

KKSSA Technique: A Novel Technique for Evaluating Human Sentiments from Text

Karamjeet Kaur, Mohammad Shabaz

Abstract: Social media become trending object, nowadays people of every age group interested in social media huge amount of live data available on social media. Social media websites have evolved to become a source of various kind of information. These social networks give the birth to sentimental analysis. Sentiments are the feelings or passion towards anything. In this research paper we tried to extract data from social networking sites by using association mining technique to know about the people's opinion about trending and upcoming matters. In this research paper we use KKSSA technique to perform sentimental analysis on the live data extracted from social source i.e. twitter. The key challenge was to perform sentimental analysis on different people's huge set of live data. Sentimental analysis performed on this data by comparing the live data set with predefines data set (negative and positive dataset). The result is depending upon the situation the sentiments are obtained for.

Index Terms: link prediction, Opinion mining, Sentimental Analysis, AS approach.

I. INTRODUCTION

Sentimental is the mental state or mental feeling which describes the human behavior towards anything. Evaluating sentiments is the process of understanding a speculation or opinion about an entity from written or spoken language. Prediction derives the relationship between a thing you know and a thing you need to predict for future reference. Evaluating Human Sentiments from text using Association Rule Mining to Perform Link Prediction from twitter data using R. Social Media websites have evolved to become a source of various kind of information. This is due to nature of Human on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these. Social media has become integral part of our emerging technological lives, as this is the age of technology where everyone tries to put forward their sentiments whether it may be anger happiness or whatever. People from different places read the posts and reply according to their state of mind having while reading, so research is about sentimental analysis of people writing on the internet i.e. on the live data. This research will perform the sentimental analysis. In this paper, we look at one such popular Social Media Website called Twitter and build

algorithm for classifying "tweets" into positive and negative sentiment. The data was fetched from twitter, compared with the already stored sentiments i: e negative and positive and the result obtained was absolutely negative. Testing was done at different stages to check the support of repository.

II. LITERATURE REVIEW AND RELATED WORK

Pandya Jalpa P and Morena Rustom D. [1]: The paper focuses on benefits of extracting the hidden data using association rule mining, this extraction opens the window of various field knowledge including stock market, health care etc, various applications of ARM algorithm are discussed. The author compares Apriori on various databases and checks the overall performance.

Raheesa Safrin and et.al. [2]: The author uses sentiment classification technique to analysis the comments, made by the customers while buying/reviewing any online product. The author extracts the data which is in the form of rating including star, like and dislike. One of the important aspects of natural language processing named POST (Part Of Speech Tagging) is used to make decision.

Muhammad Zubair Asghar and et.al. [3]: The article evaluates positive opinions on the basis of likes and negative on the basis dislikes. The author uses the data which is extracted from online web based product service providers and evaluate the product features based on customer's opinion. The features extraction is done using POST (Part Of Speech Tagging) and Lemmatization techniques of NLP(Natural Language Processing).

Deepika Jaiswal and et.al. [4]: In this paper authors performs association rule mining on data which is collected from social media, to arrange the collected data in the structured form the author applies various preprocessing techniques like tokenization filtration, stemming and indexing. And then applied the GARW (Generating Association Rule using Weighting Scheme) on structured data and find out the correlation and association between various item sets.

D. Sheila Freeda and M. Lilly Florence [5]: This research paper author worked on disease data using association rule mining and also compared the diabetes dataset. Author found the association between diseases. In this research author use data set which include the information about doctor, patient, disease, medicines etc and association mining rules are applied to find out the correlation among various ill health issues. Mohammad Shabaz and Ashok Kumar [6]: In this research paper authors counts the human sentiments range between -1 to 1 and forms 0 as neutral sentiment. The value to the left of 0 is taken as negative sentiment, whereas the value to the right of 0 is taken as

Revised Manuscript Received On April 07, 2019.

Karamjeet kaur, Computer Science Engineering, Universal Institute of Engineering & Technology, Mohali, India.

Mohammad Shabaz, Computer Science Engineering, Universal Institute of Engineering & Technology, Mohali, India.

positive sentiment. The algorithmic approach used by the author's is a novel approach named AS. Shalini T. and et.al. [7]: The author's focuses on making decisions about reviews that influence the customer to buy/purchase a product. Apriori and FP-growth algorithms of association rule mining are used to make opinion extraction. The authors compares the performance analysis of both these algorithms and finds that FP-growth algorithm is more scalable and takes less runtime during its execution.

III. DATASET

In the following **Figure 1** data set is shown. In this data set sentimental analysis is performed on 20 different items and result is achieved. Following attributes are in dataset:

Dataset Name: Dataset name contains 20 different tweets against which search are done.

Dataset Count: Dataset Count consists of the total number of words which are extracted from live twitter data after tokenization.

Sentimental count: it consists of sentimental count of fetched data after comparing that data with the stored repository. Sentimental counts range between -1 to 1.

Support of data set: it is categorized into negative data set support and positive dataset support.

Support of repository: Support of repository also categorized into two different parts i.e. support of positive repository and support of negative repository.

