iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5384

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Stock Market Analysis

Sachin Gavhane[1], Kartik Sharma[2], Akilesh Rao[3], Rushabh Shah[4], Sujay Sail[5]
[1, 2, 3, 4,5]*First-Third Department, First-Third University*

Abstract: Stock market has traditionally been the proving grounds for data mining applications. There is a lack of efficient models which can emulate the heuristic reasoning of humans based on current events/trends. The proposed system is an attempt to reconcile computed sentiments alongside traditional/more common data mining. Datasets consisting of historical data as well as recent headlines will be mined to ascertain stock price movement. This system aims to predict stock price movement more accurately by emulating instinctual reasoning by implementing sentiment analysis.
Keywords: Historical Data, Prediction, Data Mining, Stock Market, Value

I. INTRODUCTION

In a financially volatile market, as the stock market, it is important to have a very precise prediction of a future trend. Because of the financial crisis and scoring profits, it is mandatory to have a secure prediction of the values of the stocks. Stock Market Prediction is the act of trying to determine the future value of a company stock or other financial entity traded on an exchange. The successful prediction of stock's future price could yield significant profit. Quandl Stock values for a range of dates is obtained using Quandl and includes the open, close, high and low values for a given day which will be an important dataset to perform mathematical computation and generate visualizations [6]. Tweets The tweets will be imported using Tweepy into the system and that forms our second dataset that will be used to generate the sentiment of the people towards the company. Sentiment Analysis will be performed on this dataset [6].

II. LITERATURE REVIEW

A. Financial Sentiment Analysis Based on Machine Learning

The aforementioned paper attempted to figure out the best approach to perform sentiment analysis in aid of predicting stock market movement, Naive Bayes and SVM chief among them. The proposed system benefitted from this paper by observing the use of sentiment analysis in a stock market context environment as well as the efficacy of SVM over Naive Bayes as a text classifier. It assisted in the development of the premise of the proposed system as well as a tentative implementation of basic sentiment analysis [1].

B. Stock Prediction using Twitter Sentiment Analysis

The proposed system consists of 2 approaches, that is to say, data mining and sentiment analysis. This particular article/paper helped provide insight into the process of how a sentiment can be mined from textual data. It also assisted the proposed system in learning the kind of vocabulary as is entailed when dealing with the stock market [2].

C. Stock Market Prediction using Data Mining

The proposed system consists of 2 approaches, that is to say, data mining and sentiment analysis. This particular article/paper helped provide insight into the process of how conventional methods of data mining go about predicting stock market movement as well as assisted the proposed system in learning where information extracted using sentiment analysis could be helpful in aid of prediction accuracy to the traditional methods [3].

D. Prediction of Stock Market using Data Mining and Artificial Intelligence

This paper was studied so as to provide a more, well-rounded context as to the techniques related to data mining used in the stock market. It helped the proposed system ascertain sentiment analysis as the better companion to the traditional data mining approach instead of employing an artificial agent. It was noted, however, that the agent was highly effective in shaving off the delays faced by human operators in placing calls [4].

E. Stock Market Prediction using Artificial Neural Networks

As was the case with the previous paper, this paper was studied so as to provide a better-rounded context as to the techniques related to data mining used in the stock market. It helped the proposed system ascertain sentiment analysis as the better companion to the

©IJRASET: All Rights are Reserved

traditional data mining approach instead of employing a neural network in cases which called for the supervised approach. It was noted however that a neural network worked extremely well in situations which called for the unsupervised approach [5].

III. METHODOLOGIES

A. *Datasets*

There are 2 types of data sets included in the proposed system. The data sets used for regression models would be imported from online portals such as Quandl. These data sets have comprehensive list of attributes relating to stock prices and are available for download freely in .csv file formats. The reason why we have chosen this data set is because it provides all the information that is required to predict stock values in a simple csv format which makes it easier to import in the algorithm. The second set of data is for Sentiment Analysis. This data set consists of news headlines/tweets which are imported from online APIs to predict the overall sentiment related to the company stock. Both these data sets are then individually divided into two sets for training and testing purposes.

