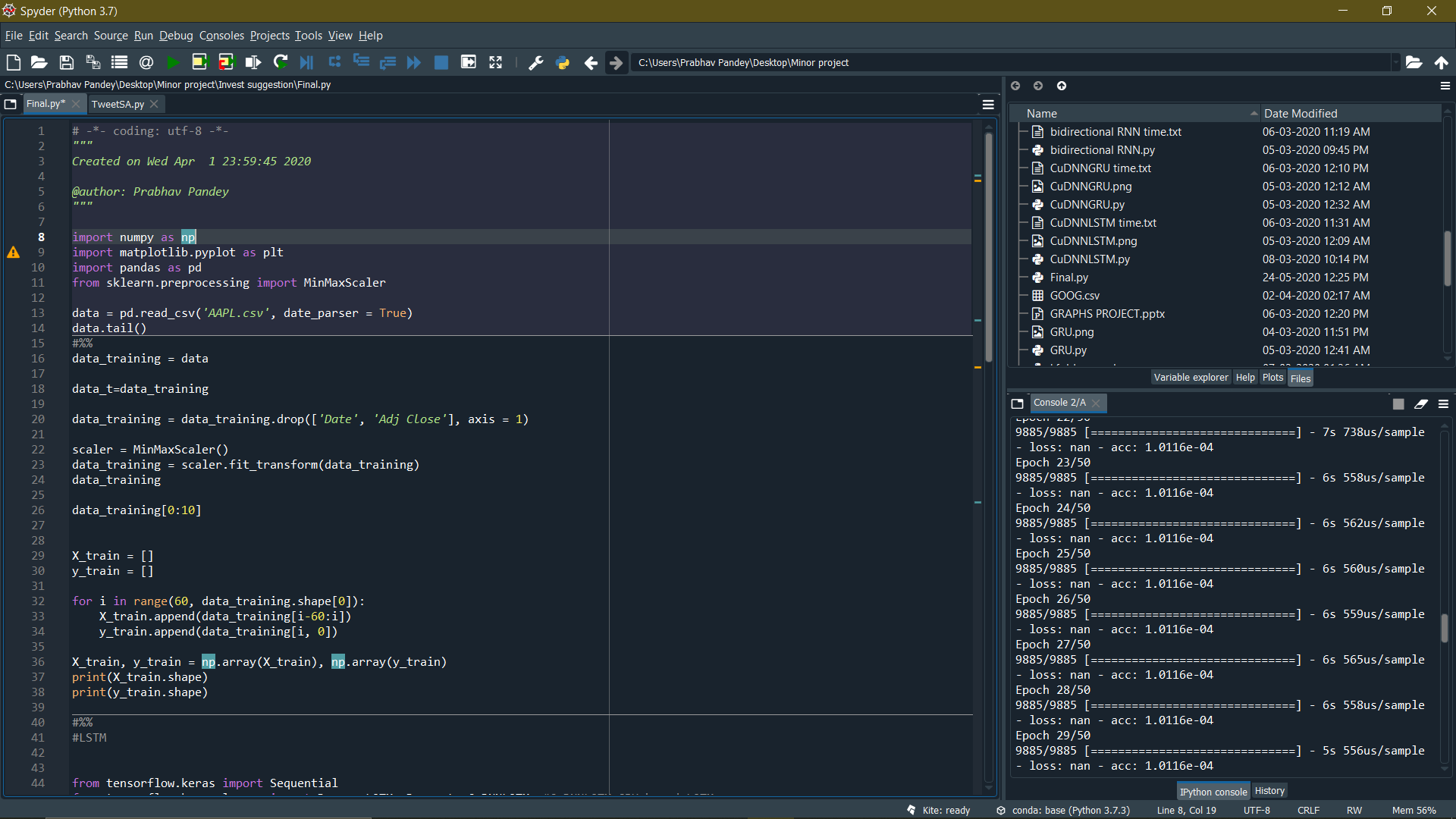


FIG-3

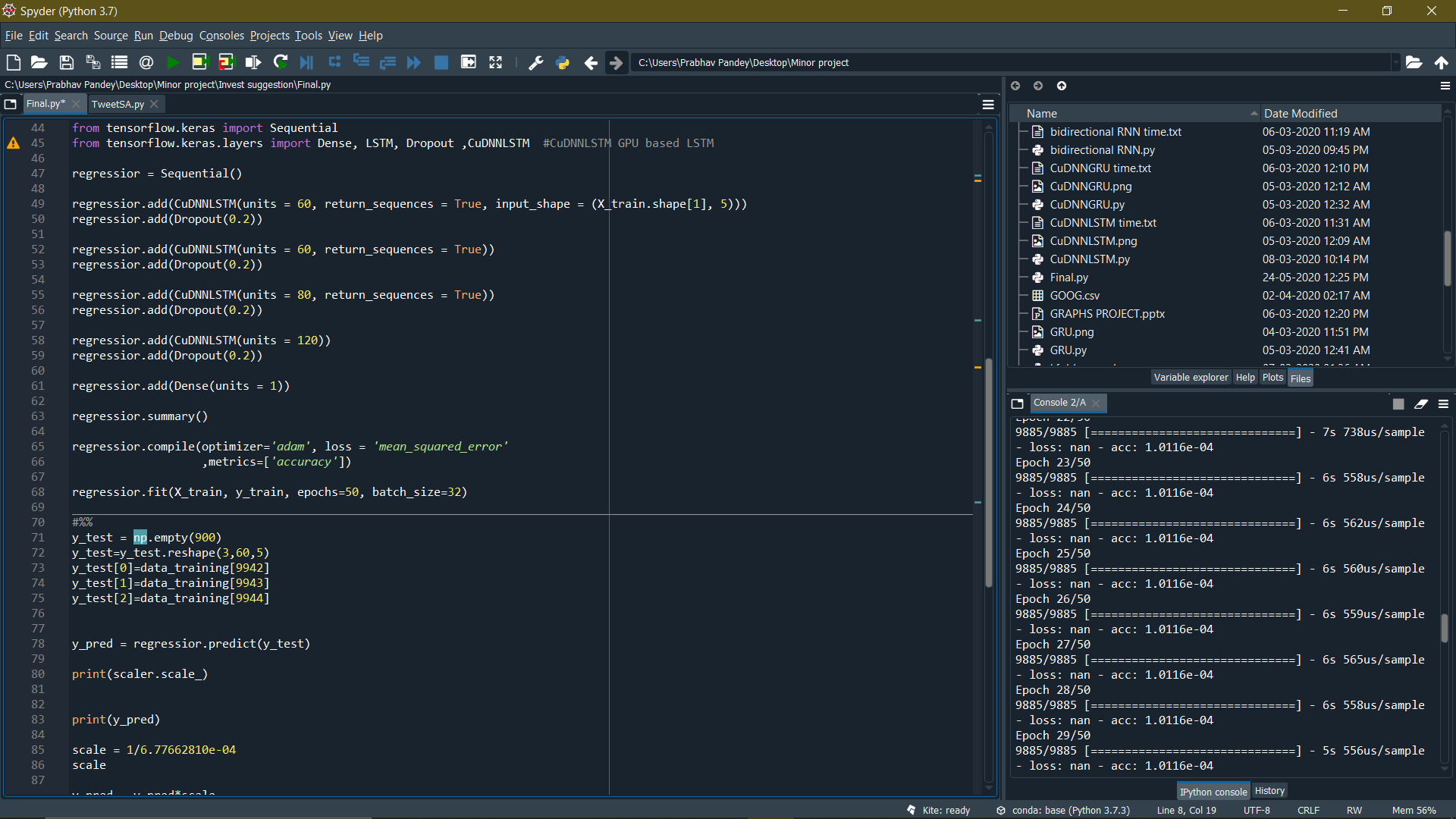


FIG-4

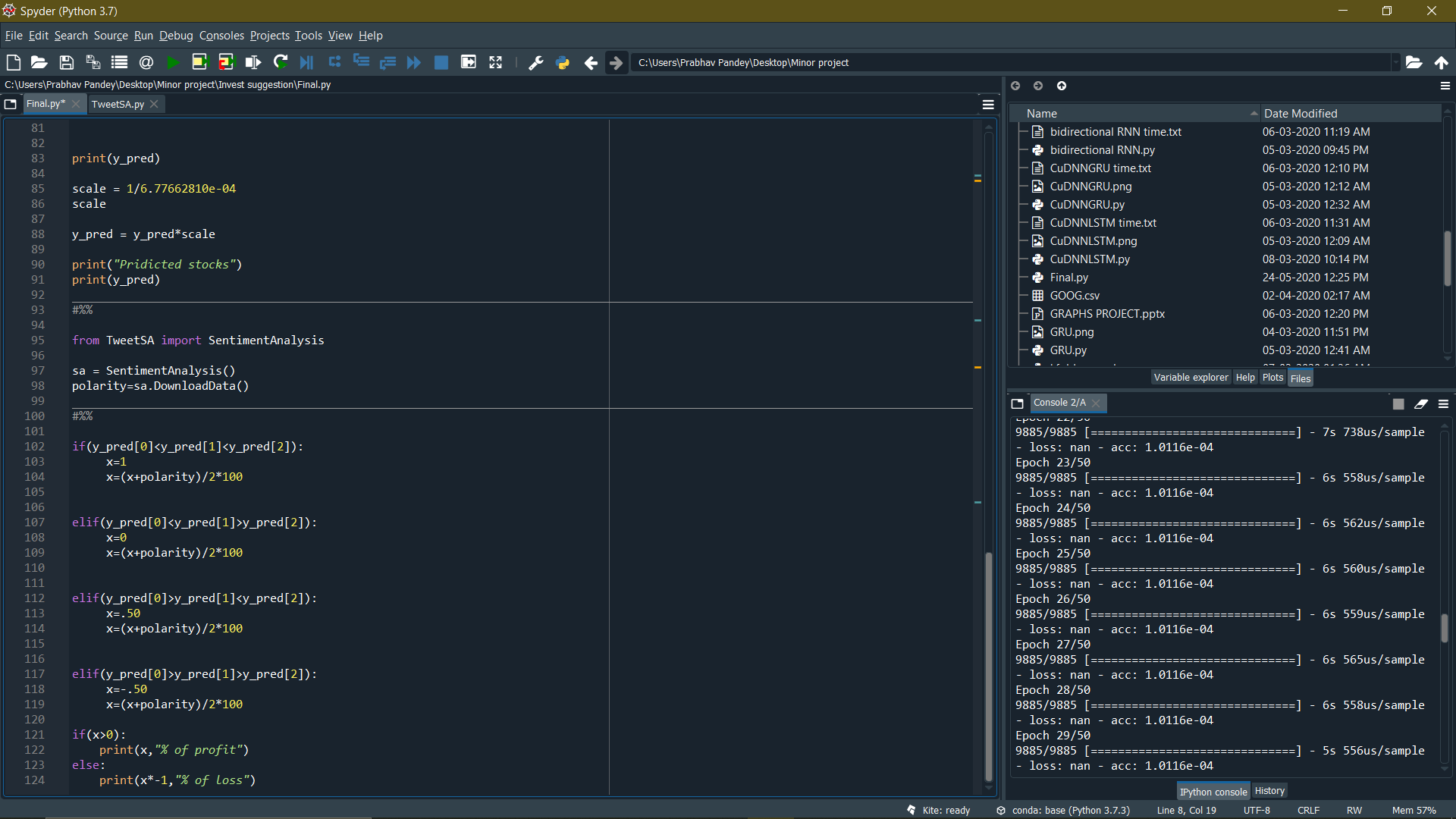


FIG-5

9.**CONCLUSIONS AND FUTURE WORKS**

This paper proposed a model based on a combination of LSTM algorithm with tweet based semantic analysis.

For RNN the predicted value is not even close to real values. For GRU the predicted value has lesser time lag as compared to the RNN but still it's not even close to real values. For LSTM the time lag between the predicted and trained values is very less; it shows the best output compared to other neural networks. As conclusion of this part we find out LSTM is the most appropriate neural network model for stock prediction.

Feasibility analysis, and verify the effectiveness of the proposed algorithm i.e LSTM and optimize results with less error and at a faster speed. Real-time sentiment analysis Of Google stock related tweets with a successful investment meter which shows profit loss of investment.

