Sentiment Analysis of Facebook Using Textblob and Vader

URMITA MEHTA, DR. DHANRAJ VERMA

Department of Computer Science, Dr. A. P. J. Abdul Kalam University, Indore(M.P.)- 452016

Corresponding Author Email: urmitasharma2020@gmail.com

Abstract— Sentiment Analysis is an emerging technology in the field of mining the text and symbol. In Social media, many information is available in many forms , either in textual form, symbolic form, image form or any other multimedia format, public gives their opinions which is calculated by mining it. In this paper vader and textblob tools are discussed and comparative study is done. In this paper different topics are used for the analysis purpose. Python programming library is used for the contents.

Index Terms— Artificial Intelligence, Machine Learning, Natural Language Processing, Opinion Mining, Sentiment Analysis.

I. INTRODUCTION

The term sentiment analysis also sometimes used interchangeably as emotion Artificial Intelligence (AI) , Opinion mining (OM). In simple term the sentiment analysis is NLP technique or a process in which it analyses the text or opinion and determines that if the text or sentence is having any emotion or not , if there is emotion, how much positive, negative or neutral it is. As Social media is considered a platform where people can put their views, opinions, experiences freely. These views or opinion may contain the information which is based on some facts. In general there are 2 types of information is present on social network, first is related to fact and the other one is related to sentiments. Fact may or may not have opinion. So for this reason we require some tool which is used to analyses those sentiments. These tools are called as sentiment analysis tolls or OM tools. These tools can generate information with high quality from some raw facts. The mining tools filter out the information and extract the important information related to particular topic. As it is merely not possible to read all articles, all messages available on social media so such tools are helpful in that case. The purpose of Sentiment analysis is to process the untrained data available in text format , determine meaningful number sequence from textual information and form the content from various algorithms.

SENTIMENT ANALYSIS THROUGH SOCIAL MEDIA

In Sentiment analysis of social media platform contains the use of social media like Facebook, Instagram, whats app and Twitter to determine the views of the crowd [4]. So for analysis Twitter as a social media is taken and put the key

words about particular subject or topic and then understand it from Natural Language Processing (NLP), to know what are the thoughts of people about that particular topic. There is a lot of challenges came across while doing this and many researchers and analyst tried to understand and figure out sarcasm, emojis, many of icons used by crowd on social media platform. Sentiment analysis is mostly used by businesses man to determine businesses product & brands for customer feedback, and find out the needs of customer, customers expectation. Which help them to understand Company's market position in the competitive edge. Mining the sentiments is very popular on social networks. These tools are used to determine and understand about reactions of people related to any incident, news or event [2][3]. It reflects the behavior of people what they think, expect, what are there thoughts and as well as ideas. So that analysis of sentiments on social media is much more representative and useful. The Natural Language Processing has many packages like textblob, sentdex, spacy and many more. All these packages are open source and provides many functionalities for Sentiment Analysis. the most popular NLP Sentiment analysis packages: Textblob and Vader.
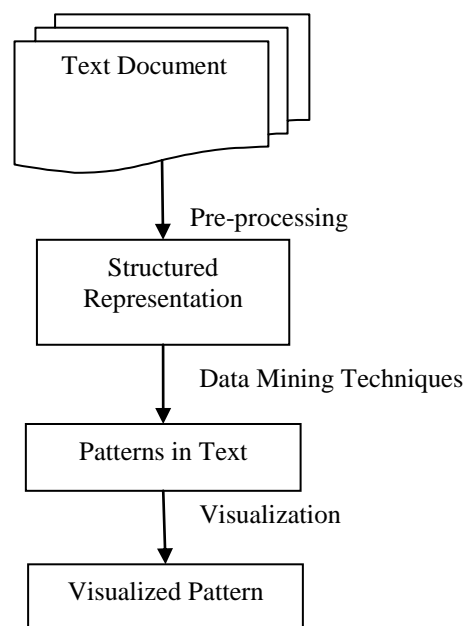

Figure 1: Text Mining Process

II. RELATED WORK

1. O. Araque et al [1] .In this paper the work is contributed in six fold. First, author have developed a classifier based on deep learning for sentiment using a word embedding's model and a linear machine learning algorithm. The classifier is taken as a base to compare the results. Second, they proposed two techniques for aggregate base classifier with other classifiers which is used in Analysis of Sentiments. Third, author also proposed two models for combining both surface and deep features to merge information from several sources. Fourth, they introduced a taxonomy for classification of various models found in the literature and they have proposed. Author, conduct several experiments to compare the performance of their models with the deep learning baseline. For this, they have used seven public datasets that were extracted from the micro blogging and movie reviews domain. Finally, as a result, a statistical study has confirmed that the performance of these proposed models surpasses that of their original baseline on F1-Score.

2 Y.Zhao et al [2], in this paper author has done preliminary research to analyze DIAC Twitter data in approaches for mining the and social network analysis. With text mining of twitter data, topics and the corresponding variations with time have been identified. The Twitter followers have been analysed and the spread of tweets over Twitter network has been studied. After analysis some initial interesting results were identified. The methodology used in this work is general, the tools used are open-source software, and the data used in this work were publicly available on Twitter.

