International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 03 | Mar 2023 www.irjet.net p-ISSN: 2395-0072

Data-Driven Approach to Stock Market Prediction and Sentiment Analysis

Shreya Bamane[1], Mitali Belge[2], Shweta Sharad Nikam[3], Ankita Umale[4], Monica Charate[5]

[1, 2, 3] Computer Science and Technology, Usha Mittal Institute of Technology, SNDT Women's University
[4 5]Professor, Dept. Computer Science and Technology, Usha Mittal Institute of Technology, SNDT Women's University

---

Abstract - *Predicting stock market prices has been an interest- ing topic because of high gains against investment over a short period of time, therefore it interests analysts and researchers for a long time. Stock prices are hard to predict because of their highly volatile nature which depends on diverse political and economic factors, change of political factors, investor sentiment, and many other factors. The method of predicting stock prices that are based on historical data or textual information alone has been proven to be insufficient. Existing studies studied by analysts in sentiment analysis have found that there is a very strong correlation between the movement of stock prices and the publication of news articles. Several sentiment analysis studies have been attempted at various levels using different machine learning algorithms such as Random Forest, etc. The prediction system in this paper shows both the results of predicting past data using Recurrent Neural Network(RNN) with Long Short- Tem Memory(LSTM) i.e RNN-LSTM algorithm and sentiment analysis using Support Vector Machine (SVM) algorithm to improve the accuracy of stock price prediction.*

Key Words: **RNN-LSTM, SVM, stock market, sentiment analysis**

1. INTRODUCTION

The financial market is dynamic because it keeps changing and is a composite system. This is a place where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms or apps supported by brokers. The stock market has allowed investors to own shares of public companies through trading, exchange or over the counter markets. This market has given investors a great chance of gaining money and having an opportunity to live a prosperous life through investing small initial amounts of money which is low risk compared to the risk of opening a new business or the need for a high salary career that demands a great amount of effort. Stock markets are influenced by factors which are in large amounts causing the uncertainty and change in the market.

The stock market appears in every newspaper, every day. Since it is quite often intensively discussed people are always keen on knowing what will happen. People could have good returns on their investments if proper methods and algorithms are used to predict. Previous methods of stock predictions involve only the use of a single method that is using historical data. But stock markets are so volatile that they are mainly affected by events happening in the world. Therefore sentiment analysis of a stock is also required. This paper combines two things, one is historical data to predict the possible values of stocks and two, sentiment analysis to understand the sentiment for a company, giving an investor a better understanding of the stocks future. In this project the problem is solved, with a system constructed to predict news polarity which may affect changes in stock trends.

Time-series prediction which uses historical data is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses continuous data that is the historical data over a period of time to predict the outcome of the next unit of specified time period. Numerous time series forecasting algorithms have been effective in practice. The most common algorithms are currently based on recurrent neural networks (RNNs) and a special form, Long-Short Term Memory (LSTM), it is a type of RNN. The stock market is a representative area that represents time series data, and many researchers have studied it and presented various models.

For sentiment analysis Support Vector Machine (SVM) algorithm is used, it is a machine learning algorithm that analyzes data and recognizes patterns or decision boundaries in a data set, and is primarily used for classification and regression analysis. SVMs can handle multiple different types of variables, this nature of SVM makes it the preferred algorithm for sentiment analysis. In Module 2, of this paper SVM algorithm is used to predict the sentiment of a particular company in the market on the basis of news headlines.