# **10.Refrences**

[1]https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/stock-market/

[2]Lai, Chun & Chen, Rung-Ching & Caraka, Rezzy. (2019). Prediction Stock Price Based on Different Index Factors Using LSTM. 1-6. 10.1109/ICMLC48188.2019.8949162.

[3]https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/

[4]https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be

[5]https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[6].Z. Kechaou, M. Ben Ammar, A.M. Alimi Improving e-learning with sentiment analysis of users’ opinions

IEEE International Conference on Global Engineering Education (2011), pp. 1032-1038

[7]. A. Sharma, D. Bhuriya and U.

Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) , Coimbatore, 2017, pp. 506-509.

[8]. R. Iacomin, "Stock market

prediction," 2015 19th International Conference on System Theory, Control and Computing (ICSTCC) , Cheile Gradistei, 2015, pp. 200-205.

[9]. Wanjawa, Barack Wamkaya(2012),

“A Neural Network Model for Predicting Stock Market Prices at the Nairobi Securities Exchange”, University of Nairobi research archive, Volume – 73404, Page- 56 .

[10]. Khan, W., Malik, U., Ghazanfar,

M.A. et al. Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. Soft Comput (2019). https://doi.org/10.1007/s00500-019- 04347-y

[11] Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8.

[12] Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of marketing, pages 74–89.

[13] Yu, S. and Kak, S. (2012). A survey of prediction using social media. arXiv preprint arXiv:1203.1647

[14] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

[15] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira, “Computational intelligence and financial markets: A survey and future directions,” Expert Systems with Applications, vol. 55, pp. 194 – 211, 2016.

[16] P. C. Tetlock, “Does Public Financial News Resolve Asymmetric Information?” The Review of Financial Studies, vol. 23, no. 9, pp. 3520–3557, 2010.

[17] Chris Wang, Yilun Xu, Qingyang Wang “Novel approaches to sentiment analysis for stock prediction,” Stanford University, USA, 2018

[18] Hakan Gündüz, Zehra Çataltepe, Yusuf Yaslan, “Stock daily return prediction using expanded features and feature selection,” Turkish Journal of Electrical Engineering & Computer Sciences, 4829- 4840, 2017

[19] Colah, Understanding LSTM Networks, 2015, [online] Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

**INDIVIDUAL CONTRIBUTION REPORT:**

**STOCK PREDICTION USING ML TECHNIQUES**

NAMAN GERA

1729036

**Abstract:** The primary objective of this project is to do prediction of stocks

using various machine learning techniques. Firstly, different ML

algorithms like LR, RNN, GRU, LSTM, Random Forest etc. were

compared to determine the best suited algorithm, i.e., algorithms which

returns least errors based on an imported benchmark dataset. Various

algorithms were compared upon the graph of their predicted data,

after training the data. A twitter semantic analyzer was created using a twitter API on a developer ID. This twitter API uses consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy was used to import tweets from twitter. Class TextBlock was used to analyse the tweets and return avg. semantic value (polarity).

**Individual contribution and findings:** After finding from weka LR and RF not giving desired prediction, moved to RNN as it was best for sequence data. Also sequence dependency is not taken into account when we use something like a feed forward neural network to model these sequence. The results were close but not upto the mark.

**Individual contribution to project report preparation:**  Prepared the diagrams for all the neural networks, added section 1.1 in introduction. Added section 3. Software requirements, added section 1.5 Recurrent Neural Network in software architecture and design.

**Individual contribution for project presentation and demonstration:** Added slides 7, 10 and 15, demonstrated the same.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………

**INDIVIDUAL CONTRIBUTION REPORT:**

**STOCK PREDICTION USING ML TECHNIQUES**

PARITOSH SINGH

1729044

**Abstract:** The primary objective of this project is to do prediction of stocks

using various machine learning techniques. Firstly, different ML

algorithms like LR, RNN, GRU, LSTM, Random Forest etc. were

compared to determine the best suited algorithm, i.e., algorithms which

returns least errors based on an imported benchmark dataset. Various

algorithms were compared upon the graph of their predicted data,

after training the data. A twitter semantic analyzer was created using a twitter API on a developer ID. This twitter API uses consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy was used to import tweets from twitter. Class TextBlock was used to analyse the tweets and return avg. semantic value (polarity).

