International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4, Issue 3, March – 2020, Pages: 13-17

Positivity Calculation using Vader Sentiment Analyser

S.Mukunda Rao, N.Monica[1], P.Nikhila, T.Tejasri, Dr.Balajee Maram[2]

Department of CSE, GMR Institute of Technology, Rajam, India
monicanakka2604@gmail.com[1], balajee.m@gmrit.edu.in[2]

Abstract: *Nowadays the data in online platforms is huge to make a reasonable analysis for providing a final view as either positive or negative. Even to buy a product, suggestions and guidance from the neighbours is needed in traditional days. Certain important events in daily life take place which are posted in online mediums. Everyone these days shows an inclination to know about a particular issue or event. In this regard, deploying the data from various users in the cloud enables us to store and perform various methods and applications on the data in the cloud. To get a view as either positive or negative and how positive the content reflects about the happenings in the world. In this paper the application of a Vader tool on the data stored in the cloud to analyse and score the polarities is provided. It gives the positivity of the content stores under a particular keyword. By measuring the polarity scores of each word, it provides the aggregated scores of overall text under the particular keyword. This application assigns scores to the emojis expressed in the content stored in cloud and provides a final view as either positive or negative and enhances the positivity score.*

Keywords —VADER ,Cloud, online mediums, Polarity scores**.**

1. INTRODUCTION

Vader Sentiment analysis is the process of identifying and designating the opinions expressed in a social media, in order to determine the user's intention towards any specific data, event, feedback, recommendations or warnings whether it can be positive, negative or neutral. The user's intention may be any one of the following situations such as his or her feedback or emotional state of the user when writing a review. Sentiment analysis can be highly useful in many real time cases like political voting behaviour, health care feedback, tourism, educational infrastructure etc. It is useful in getting feedback on health care facilities in urban or rural areas. In any organization voting, generally we can get the user feedback through online mediums about any organization's reputations or about the places and we can summarize about the overall opinions. Vader sentiment analysis is used to calculate the polarity of the written statements by the users in any domain or field. It helps the customers to know about the overall summary of a product without asking about it in traditional methods to family members, friends, neighbours. Considering various posts through social media and concluding the product or service is positive or negative. It is economical and feasible to use Vader sentiment analyzer for concluding genuine positive or negative reviews.

2. LITERATURE SURVEY

The procedure involved in the application of VADER to find the lexicon features in micro blog text is entered as feedback. It also offers brief explanations about the lexicon tools like ANEW (Affective Norms for English Words), SentiWordNet, and SenticNet which are corresponding with polarity scores for emotional intensity. We can collect reviews from the text and use it to assess how good or bad a product or service or any place is. It describes two approaches, namely the lexicon-based approach and the machine learning approach [1].

The development, validation, and evaluation of **VADER** (for **V**alence **A**ware **D**ictionary for s**E**ntiment **R**easoning) gives the working procedure in two widely used lexicons (LIWC, GI) in which words are classified as positive or negative binary classes according to their semantic orientation. It provides the application of a combination of lexical and quantitative methods and constructs and examines a list of lexical features (along with their corresponding polarity scores) [2].

It provides an overview of the current state of sentiment analysis, which is evolving very fast due to increased processing capability, access to data and advances in machine learning. It offers various descriptions of sentiment analysis and provides the challenges involved in sentiment analysis. It gives a brief description of different approaches to sentiment analysis like manual approach, lexical approach, Corpus-based approach and machine learning approaches. Pros and cons of various levels of analysis like document level analysis,sentence level analysis and phrase level analysis [3].

It explains semantic orientation on mood analysis in social media by studying the feedback reviews in micro blog text at a large scale also. It describes an individual's activity and his/her mood expression such as emojis. Emojis can be distinguished as positive or negative or neutral. People's moods are central to the expression of thoughts, opinions and altered effects can be present in behavior and feedback also. Such social media tools are used by individuals to broadcast their daily events or to report an external interest [4].

It gives a character-based approach to extract expressions from text. Semantic Orientation Calculator (SO-CAL) uses dictionaries of words with their semantic orientation.