Error difference: Error is the difference between support of dataset and support of repository and is calculated by support of dataset minus support of repository.

Dataset no	Dataset name	Dataset count	Sentimental count	Support of data set(pos)	Support of data set(Neg)	Support of Repository	Support of Repository	Error difference
1	#AbhinandanIarthaman	2530	-0.7876712	0.0707	0.242	0.0959	0.1192	0.1056
2	Brazilian school	2620	-0.6622951	0.092	0.1672	0.115	0.0851	0.0591
3	acecook and instagram down on Wednesda	2582	-0.7262	0.1449	0.5621	0.0851	0.377	0.1449
4	puhama attack	2709	-0.6624473	0.1532	0.2627	0.2059	0.1256	0.0744
5	Bjp 2019	2678	-0.6995025	0.1524	0.296	0.2042	0.1512	0.093
6	CTS bridge collapse	2671	-0.6215047	0.1135	0.2295	0.1482	0.1104	0.0764
7	congres 2019	2722	-0.6402065	0.1136	0.1916	0.1499	0.1001	0.0541
8	Auckland bomb	2643	-0.6517572	0.1605	0.2641	0.2083	0.1323	0.084
9	alibhatt birthday	2652	-0.5030303	0.1427	0.1409	0.1789	0.0708	0.1755
10	city beautiful	2589	0.5269006	0.1643	0.1496	0.2005	0.0725	0.0409
11	#NSHSLAK	2632	-0.5097007	0.1303	0.1351	0.1617	0.0677	0.036
12	#mesoolAzh	2679	-0.6699408	0.1062	0.0112	0.1372	0.1011	-0.1209
13	#ethiopianAirlineCrash	2683	-0.5400163	0.1831	0.2093	0.2369	0.1049	0.0506
14	#krypto	569	-0.7731092	0.0545	0.1694	0.0131	0.0188	0.192
15	Chowkider Narendra modi	497	-0.7818082	0.0641	0.1953	0.0131	0.01881	0.22749
16	Manoj Bajpayee	568	-0.6716418	0.0878	0.173	0.0217	0.0184	0.2207
17	Donald J. Trump	544	-0.6618705	0.1018	0.1847	0.0232	0.0188	0.2445
18	#idhummosewala	488	-0.7927938	0.055	0.22	0.011	0.0179	0.2461
19	#DelhiKaHaqAitheRakh	538	-0.7348375	0.0604	0.178	0.0136	0.0186	0.2062
20	#CharMorePardeSiya	514	-0.7905612	0.0669	0.2279	0.0137	0.0227	0.2584

Figure 1 show the dataset extracted from twitter.

IV. METHODOLOGY

The KKSSA Technique can be implemented using the following steps:

1. Extract live data from social networking sites and save them as a text document.
2. Clean the data using ETL(Extract Transform & Load)
3. Tokenize the data into words.
4. Finding the length of data set.
5. Create a repository which includes trending positive and negative words.
6. Calculate the support of Dataset:
 - a. For negative dataset $(D \cap \text{neg}) / D\text{-pos}$
 - b. For positive dataset $(D \cap \text{pos}) / D\text{-neg}$
7. Calculate the support of Repository:
 - a. For neg: $(R_{\text{neg}} \cap \text{Neg}) / R_{\text{neg}}\text{-pos}$
 - b. For pos: $(R_{\text{pos}} \cap \text{pos}) / R_{\text{pos}}\text{-neg}$
8. Improve the support of our repository to 100% support.

V. RESULTS

The results shown in Figure 2, Figure 3, Figure 4 and Figure 5 describes of sentimental count, support of positive and negative dataset and Error difference against different input data repository.

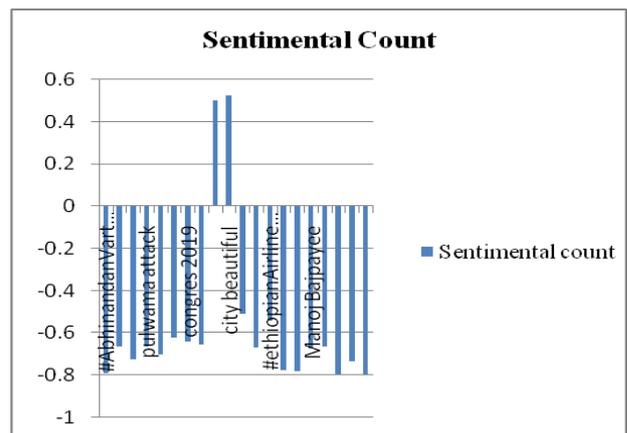


Figure 2 shows result of sentimental count against the input data repository

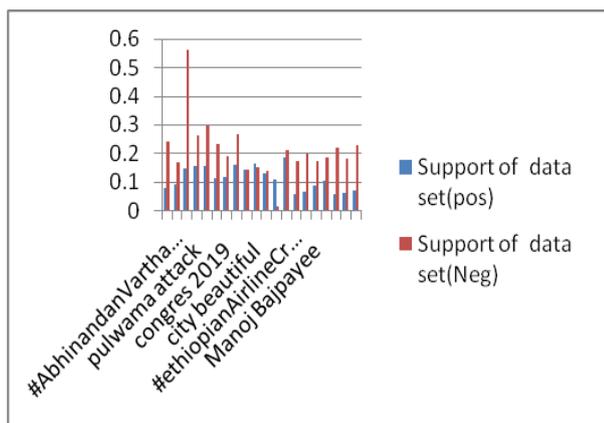


Figure 3 shows result of support of positive and negative dataset against the input data repository

