

POLITICAL PREDICTION ANALYSIS USING TEXT MINING

Dinesh Gaikwad, Snehal Kolhe, Vishwambhar Deshpande, Pratiksha Indrol



dineahg@gmail.com
Snehalkolhe10@gmail.com
vshdeshpande@gmail.com
pratikshaindrol@gmail.com

Computer Department, AISSMS IOIT, Pune

ABSTRACT

Nowadays, with a growth of social media it has become a part of every man's life so it can be used in analysis & prediction task. People share their feelings, opinion and viewpoints on the social media. Among that election is conducted to view the public opinion, where group of people choose the candidate by using votes, many methods are used to predict result. We use twitter data to predict outcome of election by collecting twitter data and analyze it to predict the outcome of the election by analyzing sentiment of twitter data about the candidates.

Keywords— Sentiment Analysis on text mining, Naïve Bayes Classifier, Machine Learning Algorithm, Natural Language Processing

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

I. INTRODUCTION

Election is most important part in terms of democracy because it is the bridge between the local citizens and ruling government. Opinion poll was introduced since the early 19th century but sometimes these polls fail to predict election results accurately. So there is an alternative way to predict election outcomes i.e. Web data. Web data provides enormous amount of information about any kind of topic which includes politics as well. In our proposed system we have used social media as a dataset because it provides data which includes politics as well.

In our proposed system we have used social media as a dataset because it provides data which is up to date to predict result with the help of user made posts about criticize, praise or stand about political party or the respective candidate. This system proposes a method to predict the winning of political party or candidate by applying sentiment analysis on textual data, deep learning algorithm and Vec2word model.

Data is collected from Twitter REST API and for that we applied for the open access under study purpose to Twitter Developer Team. Dataset imported from Twitter REST API which is for textual data and it is location based. This data is used to determine the live sentiment and we have generated graph based prediction and also we used NLP (Natural Language Processing) to segregating the views of

people's postings on social media platforms. Use of deep learning algorithm is for storing historic data like previous election data to determine future results more accurately. The aim of this system is to analyze the trends in election using location as the filter. Data collection is done on location based. By using this system Political Campaigns can be set up for improvising the chances of any political party to win and it can also be used in places like Stock Market, E-commerce websites song recommendations for better predictions and generating recommendations.

II. LITERATURE SURVEY

Prediction of Election Using SVM Support Vector Machine (SVM) is used for classification for assigning a discrete class and for clustering which separates data points to different classes.[1] The main disadvantage of SVM in terms of Prediction is that it applies for labelled data only which might sometimes is in the form of unlabelled data so it may rise conflict between the proposed systems using SVM algorithm makes the training model complicated.[2]

Model for forecasting Electoral College This model was introduced by Berry and Bickers in 2012 which did not require pre-election polls on voting intentions. however this paper also contains out-of sample errors that are lower than the within-sample errors that was reported in

manuscripts by the developers. [3] The drawback of this model is that it does not use any previous historic data to predict the result which may reduce the chance of accurate prediction.[4]

Lexicon Based Sentiment Analyser lexicon based sentiment analyser classifies the tweets based on the sentiment value. This classification is done by polarity and different measures i.e. positive, negative and neutral. The main drawback of using this system is that it cannot analyze different kinds of data and it does not support multi lingual sentiment.[5]

Machine Learning algorithm for mining text Machine Learning algorithm require massive data sets to train and that data should be unbiased and should be of good quality. If the dataset contains biased and low quality data then that leads to high error susceptibility, data acquisition that misleads the train model using machine learning algorithms may sometime occur error in predictions as dataset pairs are unlabelled and biased. so that system fails to predict accurate result. So collectively use of deep learning algorithm along with the sentiment analysis on live Twitter dataset improves the accuracy of the prediction of the result.[6]

Sentiment Analysis Using LEXIPERS. The Proposed system using lexipers were not successful in predicting result accurately due to cost of providing annotated data and this method is totally dependent on structure of language and need modification. As a result the polarity of result cannot be perfectly obtained using the lexipers.[7]

Prediction using Aggregation Generally aggregation is used to aggregate sentiments of tweets to decide the winning. But the problem comes with both statistical and computational efficiency. while using the aggregation in prediction it has been seen the loss of data and that produced inaccurate predictions.[8]

Election Prediction Based on Naive Stochastic Approach the basic idea about using this method is to pick up a vector and simulate the voting by rotating the wheel and with the preferences. But the Drawback of using this method is that this algorithm needs repetition of the values or pairs passed to it. instead of collecting the complete dataset and training the data.[9]

III. PROPOSED SYSTEM

In our proposed system first of all we take the dataset from twitter through twitter API access.