**Individual contribution and findings:** Learned Weka 3.8.4 visualisation tool for generating graphs. Provided statistical data based on visuaisation results. Originally worked on LR and RF, decided to remove them from this project as they produced undesired results due to specifications mismatch and technical failures.

**Individual contribution to project report preparation:**  Prepared abstract and handled division of topics from the same. Provided graphical results in section 7.1.1. Added section 1.1 in introduction and section 4. Requirement analysis.

**Individual contribution for project presentation and demonstration:** Added slides 4, 5, 13 and 17, demonstrated the same.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………

**INDIVIDUAL CONTRIBUTION REPORT:**

**STOCK PREDICTION USING ML TECHNIQUES**

PRABHAV PANDEY

1729045

**Abstract:** The primary objective of this project is to do prediction of stocks

using various machine learning techniques. Firstly, different ML

algorithms like LR, RNN, GRU, LSTM, Random Forest etc. were

compared to determine the best suited algorithm, i.e., algorithms which

returns least errors based on an imported benchmark dataset. Various

algorithms were compared upon the graph of their predicted data,

after training the data. A twitter semantic analyzer was created using a twitter API on a developer ID. This twitter API uses consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy was used to import tweets from twitter. Class TextBlock was used to analyse the tweets and return avg. semantic value (polarity).

**Individual contribution and findings:** Provided all technical support, handled the algorithm design and the coding involved in this project.

**Individual contribution to project report preparation:**  Provided the screenshots of the code in section 8 and added sections 7.3.

**Individual contribution for project presentation and demonstration:** Added slides 18, 19, 20 and 21, demonstrated the same.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………

**INDIVIDUAL CONTRIBUTION REPORT:**

**STOCK PREDICTION USING ML TECHNIQUES**

SHIVANGI MISHRA

1729060

**Abstract:** The primary objective of this project is to do prediction of stocks

using various machine learning techniques. Firstly, different ML

algorithms like LR, RNN, GRU, LSTM, Random Forest etc. were

compared to determine the best suited algorithm, i.e., algorithms which

returns least errors based on an imported benchmark dataset. Various

algorithms were compared upon the graph of their predicted data,

after training the data. A twitter semantic analyzer was created using a twitter API on a developer ID. This twitter API uses consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy was used to import tweets from twitter. Class TextBlock was used to analyse the tweets and return avg. semantic value (polarity).

**Individual contribution and findings:** Learned python and helped in algorithm design of the project. Supervised and reviewed the paper and report related to the projects. Also researched over works related to the topic of our choice.

**Individual contribution to project report preparation:**  Added sections 2, 4.1, 5.3 and 9.

**Individual contribution for project presentation and demonstration:** Added slides 3, 6, 9 and 16, demonstrated the same.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………

**INDIVIDUAL CONTRIBUTION REPORT:**

**STOCK PREDICTION USING ML TECHNIQUES**

SUBHAM SINGH

1729073

**Abstract:** The primary objective of this project is to do prediction of stocks

using various machine learning techniques. Firstly, different ML

algorithms like LR, RNN, GRU, LSTM, Random Forest etc. were

compared to determine the best suited algorithm, i.e., algorithms which

returns least errors based on an imported benchmark dataset. Various

algorithms were compared upon the graph of their predicted data,

after training the data. A twitter semantic analyzer was created using a twitter API on a developer ID. This twitter API uses consumer Keys, consumer secret, access tokens and access token secret. A class Tweepy was used to import tweets from twitter. Class TextBlock was used to analyse the tweets and return avg. semantic value (polarity).

**Individual contribution and findings:** Learned Python, Machine learning and Deep learning algorithms. Helped in algorithm design. Provided dataset of different stock holders exchange of previous years and refined them for our project.

**Individual contribution to project report preparation:**  Added section 5.2, and 6(all).

**Individual contribution for project presentation and demonstration:** Added slides 8, 11 and 12.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………