**INTERNATIONAL JOURNAL OF RESEARCH –
GRANTHAALAYAH**

A knowledge Repository

RACSIT - 17

SURVEY ON SENTIMENT ANALYSIS OF STOCK MARKET

Nausheen S [*1], **Anil Kumar M** [*2], **Amrutha K K** [3]
[*1, *2, 3] Computer Science Engineering, SET, Jain University, Bengaluru, India

Abstract

Sentiment analysis has seen a tremendous growth in the past few years. Sentiment analysis or opinion mining is a process of collecting users' opinion from user generated content. It has various applications, such as stock market prediction, products' review collection, etc. a large amount of work has been done in this field by applying sentiment analysis to various applications. The main goal of this paper is to study the various methods used for sentiment analysis. Further we explain the overview of various related papers and their performances.

Keywords: Support Vector Machine; NaiveBayes; K-Nearest Neighbour.

1. Introduction

With the development of technology, social media is increasingly used by people to share their views, consult for reviews, etc. This information can be used for many purposes, one of them being opinion mining. Sentiment analysis refers to identifying whether the given piece of data is positive or negative. A basic task in sentiment analysis is classifying the polarity of a given document-whether the opinion is positive or negative. Advanced polarity classification looks at emotion states such as angry, sad, happy, etc. [1]. The older methods for collecting sentiments were both tedious and less accurate. Thus, views shared by people on social media are much more accurate than that collected from questioners, which is usually filled with reluctance and without personal interest [2]. Also, an automated system is much easier to analyse than manual survey. The various opinions shared on social media can influence the buying patterns of customers [3]. It can also be used by businesses to improve their products [4]. Various methods and algorithms can be used to perform sentiment analysis based on application and dataset involved. [5] Uses social media services namely twitter to predict future stock prices. Here, they used machine learning algorithm to classify data and estimate future stock prices, and the reduced programming model was used for calculation. On the other hand,[6] uses a lexicon based approach for sentiment analysis of news comments. Here the polarity is provided using

[Nausheen et. al., Vol.5 (Iss.4: RACSIT), April, 2017]   ISSN- 2350-0530(O), ISSN- 2394-3629(P)
ICV (Index Copernicus Value) 2015: 71.21   IF: 4.321 (CosmosImpactFactor), 2.532 (I2OR)
Recent Advances in Computer Science and Information Technology   InfoBase Index IBI Factor 3.86

lexicon based approach and, these results are then fed to machine learning algorithms, namely, SVM and K-nearest neighbour. Sentiment analysis belongs to the domain of opinion mining, and hence is also referred to as opinion mining. A number of terms are used in sentiment analysis as defined by Pang and Lee [7].

Sometimes, the term sentiment analysis can also be used as it includes natural language processing. Polarity is a term that defines whether a terms or sentiment is positive negative or neutral. Subjectivity includes classifying a given text as subjective or objective. Sentences indicating facts are objective, while sentences with sentiments are usually subjective in nature. For example, "Suppose he did lie beside Lenin, would it be permanent?" is a subjective statement. Sentiments can be explicitly stated in the sentence, for example, "the movie was fantastic", can be implicit in nature.

