International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

Identification on Top Trends of Public Opinion using Location based Sentiment Analysis

Vishal C, K. Saravanan

Abstract: ***Social Media has become a platform for the users to express their opinion on the new emerging trends. It has been estimated that 80% of the data in today's world is unstructured and not been organized in a per-determined order. For the user It has become very hard to group the data because of time consumption and analyzing. Users from different parts of the world share their opinions based upon the emerging trends in social media. Such data are classified and categorization into polarities like positive, negative and neutral process using sentiment analysis and processed for the data accuracy. A model can be built by locating user's Geo location to know the connectivity and location of the users who are interested in top trending events happening in and around the world.***

Index Terms: ***Sentiment Analysis, Classification, social Networking, location, accuracy***

I. INTRODUCTION

Unstructured data or information about people expressing their opinions about the trending events happening around the world has been shared on the social media like twitter, Facebook, Instagram etc. A huge flood of information of social media data has been increasing day by day. It has become very hard to group the data based upon the user's locality and their opinions about the top trends. On the other side categorizing the data into polarities like positive, negative and neutral process for the best match of data accuracy. Different patterns of data from different location or region have been extracted and processed. A unique information from social media has become a foundation to deal with new sets of information. The data can be categorized into supervised algorithm techniques. The data will be trained and tested based upon the naves Bayes classification overdeveloping predictive model involves a supervised technique known as classification. A model can be built by locating new locations on the new trends among the users from the Twitter data and then categorize into the positive, negative and neutral process. This method removes duplicate attributes and leads to better classification for data accuracy. The classification of text is simplified and coupled using a bag-of-words model. Emerging technology can change the new dimension of the area today. The unstructured data will be processed and evaluated for the analysis. By the opinion of the user on the top trends based upon the location using sentiment analysis can be determined. Automatic Machined based algorithm is used for the text processing and classified.

Manuscript published on 30 March 2019.
***Correspondence Author(s)
Mr.Vishal C, Research Scholar, Dept. of Computer Science, PRIST University, Thanjavur, Tamilnadu, India.
Dr.K. Saravanan, Dean, Faculty of Computer Science, PRIST University, Thanjavur, Tamilnadu, India

The different levels of sentiment analysis
Document Level: Analysis of complete document or paragraph
Sentence Level: Analysis of single sentence
Sub-sentence Level: Analysis of sub expression within a sentence

II. RELATED WORK

Dimitrios Michailidis [1] Discussed about the real time classification of data based upon the customer location. Tweets regarding the customer from different location based on the air sent in united states of America and various streams of real time Twitter data have been classified into emotional contents like positive, negative and mixed terms.

Yanping Lv [2] According to the author social networking data where the users play a vital role in communication based upon the geolocation. Based on these trending events an visual analysis of the connection between network space and real physical space has been identified.

Gargi Mishra [3] A high percentage of tweets from the different location is considered by opinion mining of the users based on business, politics and media, tweets. These factors are determined by the user's current location and classified into positive, negative and neutral process.

Omaima Almatrafi [4] Users based on their location presents the different opinions for Indian general elections 2014 and the author performs data mining of different users on different location. Finally, the results of the user's opinion to the different parties of Indian politicians from different locations based upon the twitter data is categorized.

Mondher Bouazizi [5] The text classification refers to the detection of weights proposed to an approach based upon sentiment analysis. Depending on the weights, the text is based into positive, negative and mixed terms and the data accuracy reaches 81%. In the next step, sentiment quantification will perform of the tweets by defining five positive and five negative sub-classes of the existing tweets. A meaningful task and multi-class classification are performed.

III. PROPOSED METHODOLOGY

Open Authentication framework is a standard form of application which is used to gain credentials for your information without getting login information to some websites as third-party authentication provides access to protected information.

When a user registers to the twitter to access application interface, a consumer account and a secret key are provided. using the credentials as user can access the information of the twitter data.