(Polarity and energy) and includes intensity and negation. It describes how SO-CAL is applied to the polarization classification task and assigns a positive or negative label to the text that captures the textual impression of its main subject. In addition, it describes the process of creating dictionaries and the use of mechanical turks to check dictionaries for consistency and reliability.

[5].

It proposed an influential method for identifying semantic orientations of opinions expressed by user's. It is able to deal with two major problems with the existing methods, opinion words whose reviews are context dependent, and bundle of multiple opinion word meanings in the same sentence [6].

3. METHODOLOGY

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

3.1 VADER

VADER (ValenceAware Dictionary and sEntiment Reasoner) is a rule based sentiment analysis tool. The lexicon's semantic orientation classifies the user's data or reviews as positive, negative or neutral. The vader tool gives the compound score by the sum of all the lexicon ratings. Generally, the lexicon approaches does not need to train a model using labeled data. It should be accessed from the dictionary sentences. The reviews or ratings are not taken through one word, but it accesses the average of the overall ratings for each word. It depends on the context, situations, human ratings and human mood emotions. The input text calculates the semantic score which maps the lexicon features with the intensity. VADER performs very well with emojis, slangs, and acronyms in sentences.

The Vader tool gives the sensitivity between the words in the sentence.

1. Punctuation marks like the Question mark (?), Exclamation mark (!), Quotation marks (" ") etc. enhance the intensity of the words without changing the semantic meaning of the sentence. For example, the intensity of the sentence "The infrasture here is good!"is higher than" The infrastructure here is good ".
2. Capital Words :These are used to provide emphasis to the related words to enhance the value of the intensity of the words without changing the tic meaning of the sentence. For example : "The infrastructure here is GOOD ! ".The above sentence has greater intensity than "The infrastructure here is good ! ".
3. Degree Modifiers : These affect the value of the intensity of the words without changing the semantic meaning by either increasing or decreasing the intensity. For example : "The infrastructure here is extremely good". The above sentence has greater intensity than "The infrastructure here is good".
4. Differing conjunction: The differing conjunction like "but" indicates a change in the semantic meaning of the text. For example: "The infrastructure here is good but is very limited" . The above sentence provides the mixed sentiment by changing the overall rating.
5. Negation: It provides a change in the value of the polarity of the given text. For example : "The infrastructure here isn't good ". It gives the negative meaning and polarity to the sentence.

3.2 Firebase

Firebase is a real time cloud hosted NO-SQL database in which it is used to store and sync the data dynamically. Although Firebase is newer than AWS but it is advanced in technology rather than AWS. It is mostly better than heavy AWS setup and has a lower learning curve than AWS. It is not used for heavy datasets. Firebase is in SSL(Secure Socket Layer) and it is highly encrypted between server(Host) and client(Browser). Firebase is Google backup application development software which is used for developing web apps,iOS,Android. The storage of firebase saves the binary files,which stores directly from the client.

Firebase SDK has the application to store images, audio, video or various matters generated from different users. Firebase authentication is done by email/password. The firebase auth gives the authentication to the particular user's email for genuine users or developers. For static files it is easy to use a hosting service for the Firebase. Firebase database is integrated by firebase authentication. It also supports Facebook, Twitter, Google, Github etc; It is real time data, ready made API, built in security at data node level.But, The capability of queries are limited in the usage of firebase. By using traditional relational data models,NO-SQL cannot be implemented.

3.3 Compound Scores

VADER compares the user's text data with the VADER lexicon. VADER finds the polarity indices by Polarity score function. In which it returns the compound scores as negative, positive or neutral and compound for the user's text. The compound scores are calculated by the summation of all lexicon ratings that can be normalized in the range of -1 to +1. The +1 and -1 indicate the most extreme positive

and negative respectively. For multidimensional measures there are useful metrics. VADER not only analyzes data which are given by user's but also mostly concentrates on Punctuation, Capitalization, Emojis, Conjunctions and Degree modifiers. The threshold values of compound scores are as follows:

Positive Sentiment: Compound Score range from 0.5 to 1
Neutral Sentiment: Compound Score range from -0.5 to 0.5
Negative Sentiment: Compound Score range from -1 to -0.5