Sentiment analysis of stock market helps people to make informed decisions, whether to invest in a business. Stock analysis refers to analysing the trade of an enterprise or a company. Analysis shows that, online sentiment can help to predict subsequent market activity. Positive sentiments increase stock value of a company while negative remark decreases it. Stock price depends on new information significantly. The many information sources are people's opinion in social media, news, articles etc.
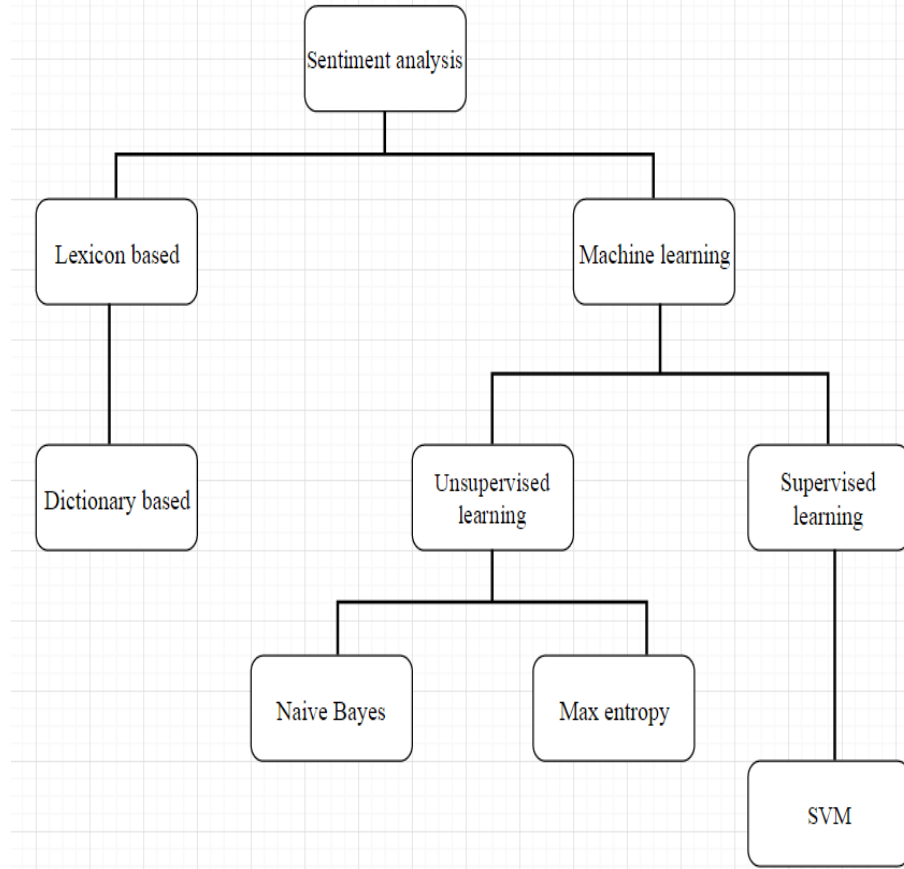
2. Challenges and Applications

Sentiment analysis or opinion mining can have various applications such as movie reviewing, stock market prediction, product feature reviews etc.
The various challenges [8] are:

- **Finding the correct dictionary:** it is difficult to find the most accurate dictionary that contains all required words, and usually we can overcome this problem by creating our own dictionary based on the requirement.
- **Detection of sarcasm in statements:** it is difficult to detect sarcasm and provide it a suitable polarity value.
- **Detection of fake reviews:** internet contains spam content as well. Effective Sentiment classification requires this spam content to be eliminated before processing. This can be done by identifying duplicates, by detecting outliers. [10]
- **Use of orthographic words:** orthographic words such as too, veeery, etc. are difficult to polarize.
- **Use of abbreviations:** short forms such as u for; you', 4 for 'for' are difficult to provide polarity for.

3. Methodology

Sentiment analysis can be carried out at three levels, namely, document level, sentiment level, and aspect level [11]. Sentiment analysis includes three main steps, identification, classification, and aggregation [12].

[Nausheen et. al., Vol.5 (Iss.4: RACSIT), April, 2017]   ISSN- 2350-0530(O), ISSN- 2394-3629(P)
ICV (Index Copernicus Value) 2015: 71.21   IF: 4.321 (CosmosImpactFactor), 2.532 (I2OR)
Recent Advances in Computer Science and Information Technology   InfoBase Index IBI Factor 3.86

Dictionary Based

A sentiment dictionary can be used to identify polarity of given word in a text. In this method, a predefined dictionary is first created manually using a set of seed words [9]. The polarity of text can then be assigned as positive, negative, or neutral based on these predefined dictionaries. This method uses word count, frequency of occurrence and other methods to provide polarity to given data [13]. The use of this approach for sentiment analysis can be explained as follows [14]: seed words with predefined polarity values are collected manually. An algorithm is then applied, which searches dictionaries such as wordnet to find more words of similar nature. These new words can then be added to the list and process can be iterated till no new words are found.

Supervised Learning

Supervised learning provides polarity to new data based on a training dataset. The training data consists of input data and output variables. One common method under supervised machine learning is using support vector machine (SVM). The support vector machine contains an algorithm that recognizes patterns from the given data and groups similar group members using the concept of decision plane [15]. SVM can give an accuracy up to80% with the correct dataset [16].