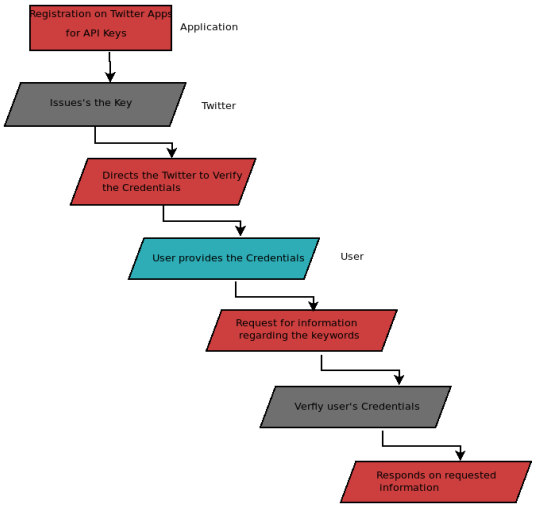

Fig 1: Twitter Application OAuth Authentication

Fig 1 shows that the twitter data is processed by the OAuth authentication. Application of OAuth is a form of authentication where a user requests the application to download the twitter data.

IV. EXPERIMENTAL SETUP

The first process is to collect data about tweets from the internet. Twitter data is generally available, and it can be collected through scraping the website or via the Twitter API. The twitter data can be collected through the credentials Keys which is provided by twitter API. Then we need to construct the outcome variable for these tweets, which means that we must label them as positive, negative, or neutral sentiment.

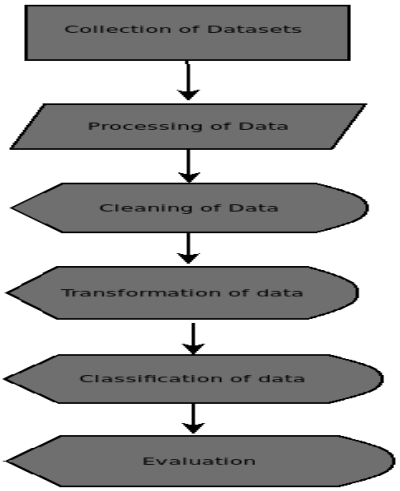

Fig 2: Process of Twitter Data sets for classification

The **figure2** explains about the complete process after the collection of datasets from twitter application. Next step to how to processes the data from twitter in RStudio is shown below

tweets=searchTwitter('narendramodi', geocode='20.593684,78.96288,1000km',n=500, lang='en')

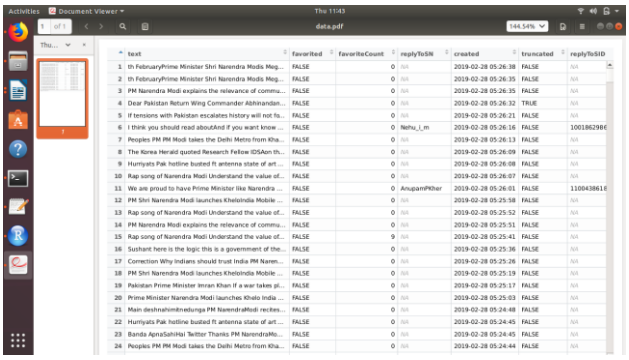

Fig 3. Collection of Tweets from Twitter Application

Fig 3 shows the Collection of the recent tweets related to our beloved prime minister regarding the likes and dislikes of people opinion by their geo location from the twitter.

A. Cleaning the Data Stop wards are more censorious in many of the application is very important to remove the repeated words which are commonly used in the document. For example, words like the", "a", "an", "in" are ignored, but when indexed for query its results in search query clean up the data for removing numbers, space, special characters etc.in this application **'gsub'** function is used.

B. Bag-of-Words (BoW)

The number of times tokens appears in the document is a collection of words. Bag of words is the process of collection of words I a document.

Number of Columns = number of distinct of tokens in the whole collection of documents

Number of rows = number of documents in the whole collection of documents.