Compound Scores	
Nature of the Polarity	***Range***
Positive	0.5 to 1
Neutral	-0.5 to 0.5
Negative	-1 to -0.5

Table -1: Vader Lexicon Compound Scores

3.4 Flask

Flask is a lightweight WSGI (Web server Gateway Interface). WSGI explains the communication between the web server and web applications. It also explains the linking procedure of the web application to process the request. Flask is used for handling complex applications which are designed to start quick and easy. It is a simple wrapper of two libraries namely Werkzeug and Jinja. Flask is a framework for developing web applications. Framework is a core library which makes it feasible to developers while building web applications which are reliable,scalable,and also for maintaining. In Python, there are a number of frameworks including Flask, Pyramid, Django, Tornado etc. Flask is the python library used to develop a micro web framework

3.5 Architecture

The working procedure involves the execution of storing the data and analysing the data to assign the polarity scores to the words as shown in a figure.

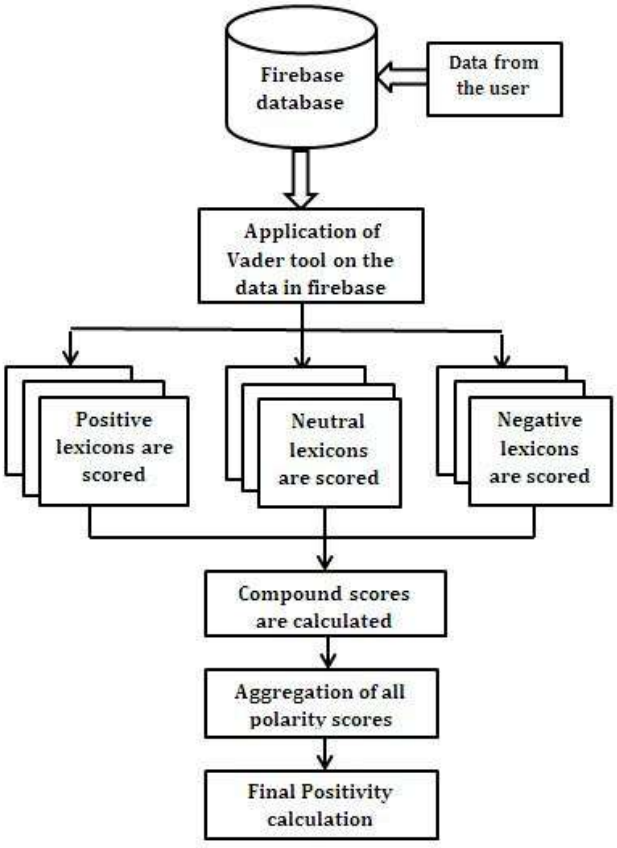

Fig - 1: The procedure involved for calculating positivity using VADER

3.6 Implementation

User's enter feedback or reviews or text which stores the data in the real time Firebase database. The data which is given by the user under a particular keyword which stores as JSON format in real time firebase database.The data from the user's text is extracted from the firebase database. The text undergoes through a VADER tool which are segregated as three different sentiments are positive, negative and also neutral. The compound scores are calculated by the summation of all the reviews given by the user as the reviews in the microblog text. Classifies emotions based on the polarity limits of text dictionaries. The aggregation of polarity scores of all statements gives the overall positivity under given particular keywords. VADER tool is used in real time scenarios like conducting surveys in obtaining data and analysing in political contests during elections, in getting the data relates to healthcare facilities

3.7 Results And Discussion

In our experiment, the data stores the in the firebase database through the process of email authentication in the cloud. The data can be entered into the firebase database through a webpage framework interface. The user can enter the data under a particular keyword into the database. Application of vader tool on the data stored in the firebase analyses the data stored under the requested keyword. It assigns the polarity points of each word to the semantic meaning of the word. Compound scores are calculated for the words stored as positive, negative and neutral words.

The results in this experiment give the positive value of the overall data stored under the particular keyword in the range of 0 to 100 percent. If the polarity scores of positive, negative and neutral words attains a value greater than 0.5 then the compound scores of positive words are assigned with a score of value 1 and the negative and neutral words are assigned with a score of 0 each. In this each particular keyword is stored with data as per the interest of the user. Aggregation of all the sentences under the particular keyword is calculated which gives the final positivity score in the range from 0 to 100 percent.