B. *Algorithm*

On the question of whether to utilize a regression, classification, clustering or association algorithm it is clear that the nature of proposed dataset, association algorithm is out of question right out the gate. Between classification and clustering again it is evident from the supervised nature of the problem, classification is better suited for the problem. That being said, both regression and clustering both have merits to them with the dividing factor being the propensity for continuous values in regression and discrete ones in classification. Given the necessity of classifying any information extracted from the data as one of three classes while also accounting for the continuous nature of the dataset, it is prudent to utilize the regression approach to benefit from the best of both worlds. It is derived from logistic regression but boasts accuracy comparable to those of the more complex algorithms out there.

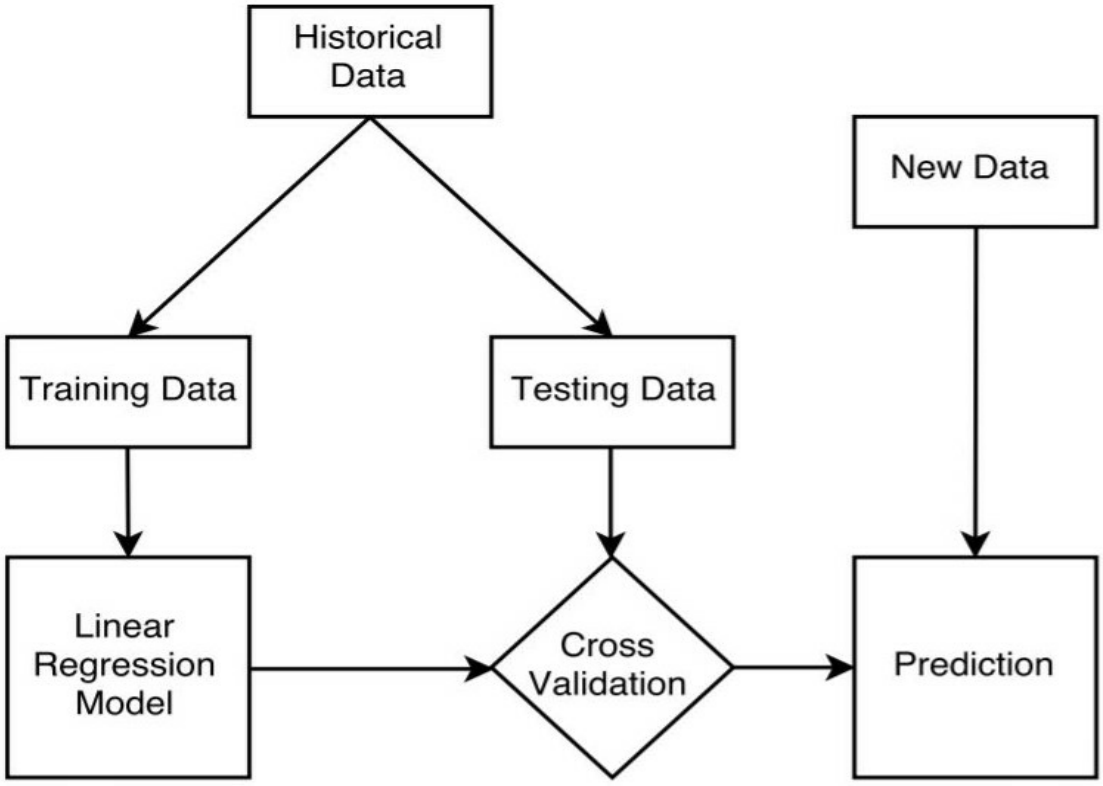


Fig 1.1

IV. RESULT

A. Run the test.py file in PyCharm IDE.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

B. First the Machine Learning part of the model starts running and the dataset of the historic values and predictions based on that appear on the output console as below.

Date	Adj. Close	HL_PCT	PCT_change	Adj. Volume	label
2018-03-01	1071.41	3.720331	-3.436559	2766856.0	1115.04
2018-03-02	1084.14	0.253657	2.472637	2508145.0	1129.38
2018-03-05	1094.76	0.586430	1.542486	1432369.0	1160.84
2018-03-06	1100.90	0.429648	-0.108883	1169068.0	1165.93
2018-03-07	1115.04	0.104032	2.033272	1537429.0	1139.91
2018-03-08	1129.38	0.182401	1.090226	1510478.0	1148.89
2018-03-09	1160.84	0.013783	1.872751	2070174.0	1150.61
2018-03-12	1165.93	1.048948	0.075533	2129297.0	1134.42
2018-03-13	1139.91	3.341492	-2.723945	2129435.0	1100.07
2018-03-14	1148.89	0.946131	0.269681	2033697.0	1095.80
2018-03-15	1150.61	1.033365	0.090469	1623868.0	1094.00
2018-03-16	1134.42	1.973696	-1.811572	2654602.0	1053.15
2018-03-19	1100.07	1.754434	-1.582630	3076349.0	1026.55
2018-03-20	1095.80	0.889761	-0.236708	2709310.0	1054.09
2018-03-21	1094.00	1.343693	0.130884	1990515.0	1006.94