2. RELATED WORK

2.1 Literature Survey

The Stock Market Price Prediction Using LSTM RNN

[1] has created the system for stock price prediction using RNN with LSTM. It takes the historic time series data of different stocks from Yahoo Finance. The data is preprocessed and regularized. The set LSTM model is trained with the train data, the optimizer used is Adam optimizer. After the training is done, the LSTM model is tested using test data of different stocks. The design has enforced LSTM for stock price vaticination and it has redounded in RNN- LSTM giving better results than the traditional machine learning algorithm. The Stock Market Prediction using ANN [2] proposed a design that tries to prognosticate stock prices using machine learning ways on the NSE. It uses direct retrogression and SVM regression. Linear regression is used for prognosticating the open price of the stock for the coming day using the close price of the stock for the former day. SVM regression is used for prognosticating the difference between close and open prices of the stock for the coming day. They trained the ANN model by using stock data. colorful features similar as stochastic pointers, moving pars, and RSI are uprooted from the literal stock data. External factors like foreign exchange rate, NSE indicator, moving parts, relative Strength indicator etc are used to get accurate results. The dataset is also divided into training and testing sets which are used for training and testing the delicacy of the ANN model. The design attempts to prognosticate whether a stock price occasionally in the future

will be advanced or lower than it's on a given day.

The Sentiment Analysis of Twitter Data for Predicting Stock Market Movements [3] have applied sentiment analysis and supervised machine learning principles to the tweets taken from twitter and dissected the correlation between stock request movements of a company and sentiments in tweets. This design employed two different textual representations, Word2vec and Ngram, for assaying the public sentiments in tweets. Then they actually contributed to the field of sentiment analysis of twitter data. Sentiment analysis is the task of judging opinion in a piece of textbook as positive, negative or neutral. The disadvantage of this design is that they've considered only twitter data for assaying people's sentiment which may be poisoned because not all the people who trade in stocks partake their opinions on twitter.

The Twitter Sentiment Analysis Approaches A Survey [4] has surveyed different exploration papers of different algo- rithm and styles used to achieve sentiment analysis in twitter. Social media has proven to have its affect on stock prices. In this check, they've distributed the approaches into four orders machine learning, wordbook- grounded, mongrel (combines machine learning and wordbook-grounded approaches) and graph- grounded approaches. Over 40 papers of recent explo- ration on TSA (Twitter Sentiment Analysis) were compactly reviewed and categorized. From the discussion it was con- cluded that TSA will be an active exploration area in the coming times.

In An SVM-based approach for stock market trend predic- tion [5] they have used Hidden Markov Model to predict the close price of the stocks of next day. They have used a ton of historical stock prices data of companies that are Apple Inc., IBM Corporation, TATA Steel and Dell Inc. Inputs that were given were the High Price, Low Price, Open Price and Close Price of the next day. The model had been first trained for 7 months. Model testing was done using MAPE values.

In Stock market prediction using hidden markov models

[6] they have taken the SVM based way for the prediction price of stock market trends. The problem has been solved in two parts, i.e. selection of features and direction prediction of trends that are going on in market. They have used SVM correlation for selection of the features which affects the price. Linear SVM has been applied to the data series to predict direction. The system has shown to select good feature and control the overfitting that occurs on stock market prediction. In Ensemble model for stock price movement trend predic- tion on different investing periods [7] first they have selected the most relevant features for the prediction outcome of the stock price by using calculations of the maximal information coefficient. They have built their assembler model by using the three different outstanding classifiers on stock market trend prediction SVM, Random forest and AdaBoost have been intelligently named as the SRAVoting. Their model has been used on Chinese Stock Market and the conclusion is that SRAVoting gives higher accuracy than SVM but at the same

time lesser buy/sell strategies than SVM.

H. Gunduz, Z. Cataltepe and Y. Yaslan [8] have used deep neural network techniques for stock price prediction. Simi- larly, in Stock market prediction using an improved training algorithm of neural network [9], they have suggested some further improvements for stock price prediction using neural networks using a good performing training algorithm which they have designed by themselves. In Stock index prediction using regression and neural network models under non nor- mal conditions [10] have suggested techniques themselves on handling a non-normal situation which often rises during the performance of the system and cause great disruptions or lead to incorrect predictions. In Stock transaction prediction mod- elling and analysis based on LSTM [11] they have conducted similar works and have designed a model for

applying the LSTM to stock prediction with a good amount of scope for improving the stock prediction accuracy.