For example the results of positivity scores of the sentences below shows the corresponding scores by the application of vader sentiment analyzer.

The sentences "The furniture here is good", "The furniture here is not good" ,"The furniture here is bad" have a final positivity score as shown in the table.

Keywords	Statements	Scores	Positivity
Keyword1	The furniture here is good	1	100%
Keyword2	The furniture here is not good	0	0%
Keyword3	The furniture here is bad	0	0%

Table -2: Positivity Score of Individual Statements

In this experiment, for example a particular keyword "furniture" is given. Under this keyword, different users provide their content as "The furniture here is good", "The furniture here is not good " and "The furniture here is bad". When the keyword furniture is requested from the web page to the firebase database, the overall positivity score is provided by calculating the aggregation of all the individual polarity scores. For instance, the results of an example are shown in the table.

Keywords	Statements	Score	Positivity
Keyword1	The furniture here is good	1	33%
Keyword2	The furniture here is not good	0	
Keyword3	The furniture here is bad"	0	

Table -3: Positivity Score of Multiple Statements given under a single keyword

3.8 Conclusion

Vader provides the words consisting of the corresponding semantic meaning. Vader works well on the content available in the microblogs in the online medium. As the vader consists of the lexicons, it works and gives the results in less time. The performance of vader is good when compared with the performance of Textblob and NLTK. Considering a limitation value as 0.5 for the compound scores on positive, negative and neutral words, Vader provides the good final positivity scores. By using expressions, emphasis and various modes of grammatical styles, can enhance the magnitude of the sentiment intensity.

4. ACKNOWLEDGMENT (HEADING 5)

This research was supported by GMR Institute of Technology. I am thankful to my guide associate Prof. **Dr**.**Balajee Maram** who provided expertise that greatly assisted the research.

5. REFERENCES

[1] Venkateswarlu Bonta , Nandhini Kumaresh and N. Janardhan. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis .Vol.8 No.S2, pp. 1-6

[2] C.J. Hutto, Eric Gilbert. (january 2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Georgia Institute of Technology, Atlanta, GA 30032.

[3] Ge, J., Alonso Vazquez, M. & Gretzel, U. (2018). Sentiment analysis: a review. In Sigala, M.& Gretzel, U. (Eds.), Advances in Social Media for Travel, Tourism and Hospitality: New Perspectives, Practice and Cases, pp. 243-261. New York: Routledge.

[4] De Choudhury, M.; Counts, S.; and Gamon, M. 2012. Not All Moods are Created Equal! Exploring Human Emotional States in Social Media. In Proc. ICWSM '12. .

[5] Maite T., Julian B., Milan T., Kimberly V., Manfred S.(2011). Lexicon-Based Methods for Sentiment Analysis. Article in Computational Linguistics

[6] Xiaowen Ding; Bing Liu; Philip S. Yu. (2016). A Holistic Lexicon-Based Approach to Opinion Mining

[7] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R.(2011). Sentiment analysis of Twitter data. In Proc. WLSM-11s.

[8] Akkaya, C., Wiebe, J., & Mihalcea, R. (2009). Subjectivity word sense disambiguation. In Proc. EMNLP

[9] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0. In Proc. of LREC-10.

[10] Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings.

[11] Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2. In Proc. AAAI IFAI RSC-12.

[12] Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet. In Proc. of AAAI SCK-10.

[13] Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced Senti-ment Learning Using Twitter Hashtags and Smileys. ICCL-10.

[14] De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting Postpartum Changes in Emotion and Behavior via Social Media.

[15] In Proc. CHI-13. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013).Predicting Depression via Social Media. In Proc. ICWSM-13.

[16] Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In Proc. ICWSDM-08.

[17] Hancock, J. T., Landrigan, C., & Silver, C. (2007). Expressing emotion in text-based communication. In Proc. CHI-07.

[18] Nielsen, F. A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In Proc. ESWC-11.

[19] Pang, B., & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization. In Proc. ACL-04.

[20] H. Han, Y. Zhang, J. Zhang, J. Yang, and X. Zou, "Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias"„ pp. 1–11, 2018