Fig. 9.1(Quandl Result Data Sets)

This dataset is of the stock Google from 1st March, 2018 till 21st March, 2018. The label column depicts the predicted output just from the historical stock prices of Google.

C. After the historical prediction is made, tweets related to Google start loading in on the output console and it shows the polarity of the tweet between values -1 and 1 as below

```
Google is an amazing website and I'm learning a lot from it.        Polarity = 0.600000, Subjectivity = 0.900000

Google ad policy is stupid.                                         Polarity = -0.800000, Subjectivity = 1.000000
```

Fig. 9.2(Textblob sentiment analysis performance)

D. When both of the module processing is done, a single screen plot of both the predictions appear (graphically) as below

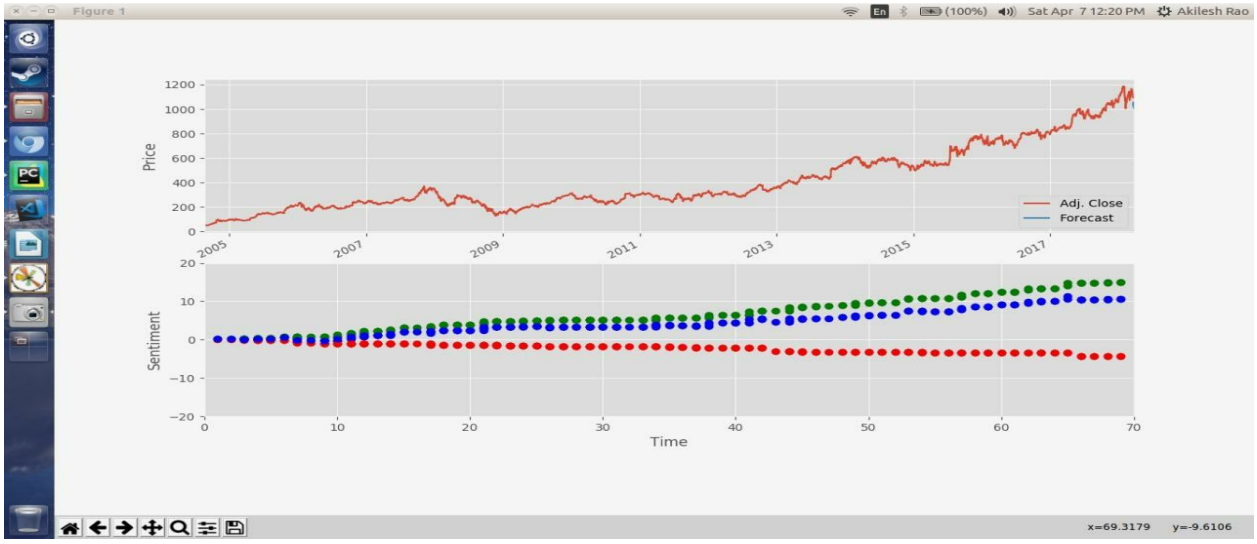


Fig. 9.3(Compound Sentiment Plots)

The first graph shows predictions made on the basis of historical data from Quandl and the second graph shows us the sentiments where blue denotes the overall sentiment about Google, which here, is positive (above 0).