[Nausheen et. al., Vol.5 (Iss.4: RACSIT), April, 2017]   ISSN- 2350-0530(O), ISSN- 2394-3629(P)
ICV (Index Copernicus Value) 2015: 71.21   IF: 4.321 (CosmosImpactFactor), 2.532 (I2OR)
Recent Advances in Computer Science and Information Technology   InfoBase Index IBI Factor 3.86

4. Results and Discussions

Paper	Summary	Performance
Sentiment Analysis on News Articles for Stocks [17]	This paper analyses the sentiments of data collected from news articles. To get the news links, the Bing API was used and a sentiment dictionary was then used to analyse the articles. A comparison between two different machine learning algorithms was made, and also, the predicted results were compared with the actual changes of stock prices in the market. Mapping the analysis to immediate changes in the market was discovered as scope for future improvement.	Accuracy for normal equation- 53.2%, and for gradient decent-59.5%.
Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets [18]	This paper deals with the feature sparse problem resulting from sentiment analysis using tweets. In order to overcome this, a model called text-sentiment-based stock trend prediction model was used. This model uses an SVM classifier model. To improve accuracy, the hybrid feature selection method was added to the model. Here a SentiWordNet is used to add weightage to selected features. This paper compares many learning algorithms, concluding that SVM is the most accurate one.	An accuracy of 90.34%.
Twitter mood predicts the stock market [19]	This paper is used to check whether or not the moods of people correlate to the Dow Jones Industrial Average (DJIA) value. For doing so, two tracking tools are used, namely, opinion finder that measures positive versus negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) for measuring the response to presidential election and Thanksgiving. This paper also throws light on the hypothesis that peoples' moods measured using opinion finder and GPOMS predict the DJIA closing values.	An accuracy of 87.6% was obtained and Mean Average Error was reduced by more than 6%.
Collective Sentiment Mining of Micro blogs in 24-hour Stock Price Movement Prediction [20]	This paper is used for collective sentiment analysis to predict and analyse the stock price change for the next day. it includes the use of a two-stage process which uses NLP approach and a statistical analysis approach. It is based on the SVM approach. It uses the knowledge of machine learning and natural language processing. The result of prediction is tested using Granger Causality test.	An accuracy of 71.84% for positive sentiments and 74.3% for negative

[Nausheen et. al., Vol.5 (Iss.4: RACSIT), April, 2017]   ISSN- 2350-0530(O), ISSN- 2394-3629(P)
ICV (Index Copernicus Value) 2015: 71.21   IF: 4.321 (CosmosImpactFactor), 2.532 (I2OR)
Recent Advances in Computer Science and Information Technology   InfoBase Index IBI Factor 3.86

Stock Price Prediction using Linear Regression based on Sentiment Analysis [21]	This paper stresses on the fact that according to efficient market hypothesis (EMH) stock prices depend on a number of factors, one of them being peoples' opinion or sentiment. This paper surveys the Indonesian stock market using sentiment analysis. Two algorithms are used for analysing the tweets, namely, Naïve Bayes and Random Forest algorithms. As future work, the authors suggest increasing the data collection period, using non-linear prediction models and including features such as POS tagging, word weighting, etc. in classification.	The Random model algorithm yields an accuracy of 60.39% and the Naïve Bayes algorithm gives 56.50% accuracy.
Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis [22]	This paper tests the assumption of increasing accuracy of stock market prediction by analysing the psychological moods of twitter users. Eight different emotions can be analyzed by using lexicon-based approach to classify peoples' psychological states. Support vector machine algorithms are used to predict the DJIA indicators. Focusing on the increase of training period and efficiency of algorithms was saved as future work.	The accuracy rate of 64.10% was achieved using Support Vector Machine algorithm to predict DJIA indicator.
A Hybrid Approach to Sentiment Analysis of News Comments [23]	This research uses two main algorithms, that is, Support Vector Machine (SVM) and K-Nearest Neighbour (KNN), to perform sentiment analysis of news comments. For this, a hybrid approach is adopted, wherein, a sentiment lexicon is used to classify the comments and calculate polarity. The results of this are then used to train the machine learning algorithms.	The accuracy of 60.96 for SVM
Prediction Relying on Text mining and Sentiment Analysis with Tweets [24]	This paper focuses mainly on the twitter feature sparse problem and also the unreliability of using average sentiment score. This is overcome using a hybrid approach consisting of sentiwordnet to give additional weightage. SVM cross validation is used.	Accuracy without Feature selection (FS) was 53.62%, whereas with FS, an accuracy of up to 84.06% was obtained.
Stock Trend Forecasting Method Based on Sentiment Analysis and System Similarity Model [25]	This paper proposes a system wherein the Bayesian classifier is used based on the system similar model to predict stock movement. The system is tested using inter cross and turnover test.	An accuracy of up to 78.5% is obtained