In Evaluation of bidirectional LSTM for short and long-term stock market prediction [12] they have further contributed to the field by creating better staging experiments and simula- tions for having good feasibility of applying deep learning techniques for the prediction of stock price.

In [13] they have shown that ANN model works well even if there is no clear relation between attributes and output. But disadvantages that are unignorable are that time required for prediction is far more than other methods. ANN faces an issue with overfitting problem.

A comparative literature survey is done on Support Vector Machine to prove that results of ANN are more accurate than

SVM. But disadvantages found are only attributes having clear relationship can be supplied as attributes[14]. If not then the accuracy will be reduced.

In Predicting the Future with Social Media.[15] they found out that there is a correlation between trading volumes of stocks traded in NASDAQ-100 and query volumes (i.e., the number of user requests submitted to search engines on the Internet).

In Twitter mood predicts the stock market[16], they conducted a research in which their findings were found about Dow Jones Average Index, in that they noticed that there is a negative correlation between sentiments like hope, fear and worry in twitter tweets.

In Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation[17] they have used Pearson correlation coefficient for investigation of the correlation of sentiments of the public with stock increase and decrease for stocks.

B. Existing System

The Existing system for Stock Price Prediction [1] uses RNN with LSTM. It uses Historical Stock Data i.e. time series dataset of the stock for prediction. A time series dataset is the data taken in a series of data points indexed in time order. This dataset works well with the RNN for prediction. In this System, we have taken all the required data from Yahoo finance to collect the stock datasets of various companies such as Apple, Google, and Tesla. The Dataset is then preprocessed which includes data cleaning and normalization. The collected Historical Stock Data is used to get Predicted Stock Prices for n Data Points. Then Adjacent Close values are retrieved from Data. The Data is Normalized using the Adjacent Close values. After the Normalization of the Data, the Data is then split into train data and test data. The Train model is obtained using the RNN-LSTM input layer and hidden layers, using the loss function and the optimizer. Then the obtained model is trained and tested and is used to plot a graph of Predicted v/s True Data.

The model is tested using varied trial values for loss function of mean squared error and the Adam optimizer is used- Adam is an optimization algorithm which is used to update the weights of the network iterating based on training data. Several LSTM architectures are tested to find the best model with the lowest loss value. The loss observed, which is the lost value for the LSTM architecture after three epochs, is 0.5770. The loss observed for the Deep LSTM after three epochs is 3.1464e−04, which is quite better than the other models. As the LSTM architecture has got the best loss value, the behavior is also checked by varying the inner architecture, i.e. the number of cells and the number of layers keeping constant the activation function of the hidden layer, here hidden layer as hyperbolic tangent function and of the output layer as the rectified Linear Unit function are observed to give the best results. Then the Historical data of Apple, Google and Tesla are tested using 10 epochs and the graph of Predicted vs True Data is plotted for each of the stocks. The results show that the RNN-LSTM model is prone to give more accurate results than the traditional machine learning algorithms.

3. RELATED WORK

In this project, two main algorithms are used, Long Short Term Memory (LSTM) and Support Vector Machine (SVM).

3.1 Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) [18] was introduced by Hochreiter and Schmidhuber in 1997 to cope with the problem of long-term dependencies. LSTM is capable of processing sequential data. A very large architecture can be successfully trained using LSTM thus used in deep learning. LSTM is a type of RNN network structure which has three “gate” structures. The three gates placed in an LSTM unit are called input gate, forgetting gate and output gate. While information is used in LSTM’s network, it can be selected by rules. One gate is dedicated for reading out the entries from the cell, the output gate. Another gate like a input to the program is needed to decide when data should be read into the cell, this is called the input gate. Finally a gate that will help to reset the content collected by the cell, that is the forget gate. This design was used in order to decide when to remember and ignore inputs at the hidden state. Only the information accepts to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate. By using gates they were tremendously able to improve memory capacity and control the memory cell.