C. Matrix representation of Bag of Words: The Document Term Matrix

Document term matrix (DTM) is the collection of positive and negative terms. The rows represent the sentences in the document in the collections and columns represent the main terms of the documents and it is known as term frequencies. 'tm' built in function package to create DTM

df.train <- df[1:1500,] df.test <- df[1501:2000]
dtm.train <- dtm[1:1500,] dtm.test<- dtm[1501:2000]

Docs	strike	transformation	attack	news	piece	angry
31	0	0	0	1	0	1
32	0	0	1	0	0	0

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

33	0	0	0	0	0	0
34	0	0	0	0	0	0
35	1	0	0	1	0	0
36	0	0	0	0	0	0
37	0	0	0	0	0	1
38	0	0	1	0	0	0
39	0	0	0	0	1	0
40	0	0	0	0	0	0

Table1: Document Term matrix of main terms calculation

The document tree matrix of main terms of positive and negative terms is shown in **table 1**.

To train the model we use the naive Bayes function from the 'e1071' package. Since Naive Bayes evaluates products of probabilities, we need some way of assigning non-zero probabilities to words which do not occur in the sample.

table ("Predictions"= pred, "Actual" = df.test$class)

V. RESULT ANALYSIS

The data after the process are categorized into positive and negative categorist table shown below how's the number of tweets process and categorized. The **table 2** shown below shows the actual and predicted terms of positive and negative frequency with data accuracy.

		Actuals		
		Neg	**Pos**	**Accuracy**
Predictions	**pos**	**224**	**54**	**0.81**
	Neg	**41**	**181**	

Table 2: The Actual and Predicted frequency terms with data accuracy.

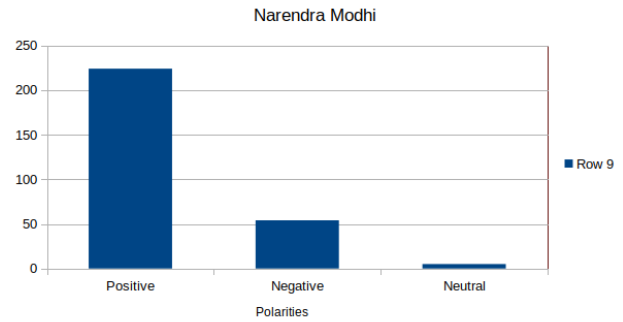

Fig 4: Classification of tweets into positive, negative and neutral

The tweets are classified into positive, negative and neutral polarities is shown in the figure 4.

D. Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known**. Table 3** shows the values preferred for sensitivity, specificity and positive values

Sensitivity	Specificity	Pos Pred Value
0.8452830	0.7702128	0.8057554

Table 3: Confusion Matrix table of sensitivity, specificity and positive preferred value

The prediction accuracy of a classification model is given by the proportion of the total number of correct predictions.

The accuracy for this model turns out to be **81%.**

Ge-location of the tweets the location of the user can be identified by providing the longitude and latitude. User opinion about the top trends has been displayed over the map.

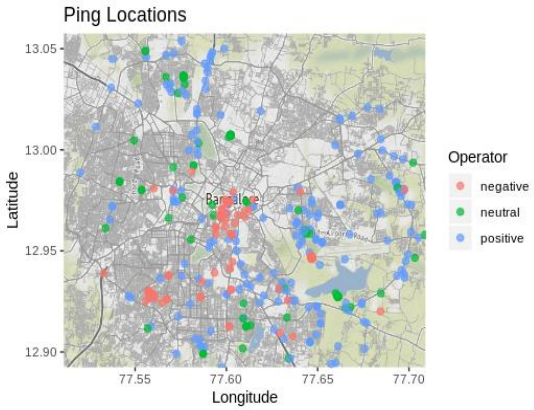

Fig5: Geo-location of users of Bangalore region

Figure 5 shows the geo location of the user from the Bangalore region by providing the longitude and attitude

VI. CONCLUSION

The information from the twitter data results in attractive source of information for opinion and sentiment analysis. Location based sentiment analysis over 10000 twitter data gives different pattern of information to the user about the top trends. The prediction of tweets over the top trends helps the user to know the opinion of other users' human behaviour. Hence, our classifier will make use of classification techniques, aspect extraction and supervised machine learning algorithms. By classifying the data based upon the user location, it was found that the accuracy of classifier data on different region can be determined.