Then loading data into database using sqlite server in the form of text data & tabular data. After that data preprocessing carried out on that text & tabular data. In data preprocessing hashtags, punctuation marks, white space will be removed. Once data preprocessing done then data classification Will be carried out on preprocessing data naive bayes classification. In Simple term, a naive Bayes classifier assumes that the presence of a particular feature of a class is unrelated to the presence of any other features given in class variable. And here we use data from Twitter data set.

Here analysis use of Vader sentiment analysis. To classify either word is positive or negative. Vader not only tells

about the Positivity and negativity score but also tells us about how positive or negative a sentiment is.

Natural Language Processing(NLP) And Naive Bayes algorithm also used for sentiment analysis.

And we get political Prediction in the form of Graph and Piechart.

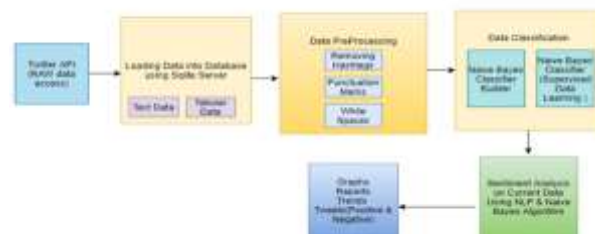


Fig 1: Architecture of Prediction using sentiment analysis

IV. APPLICATIONS

1. To know public opinions for political leaders and their Activities in terms of Development.
2. In business and Government intelligence for knowing customer attitudes and trends in market.
3. Detection of Insensitive data on social media platforms like Facebook, Twitter, Instagram Etc.
4. Resolving Customer Experiences for growing sales and profit.
5. For Analyzing Social Media return of investment on social media marketing

V. RESULTS & OUTCOME



Fig 2. Main Application Screen



Fig 3. Positive & Negative Tweet Analysis



Fig 4. Scatter Plot on Live Data



Fig 5. After Sentiment Analysis & Classification

VII. CONCLUSION

Our system generates location based prediction in the form of statistical data. the system uses Naive Bayes classifier, Natural Language Processing and Sentiment analysis on Live data so it Improves Accuracy in Prediction Analysis and Reduces Loss of Data. use of unsupervised learning algorithm reduces model training efforts as well.

REFERENCES

- [1] Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, Gurpreet Kaur, Opinion Mining And Sentiment Analysis, IEEE International Conference On Computing For Sustainable Global Development, October 2016.
- [2] Y. Yang and F. Zhou, "Microblog Sentiment Analysis Algorithm Research and Implementation Based on Classification", 2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), 2015.
- [3] Hutto, C.J. Gilbert, E.E. (2014). VADER: A Parsimonious Rule based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- [4] Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013
- [5] S. Mukherjee and P. Bhattacharyya, "Sentiment analysis in twitter with lightweight discourse analysis,"

Proceedings of the 24th International Conference on Computational Linguistics (COLING), pp. 1847–1864, Dec. 2012.

[6] A. Bakliwal, P. Arora, A. Patil, and V. Varma, "Towards enhanced opinion classification using NLP techniques," Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP, pp. 101–107, Nov. 2011.

[7] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford University, pp. 1–12, 2009.

[8] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp 115–120, July 2012

[9] Q. Wu, S. Tan, H. Zhai, G. Zhang, M. Duan, and X. Cheng, "Sent iRank: CrossDomain Graph Ranking for Sentiment Classification," presented at the Proceedings of the 2009IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, 2009

[10] K. Mao, J. Niu, X. Wang, L. Wang and M. Qiu, "Cross-Domain Sentiment Analysis of Product Reviews by Combining Lexicon-Based and Learn-Based Techniques", 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, 2015.

[11] Alexander Pak and Patrick Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining". In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), may 2010.

[12] Pritiee Salunkhe and Sachin Deshmukh. "Twitter Based Election Prediction and Analysis", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 10 October 2017.