Min-max normalization (commonly referred to as feature scaling) performs a linear transformation on the original data. This method gets all scaled data in the range (0, 1). The formula to achieve this is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

In this project, historical stock data table containing the information of open, the high, low, close, volume of a particular day to predict the coming day's stock prices is taken. The data in the dataset are explained as below

OPEN: The price of the first trade for any company or a script in a stock exchange is its opening price for that day.

CLOSE: The volume ladened normal of all the trades happened during the last half an of trading price.

HIGH: The loftiest price for which a stock is traded on a script in a day.

LOW: The smallest price for which a stock is traded on a script in a day

VOLUME: Number of stocks traded in a day

Here, LSTM in layers with timesteps of 120 days is being used.

3.2 Support Vector Machine (SVM)

Support Vector Machine or SVM [19] is a Supervised learning algorithm, which is used for Classification as well as Regression problems. Primarily, it's used for Classification problems in Machine Learning. Developed at AT&T Bell Laboratories, SVMs are one of the most robust prediction styles, being grounded on statistical learning frameworks. The purpose of the SVM algorithm is to produce the best line or decision boundary that can insulate n- dimensional space into classes so that it can effortlessly put the new data point in the correct order in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is nominated as Support Vector Machine. SVM can be of two types

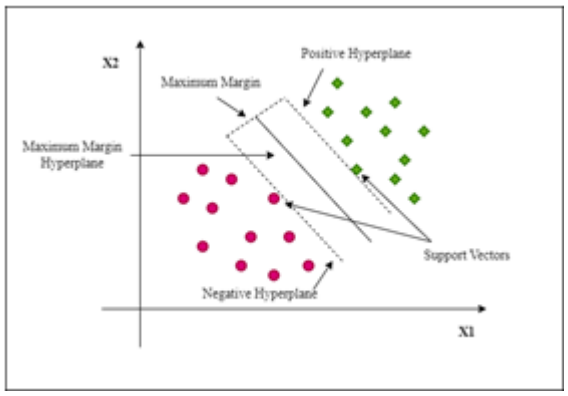

Fig -1: SVM

Linear SVM: Linear SVM is used for linearly divisible data, which means if a dataset can be classified into two classes by using a single straight line, also similar data is nominated as linearly divisible data, and classifier is used as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset can not be classified by using a straight line, also similar data is nominated as non-linear data and classifier used is called as Non-linear SVM classifier.

The data is news captions, attained using a web news scraper targeting a particular stock. The data is also collected from the news scraper, it's pre-processed to be used for the bracket. After the word vectorization of the data is completed, the training dataset is also fed to the SVM algorithm with a direct kernel for the classification and training. also the test dataset is used with the SVM for prediction of the sentiment.

4. PROPOSED WORK

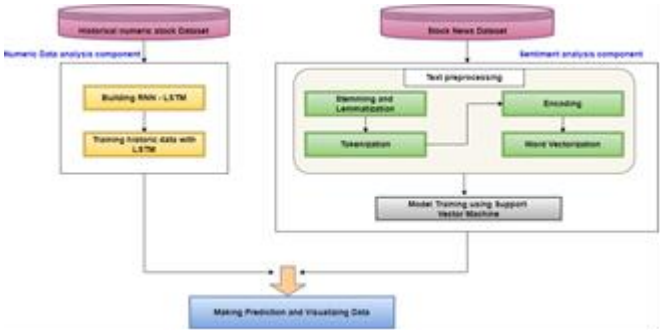

Fig -2: Mechanism of Proposed System

For years, research has been done on predicting stock prices based on historical stock price data alone. In this system, financial news articles from well-known sources are used to avoid fake news that may be prevalent on social media. It includes both stock price prediction and sentiment analysis. This approach is better because

financial news related to the company has a significant effect on its stock price. Hence, taking into consideration the financial news of a company instead of only considering the past stock prices can lead to a better prediction system. This system will give better accuracy and will be robust as compared to the other systems. This system uses Recurrent Neural Network with its special type Long Short Term Memory (LSTM) for the historic data prediction and Support Vector Machine for the sentiment analysis part.