Stock Market Forecast Using Sentiment Analysis [26]	In this paper they have used twitter and BSE to calculate public moods and used granger causality to predict the results, they also implemented SOFNN (Self organised fuzzy neural network) for decoding non-linear time series.	An accuracy of 75.56% was obtained

Acknowledgement

References

[1] Manisha Shinde-Pawar,"Formation of Smart Sentiment Analysis Technique for Big Data",International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 12, December 2014

[2] Kim Schouten and Flavius Frasincar, "Survey on aspect-level sentiment analysis", IEEE Transactions on Knowledge and Data Engineering, Volume: 28, Issue: 3, March 1 2016

[3] E.Van Kleef, H. C. M. Van Trijp, and P. Luning, "Consumer research in the early stages of new product development: a critical review of methods and techniques," food quality preference, Vol. 16, no. 3, pp. 181–201, 2005.

[4] Y. Chen and J. Xie, "Online Consumer review: word-of-mouth as a new element of Marketing Communication mix," manage. sci., Vol. 54, no. 3, pp. 477–491, 2008.

[5] Michał Skuza, Andrzej Romanowski, "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction", Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1349–1354 ACSIS, Vol. 5.

[6] Addlight Mukwazvure, K.P, Supreethi, Addlight Mukwazvure, K.P Supreethi," A Hybrid Approach to Sentiment Analysis of News Comments".

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," found. trends inf. retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[8] Asmita Dhokrat, Sunil Khillare, C. Namrata Mahender, "Review on techniques and tools used for opinion mining", international journal of computer applications technology and research volume 4– issue 6, 419 - 424, 2015

[9] Laura Cruz, Jose Ochoa, Mathieu Roche, Pascal Poncelet, "Dictionary-based Sentiment Analysis applied to specific domain using a Web Mining Approach"

[10] Indurkhya, N. & Damerau, F.J. [eds.] Liu, B., 2010 "Sentiment Analysis and subjectivity," appeared in handbook of natural language processing.

[11] Bing l., 2012 "Sentiment analysis and opinion mining," Morgan & Claypool publishers.

[12] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," data mining know. discovery, vol. 24, no. 3, pp. 478–514, 2012.

[13] Douglas R, Rice Christopher, "Corpus-based dictionaries for Sentiment Analysis of specialized Vocabularies"version 0.1 September 19, 2013

[14] Reshma Bhonde, Binita Bhagwat , Sayali Ingulkar, Apeksha Pande, "Sentiment Analysis based on dictionary approach", International Journal of Emerging Engineering Research and Technology Volume 3, issue 1, January 2015

[15] Pravesh Kumar Singh, Mohd Shahid Husain," Methodological Study of Opinion Mining and Sentiment Analysis Techniques", International Journal on Soft Computing (ijsc) vol. 5, no. 1, February 2014

[16] Phayung Meesad and Jiajia Li, King Mongkut's, "Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets"

[17] VaanchithaKalyanaraman, Sarah Kazi, Rohan Tondulkar,"Sentiment Analysis on News Articles for Stocks".

[18] Phayung Meesad and Jiajia Li," Stock Trend Prediction Relying on Text Mining and Sentiment Analysis withTweets"

[19] Johan Bollen,Huina Mao,Xiao-Jun Zeng," Twitter mood predicts the stock market"

[20] Feifei Xu and Vlado Keˇselj,"Collective Sentiment Mining of Microblogs in 24-hour Stock Price MovementPrediction", 2014 IEEE 16th Conference on Business Informatics

[21] Yahya Eru Cakra,Bayu Distiawan Trisedya," Stock price prediction using linear regression based on sentiment analysis", Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference.

[22] Alexander Porshnev, Ilya Redkin, Alexey Shevchenko,"Machine learning in prediction of stock market indicators based on historical dataand data from Twitter sentiment analysis", 2013 IEEE 13th International Conference on Data Mining Workshops

[23] Addlight Mukwazvure,K.P Supreethi,"A Hybrid Approach to Sentiment Analysis of NewsComments"

[24] Phayung Meesad, Jiajia Li, "Stock trend prediction relying on text mining and sentiment analysis with tweets",2014 4th World Congress on Information and Communication Technologies (WICT 2014)

[25] Kaihui Zhang, Lei Li, Peng Li,"Stock trend forecasting method based on sentiment analysis and system similarity model",Strategic Technology (IFOST), 2011 6th International Forum.

[26] Rajat Ahuja, Harshil Rastogi,Arpita Choudhuri," Stock market forecast using sentiment analysis", Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference

*Corresponding author.
E-mail address: nausheensalauddin@gmail.com