International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Sentiment Analysis with Conventional Data Mining for Predicting Share Market Values

Rohan Kapre[1], Rohan Jawale[2], Akshay Dhole[3], Manisha Sonawane[4]

[1, 2, 3, 4]Computer Engineering, Shivajirao S Jondhale COE, Mumbai University, India

Abstract: ***Proposed system aims to predict stock price movement more accurately by emulating instinctual reasoning by implementing sentiment analysis. Stock market has usually been the proving grounds for data mining applications. There is a lack of efficient models which can imitate 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 predict future stock price.***

Keywords: Sentiment Analysis, data mining, Historical Data, Prediction, Stock Market, Stock Value

1. Introduction

Stock market prediction is the act of trying to determine the future value of a company stock or other financial institution. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information. 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. 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. The successful prediction of stock's future price could yield significant profit.

1.1 Quandl

The premier source for financial, economic, and alternative datasets, serving investment professionals. Quandl's platform is used by over 250,000 people, including analysts from the world's top hedge funds, asset managers and investment banks. 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].

1.2 Headlines

The headlines or news will be imported using news APIs 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 [6].

2. Literature Review

2.1 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].

2.2 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].

2.3 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].

2.4 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].

2.5 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 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].

3. Methodologies

3.1 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.

3.2 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.

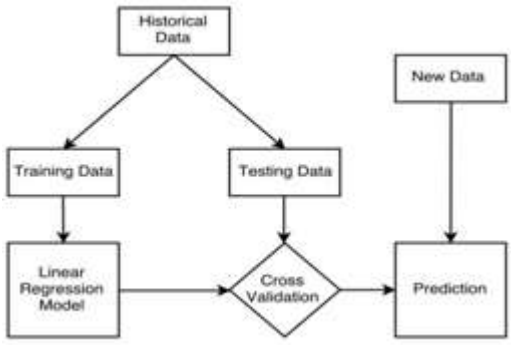
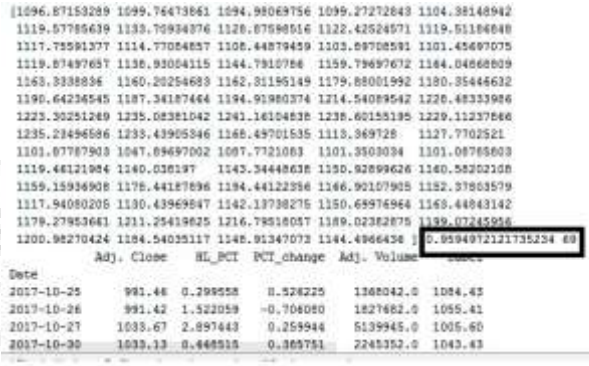


Figure: Flow of Proposed System

3.3 Result

Figure: Conventional Data mining Approach



Date	Adj. Close	HL_PCT	PCT_change	Adj. Volume
2017-10-25	991.46	0.299558	0.528225	1368042.0
2017-10-26	991.42	1.522059	-0.706080	1827682.0
2017-10-27	1033.67	2.897443	0.259944	5139945.0
2017-10-30	1033.13	0.448515	0.365751	2245352.0

Figure: Data mining Approach for Stock market analysis

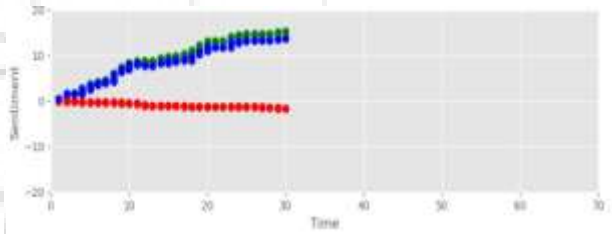


Figure: Sentiment Analysis Approach for predicting stocks

4. 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.

5. Future Scope

Going forward, the system would heavily benefit from fine-tuning the sentiment target to improve accuracy and value of sentiment analysis to the process. In addition, it would be highly beneficial to improve upon the historical dataset's availability to smooth over even Quandl's minor network/key issues to minimize data unavailability even more. Efforts should be made to incorporate alternative dataset sources to test out their feasibility, for example, Google/Yahoo Finance News, instead of Twitter.

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
[3] Intelligent Systems and Technology, 2:27:1–27:27, 2011.
[4] 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.
[5] Lapedes and R. Farber, "Nonlinear signal processing using neural network: Prediction and system modeling." In Los Alamos National Lab Technical Report.
[6] E. Stefano Baccianella and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." In LREC. LREC.
[7] https://www.quandl.com