This section describes each component of the proposed model:

In the historic data analysis component, the stock related dataset of companies will be taken from Yahoo Finance. This dataset for the respective company includes the open, close, high and low values for a given day. It will take 120 days of stock data of the company. Further it divides the data into two parts, training data and testing data, where 75% of the data will be used for training and 25% of the data will be used for testing. Then the data is preprocessed, i.e. feature scaling, creating an array data structure to store the data. Now as the dataset is ready it can be used for the RNN- LSTM model and train the historic data with LSTM and predict the output. The output will be plotted in the form of a graph.

In Sentiment Analysis component, the analysis of stock news data is performed as follows:

For news data, the objective is to classify news to be either positive or negative sentiments. So news headlines are extracted from the websites. Then these headlines are preprocessed, followed by news classification using the Support Vector Machine algorithm.

The following section describes the details of the proposed steps.

Text preprocessing: There are a several preprocessing steps that are performed as follow:

Stemming: Porter stemmer is applied on the data to return each word to its stem and remove suffixes such as (-Ed,-ing,-ion…etc.) to reduce the complexity in the document and minimize the processing time which improves the model performance.

Lemmatization: It is a text normalization technique that switches any kind of a word to its base root mode. Lemmatization is mostly responsible for grouping or categorizing different inflected forms of words into the root form, which has the same meaning (Caring returns Care).

Tokenization: Each news article or financial report document is split into meaningful words called tokens (For example, the text “It is raining” can be tokenized into ‘It’, ‘is’, ‘raining’).

Encoding: Text encoding is a conversion process to convert meaningful text into number vector representation so as to preserve the context and relationship between words and sentences, such that a machine can understand the pattern associated with any text and can make out the context of sentences.

Vectorization: It is a classic approach of converting input data into vector form to reduce the complexity of dealing with text data.

After data preprocessing, the Support Vector Machine features algorithm is used. Each headline is labeled as positive or negative. Then, the feature extraction is done and SVM is trained to predict the sentiment on news headlines.

4.1 Workflow of Proposed Model

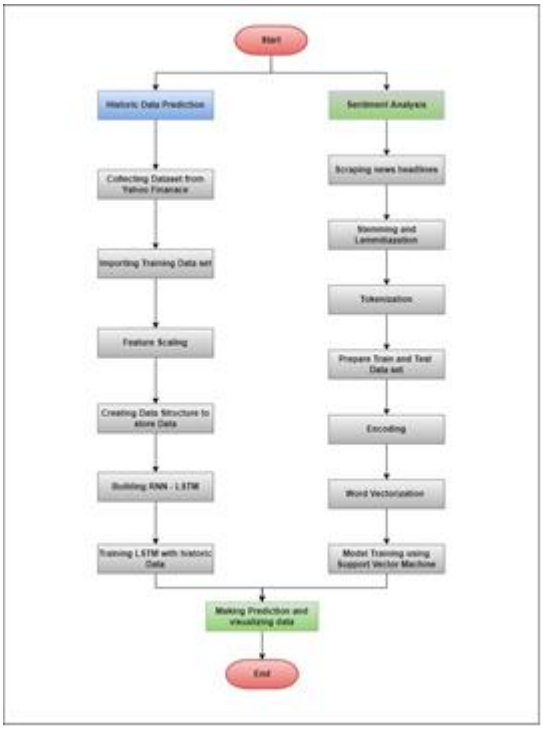

Fig -3: Workflow Diagram

The Stock Market Trend Prediction and Analysis system is divided into two major modules. The first module is Historic Data Prediction and the second module is the Sentiment Analysis. Investors will therefore benefit from an automated system that forecasts stock prices. An automated system can gather financial news related to the companies of interest in real time and can execute a machine learning model on those data to predict the sentiment, along with historical stock price information, to

predict price. The workflow for both these modules is explained in detail below:

1) **Historic Data Prediction:** Historic data analysis is an inevitable part of stock market prediction. For this system, historical stock price data is used. Dataset of respective companies is gathered from Yahoo Finance website which includes the open, close, high and low values for a given day. The system takes 120 days of stock data of the company. Further it divides the data into two parts, training data and testing data, where 75% of the data is used for training and 25% of the data is used for testing. The system then preprocesses the data i.e feature scaling, creating an array data structure to store the data is executed. After the preprocessing the dataset is ready for the constructed RNN- LSTM model, for training the LSTM model with historic data and predict the output. The output of this model will be the next day's open, close, high and low prices of a particular stock that are predicted.

2) **Sentiment Analysis:** Sentiment analysis is a crucial component of this approach because the output of this module is used to determine the trend of a specific stock. Social media sentiment analysis is a great source of data and can offer insights that can show positive or negative views on stocks and trends.

   In this module, news headlines are extracted from the web and the data is pre-processed and cleaned (i.e. Stemming, Lemmatization, Tokenization). The input for this module is the news headlines of the stock, each headline is then classified as positive, negative and neutral, afterwards the feature extraction is done and then the data is used to train the Support Vector Machine to predict the sentiment.

5. IMPLEMENTATION DETAILS

The Dataset is gathered from Yahoo Finance website for the historical stock prices of companies.

The first dataset is to train the RNN-LSTM model and the second dataset is to test and predict the values. The training dataset would have 5 years of data, which is from 1st January 2017 to 1st January 2022. The Testing dataset would have 120 days of data that is from 1st January 2022 to 1st May 2022. 120 days period is used for testing because the average price of a particular stock is calculated for 120 days and as it is a good point in between long and short time periods. After all the unnecessary columns are deleted and only the ones which are required are kept which are open, high, low, close and volume. The code is executed in Google Colab and then the csv files are uploaded to the session storage. Then the libraries numpy, matplotlib. pyplot and pandas are imported.

The training set is then imported from session storage to print it. Feature scaling is the process of normalizing the range of features in a dataset. The datasets usually contain features that are varying in magnitude, range and units. Therefore, for this machine learning model to interpret these features on the same scale feature scaling is implemented. After feature scaling, RNN-LSTM model is constructed. In this the LSTM layers are added to the stack with dropout out regularization. Then model is then trained and tested with training and testing dataset which were imported. The predicted values are visualized using graphs for clarity and better understanding of the results.

Sentiment analysis module first collects news headlines from the website www.finviz.com[20]. This data is collected in an excel sheet and then it is labeled into two labels: label 1

= negative; label 2 = positive. This labeled data then goes through word preprocessing: stemming, lemmatization, word vectorization. After the data is preprocessed it is split into train and test datasets where the ratio of train and test data set is set to 80:20 respectively. This data is then used to train and test the SVM algorithm for sentiment analysis.

6. RESULTS

The module 1 of the system results a graph of the company stock created on the basis of prediction value from the RNN-LSTM model. The module 2 of the system which performs sentiment analysis results into the sentiment prediction of news headlines using the SVM machine learning algorithm to predict sentiment as negative or positive for a particular headline and the accuracy for apple stock.