REFERENCES

1. Dimitrios Michailidisi, Tomoaki Ohtsuki, "Real Time Location Based Sentiment Analysis on Twitter The AirSent SystSETN '18, July 9–15, 2018, Rio Patras, Greeceem
2. Yanping Lv , Xiao Xiao , Dazhen Lin , Donglin Cao ,"Public Opinion Analysis Based on Geographical Location", 2015 8th International Congress on Image and Signal Processing (CISP 2015)

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

3. Gargi Mishra,Shivani Varshney," Location Based Opinion Mining of Real Time Twitter Data", Gargi Mishra et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (4) , 2016, 1831-1835
4. Omaima Almatrafi, Suhem Parack, Bravim Chavan, "Application of Location-Based Sentiment Analysis Using Twitter for Identifying Trends Towards Indian General Elections 2014", IMCOM '15 Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication
5. Mondher Bouazizi, Tomoaki Ohtsuki, "Sentiment Analysis in Twitter: From Classification to Quantification of Sentiments within Tweets", IEEE Global Communications Conference (GLOBECOM), 2016.
6. Jyoti Jain, Archana Shinde, Prachi Panchal, Nihar Suryawanshi, "Sentiment Analysis using Machine Learning", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, ISSN: 2277 128X, January 2016, pp 451-453.
7. Neethu M S, Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT),2013.\
8. Ankit Pradeep Patel, Ankit Vithalbhai Patel, Sanjaykumar Ghanshyam Bhai Butani, Prashant B. Sawant, "Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches", International Journal for Innovative Research in Science & Technology, Volume 3, Issue 10, ISSN (online): 2349-6010, March 2017,pp 19-21.
9. John Ross Quinlan. C4. 5: programs for machine learning, volume 1. Morgan Kaufmann, 1993.
10. Balakrishnan Gokul Krishnan, P Priyantha, T Raghavan, N Prasanth, and A Perera. Opinion mining and sentiment analysis on a twitter data stream. In Advances in ICT for Emerging Regions (ICTer), 2012 International Conference on. IEEE, 2012
11. Martin F Porter. An algorithm for suffix stripping. Program: electronic library and information systems, 40(3): pages 211-218, 2006.
12. Hassan Saif, Yulan He, and Harith Alani. Semantic sentiment analysis of Twitter. In the Semantic Web-ISWC 2012, pages 508-524. Springer, 2012.
13. Isaac G Council, Ryan McDonald, and Leonid Velikovich. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In Proceedings of the workshop on negation and speculation in natural language processing, pages 51-59. Association for Computational Linguistics, 2010.
14. Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! ICWSM, 11: pages 538-541, 2011.
15. Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. volume 2010, pages 1320-1326, 2010.
16. Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments.

Dr. K. Saravanan received his M.Sc. in Computer Science from A. V. C. College (Autonomous), Mayiladuthurai in 1992, M.S. in Software Systems from B.I.T.S. Pilani in 1998, M.Phil. in Computer Science from M.S. University, Tirunelveli in 2003, Ph.D. in Computer Science from PRIST University, Thanjavur in 2011 and M. Tech., in CSE from PRIST University, Thanjavur in 2013. He is having 24+ years of teaching experience and 13+ years of research experience. He guided more than 70 candidates at M.Phil level and guiding 8 candidates at Ph.D. level. Right now, He is working as Dean, Faculty of Computer Science, PRIST University, Thanjavur. His research interest includes in the areas of Big Data Analytics, Data Mining, Cloud Computing, Wireless Sensor Networking and Computational algorithms.

Mr.Vishal C received M.C.A. Degree in Computer Application from R.N.S.I.T Bangalore in the year 2006, perusing (Ph.D.) in Computer Applications from PRIST University, Thanjavur. Having 9+ years of teaching experience and 2 years of industry experience. Working as Assistant Professor, Faculty of Master of Computer Applications, RV College of Engineering, and Bangalore. My area of research interest includes Big Data Analytics, Data Mining, Cloud Computing.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)