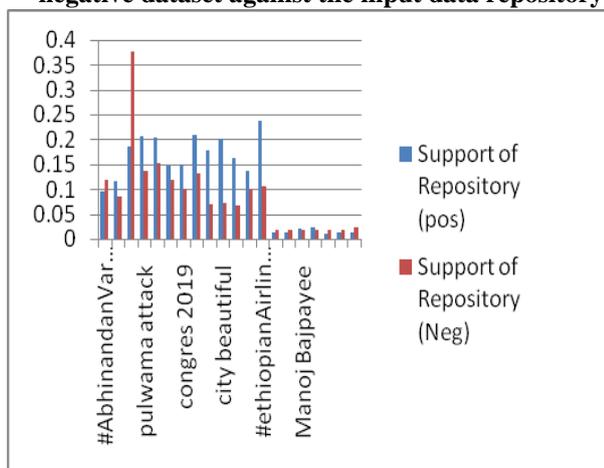


Figure 4 shows result of support of positive and negative repository against the input data repository

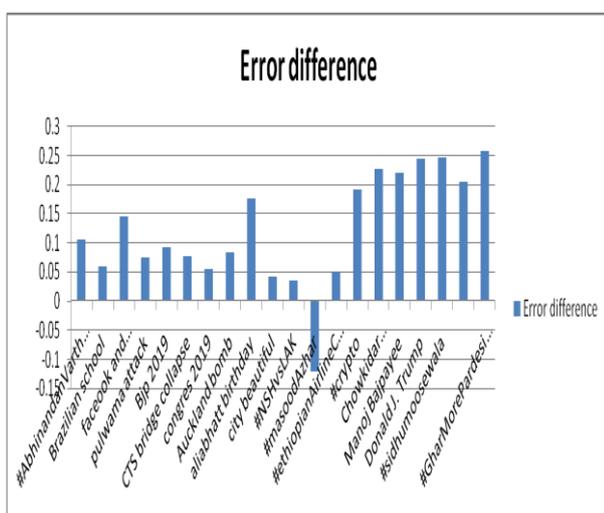


Figure 5 shows the result of Error difference against the input data repository

VI. CONCLUSION

The proposed methodology adopted is very simple technique to perform the sentimental analysis on live data to check positive or negative of any ongoing and upcoming situation.

For example the research has performed sentimental analysis on the live data of pulwama attack fetched from the twitter and result obtained was totally negative, similarly anyone can apply KKSSA technique to achieve the sentiment accordingly over the live data. Live data which is fetched from twitter compared with the data stored in repository to get the support of repository. In this paper we have tested 20 different dataset to analysis the result .The only drawback of our research is that we cannot perform analysis on every language as people from different places tweet in different languages and our research is restricted to English only.

ACKNOWLEDGMENT

We would like to thank with whole heartedly to Prishi Kapoor and all others who are either directly or indirectly provide their support and encouragement.

REFERENCES

1. Pandya Jalpa P,Morena Rustom D,," A survey on association rule mining algorithms used in different application areas", International journal of Advance Research in Computer Science may-june 2017.
2. Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal,," Sentiment Analysis on online product review", International Research Journal of Engineering and Technology (IRJET), Apr -2017.
3. Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad1, Fazal Masud Kundi1,,"A Review of Feature Extraction in Sentiment Analysis",Journal of Basic and Applied Scientific Research,2014.
4. Deepika Jaiswal1, Shilpa Singh2 , Suhasini VijayKumar3,,"Generating Association Rules for Social Media Analysis",IOSR Journal of Computer Engineering (IOSR-JCE),2018
5. D. Sheila Freeda, and M. Lilly Florence ,,"An overview of Disease Analysis using Association Rule Mining",International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
6. Mohammad Shabaz and Ashok Kumar ,,"AS: a novel sentimental analysis approach" International journal of Engineering &Technology,2018.
7. Shalini Tripathi, Maruska Mascarenhas,,"Survey on Association Rule Mining Algorithms for Opinion Mining",International Journal of Electrical, Electronics and Computer Systems (IJECS),2017.

AUTHORS PROFILE



Karamjeet kaur has done B.Tech,and currently Pursuing M.Tech in Computer Science Engineering from Universal Institute of Engineering and Technology (UIET), Lalru.



Mohammad Shabaz has done B.Tech, M.Tech and currently Pursuing Ph.D in Computer Science Engineering from Chandigarh University, Mohali. He is Currently Head of Department in Computer Science Engineering at Universal Institute of Engineering and Technology (UIET), Lalru. His Publication includes many UGC, Scopus and ESCI.