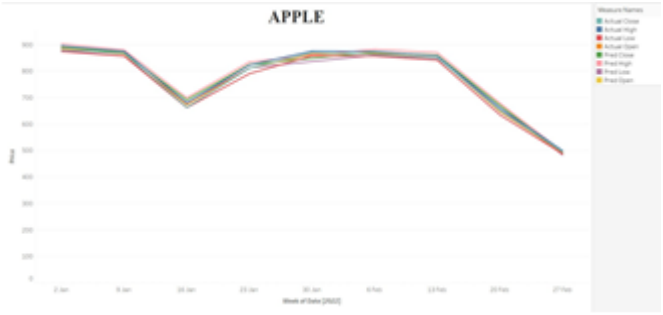

Fig -4: Tableau Representation of Apple Price

	A	B	C
1	Headlines	Labels	Predictions_SVM
2	apple airtags and bluetooth trackers be officially a billion-dollar industry here's what to know, trends, and the best ways to invest	1	1
3	apple introduce faster macbook pros and mac minis	1	1
4	2 top stock to buy in 2023 and hold forever	1	1
5	the zacks analyst blog highlight lenovo group, hp, dell technologies and apple	1	1
6	qualcomm: a rare technology value play	0	1
7	history say the nasdaq could soar in 2023 -- 5 stock you'll wish you'd buy if it do	1	1
8	here be 2 technology stock of the future you can buy right now	1	1
9	great week for apple inc. (nasdaq:aapl) institutional investors after lose 22% over the previous year	1	0
10	be it time to buy the dow jones' 3 worst-performing december stocks?	0	0
11	apple debut macbook pro and mac mini with new high-powered m2 pro and m2 max chip	1	1
12	apple will delay ar glasses, focus instead on cheaper mixed-reality headset: report	0	0
13	apple ceo tim cook's pay cut in half	0	0
14	apple do audit relate to its human right policy and labor practice this year	1	1
15	apple be take on meta in virtual reality. its not go to be easy.	1	0
16	281 billion reason why you may regret not buy apple stock	1	1
17	stock trend in after hours: alcoa, discover financial, vroom, apple	1	1
18	spotify join media firm to urge eu action against apple's 'unfair' practice	0	0
19	why apples mac price be bad for customersand bad for business	0	0
20	apple get a boost in india as chinese suppliers give clearance	1	1

Fig -5: Sentiment Predictions using SVM

Fig -6: SVM Accuracy for APPLE

Both modules are incorporated in the website for the consumer. The consumer first has to sign up to use the system. After that they can search for the company in the search bar which they want to know about.

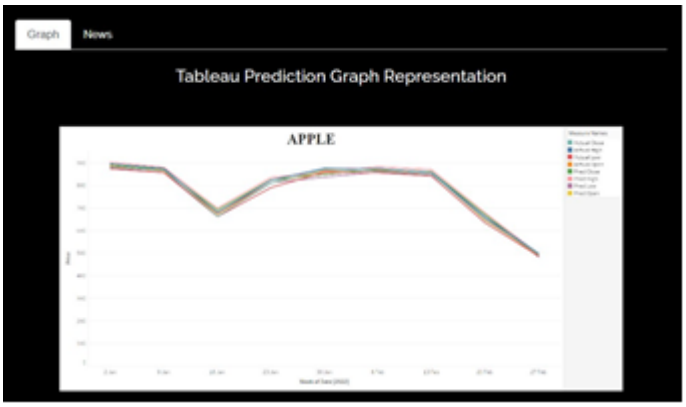

Fig -7: Historical data Tableau Representation

Fig -8: News headlines sentiment analysis

7. CONCLUSION

The aim of this system is to analyze the historic data of the stocks, and train the machine learning algorithm with it and make predictions on the future trend of the stock and Taking the sentiments into consideration via news headlines and perform sentiment analysis on them using SVM machine learning algorithm and give the current sentiment about the stock. Unlike traditional stock market forecasting systems, the new approach incorporates the sentiment of the stock market participants via news feeds and moving averages of stock prices. This web application would help the user to take in consideration two factors for a particular stock i.e. the future trend and the sentiment for it, before considering making a decision for investment in stock market. This system would be useful for users who are unaware of the stock market. This helps the new users to make a decision who don't have go through the hassle to take a look into the history of a stock to make an investment and research the news about it as they can use this system to achieve that easily.