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

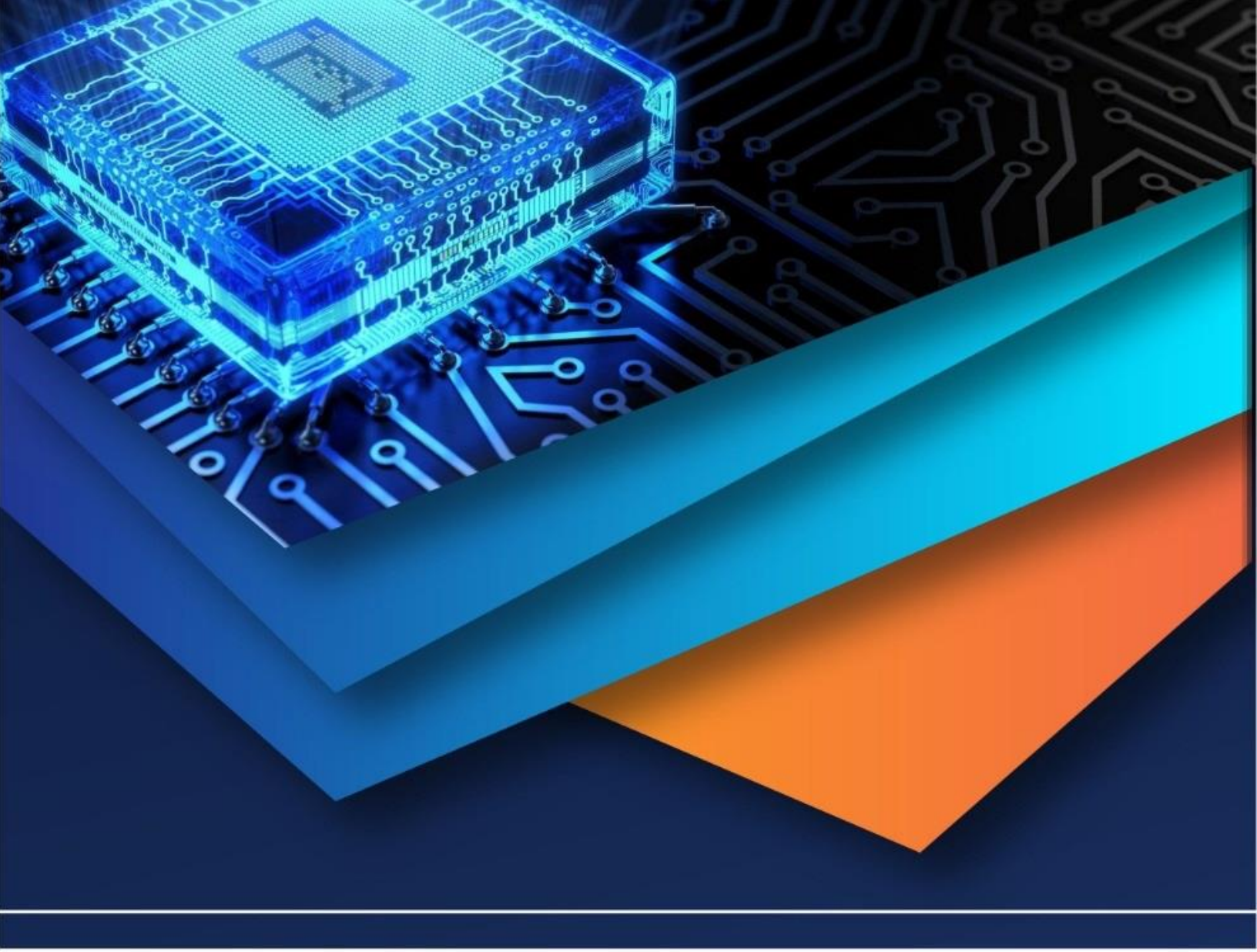
V. CONCLUSION

The proposed algorithm worked on the views/ opinions of their viewers on the shares. Stock Market is such a field where views of the users matter. The views of the experts affect a lot to the traders who want to enter into the market. The unsupervised and supervised learning depend methods help to find the results in a better way. The combinational study is done to get better accuracy. Further, optimizations can be done in sequence to get improved results.

REFERENCES

[1] J. Bollen and H. Mao, "Twitter mood as a stock market predictor". IEEE Computer, 44(10):91–94.
[2] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011
[3] G. P. Gang Leng and T. M. McGinnity, "An on-line algorithm for creating self-organizing fuzzy neural networks." Neural Networks, 17(10):1477–1493
[4] A. Lapedes and R. Farber, "Nonlinear signal processing using neural network: Prediction and system modeling." In Los Alamos National Lab Technical Report
[5] A. E. Stefano Baccianella and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." In LREC. LREC
[6] https://www.quandl.com

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)