8. FUTURE WORK

The future work for this system is to improve the accuracy of the stock prediction using more hybrid algorithms as they might yield better accuracy. A feature like Risk analysis can be added to check how performance of other stocks affect a particular stock, which can be used to consider the future of the stock. More visualization tools can be added for more pictorial representation of the predictions, which would make understanding the stock trend more understandable. A combination of results can be used to produce a more realistic and hybrid result for prediction, that will give more points to take into consideration for making an investment as well as analyzing the stock.

REFERENCES

[1] K. Pawar, R. S. Jalem, and V. Tiwari, "Stock Market Price Prediction Using LSTM RNN," in *Emerging Trends in Expert Applications and Security*, ser. Advances in Intelligent Systems and Computing, Springer, vol. 841, 2018, pp. 29–38.

[2] M. Gurjar, P. Naik, G. Mujumdar, and T. Vaidya, "Stock market prediction using ANN," *International Research Journal of Engineering and Technology*, vol. 5, no. 3, pp. 2758–2761, 2018.

[3] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *2016 international conference on signal processing, communication, power and em- bedded system (SCOPES)*, 2016, pp. 1345–1350.

[4] O. Adwan, M. Al-Tawil, A. Huneiti, R. Shahin, A. A. Zayed, and R. Al-Dibsi, "Twitter sentiment analysis approaches: A survey," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 15,pp. 79–93, 2020.

[5] Y.-H. Lin, H.-W. Guo, and J. Hu, "An SVM-based approach for stock market trend prediction," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, IEEE, 2013, pp. 1–7.

[6] A. Gupta and B. Dhingra, "Stock market prediction using hidden markov models," in *Engineering and Systems (SCES), 2012 Students Conference on*, IEEE, 2012,pp. 1–4.

[7] J. Yang, R. Rao, P. Hong, and P. Ding, "Ensemble model for stock price movement trend prediction on different investing periods," in *Computational Intelligence and Security (CIS), 2016 12th International Conference on*, IEEE, 2016, pp. 358–361.

[8] H. Gunduz, Z. Cataltepe, and Y. Yaslan, "Stock mar- ket direction prediction using deep neural networks," in *2017 25th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2017, pp. 1–4.

[9] M. Billah, S. Waheed, and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," in *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, IEEE, 2016, pp. 1–4.

[10] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," in *INTERACT-2010*, IEEE, 2010, pp. 59–63.

[11] S. Liu, G. Liao, and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, 2018, pp. 2787–2790.

[12] K. A. Althelaya, E. M. El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in *2018 9th International Conference on Information and Communication Sys- tems (ICICS)*, IEEE, 2018, pp. 151–156.

[13] K. Lucas, C. Lai, N. James, and K. Liu, "Stock forecasting using support vector machine," in *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, IEEE, 2010, pp. 1607–1614.

[14] R. I. Rasel, N. Sultana, and N. Hasan, "Financial Instability Analysis using ANN and Feature Selection Technique: Application to Stock Market Price Prediction," *IEEE*, 2016.

[15] S. Asur and B. A. Huberman, "Predicting the Future with Social Media," in *Proceedings of the ACM International Conference on Web Intelligence*, 2010, pp. 492–499.

[16] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[17] L. Zhang and B. Liu, "Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation," *IEEE*, p. 130, 2013.

[18] S. Hochreiter and J. Schmidhuber, "Lstm can solve hard long time lag problems," *Advances in neural information processing systems*, vol. 9, 1996.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[20] M. P. Cristescu, R. A. Neris¸anu, and D. A. Mara, "Using data mining in the sentiment analysis process on the financial market," *Journal of Social and Economic Statistics*, vol. 11, no. 1-2, pp. 36–58,