3. K.Rizwan et al [3], In This paper authors have focused in providing the sentiment analysis on corona disease and the reaction of people about the decisions has taken by the government or other authorities on Twitter. They have proposed a system tool for twitter data which automatically analyzes the twitter post and divide them into positive, negative or neutral sentence. To achieve accuracy natural language processing is used. On the basis of pattern, division is done with natural language tool kit

4. B. Venkateshwarlu et al [4] In This article author have used tools like NLTK, VADER and Textblob for calculation of Sentiments of reviews for movie which is taken from the Cornell University data, and then in this paper they have done a comparison of the tools to find out the more efficient tools   for classification of sentiment. The final results of this work came out that VADER performs well then the Text blob.

5.S. Mrityunjay et al [5] In this paper, the authors have focused on opinions of people in India as well as people in the world so that they can find out the situations whether it is favorable or not. So the goal was to find out sentiment analysis related to corona disease. Author have observed that from many social media people have participated very actively .The author in their work used the twitter, as a social media platform, and opinion of the people have collected as post or reviews on covid. Authors have proposed a model and for data extraction twitter API is used. For their work 2 data sets have prepared .one for specification of world and other for India specifically. After that analysis of the sentiment is done using various matrices like Subjectivity, Polarity &

intensity. Author have also used BERT model that is Bidirectional Encoder Representations from Transformers so that on corona virus opinions of the people can be classified.

6. C.Hutto et al [6], In this paper author have presented VADER, which is model for the analysis of sentiments and then authors have done comparison to find how much this system is effective with the different practices. Author have used different techniques and tools which is machine learning based techniques and algorithms. In this paper Authors have taken combination of both the methods as qualitative and quantitative, firstly they have constructed list of features of lexical methods with the intensity of sentiments mainly tuned with micro blogs context. To emphasize and express the intensity Authors have combined the features of lexical taking five considerations of general rules for convention of grammar and syntactic. Interestingly, they found that to access the tweets for sentiments analysis   after using their model, VADER outperforms well in Accuracy that is 0.96 and 0.84.

7. C. Kaur et al [7] in this paper, sentiment analysis is done related to corona virus. As various countries and peoples get affected with covid-19 which is now a crucial issues in these days. So Authors have analysed sentiment of various people and took their opinion related to this disease. For this twitter streaming API is used for the analysis of sentiments on twitter data and for calculation of tweets as neutral, negative or positive. For analysis machine learning based algorithms and tools is used. In this paper, author have done their experiments using Python programming on various tweets. For this purpose twitter API and Natural Language Tool Kit library is used. Textblob is used for tweets or data on twitter. And through various visualizations results came out as negative, positive and neutral.

8. S.Kiritchenko[8] in his paper authors have described a system for the analysis of sentiments called state-of-the-art which that detects the sentiments in textual messages which are short informal such as tweets and Short Message Services and words with sentiments or a phrases in a message. According to author, The system is based on statistical text of a supervised approach for the classification. The features of Sentiments were mainly taken from twitter data which is based on lexicons specification. These lexicons were generated from the twitter data automatically which has emoticons and hash tags. To capture the words with sentiment author has done with negated contexts, and for that negated words a different lexicon of sentiment is generated.

9. S. Tiwari et al [9], In this paper, authors have proposed a model that will analyse the sentiment of people and emotions on social media. Authors have used deep learning and machine learning techniques. According to the author due to the long time and high use of social media, teenagers are stressed, depressed and emotional became weak. The various information which are taken in context of big data & data science. In this paper, author has shown that how meaningful information can be extracted from the data sets. In This paper data from social media is accessed and processed for the analysis of social media. In this paper for correct sentiment analysis various techniques of Natural Language Processing are used.

III. METHODOLOGY

FACEBOOK:SENTIMENT ANALYSIS OF FACEBOOK DATA

To analyze Facebook data , we need to perform following steps:

1. Gather the Data
2. Prepare the Data
3. Analyze the Data
4. use a Sentiment Analysis tool

1. Data Gathering: Collecting Facebook Data

To collect data from Facebook or from any other social media sites. There are many tools available for the public data sets, and APIs.

a) Zapier : Zapier allows to extract the data from an app and connect to it , using a "zap." One can extract Facebook posts that are mentioned company's name.

b) Dexi : Dexi web crawler allows user to export the data from Facebook to a CSV file, and offers direct integration with the tool.

c) Scrape Storm : user have to Just enter the URL, & hit ' *Start* ' & scrap storm will download the text to the file of users choice.

d)Content Grabber :With the Content Grabber web scraping tool user can pull text from any website and save it as a CSV file to upload to analyse the sentiment .

e) Pattern : Pattern is a GitHub web mining module for Python that includes tools for scraping or direct natural language processing. It has datasets for Facebook, Twitter, YouTube, and more.

f) API : The Graph API offers a fairly simple setup for Face book data extraction. The Graph API is the primary way for apps to write and read to the Face book social graph.
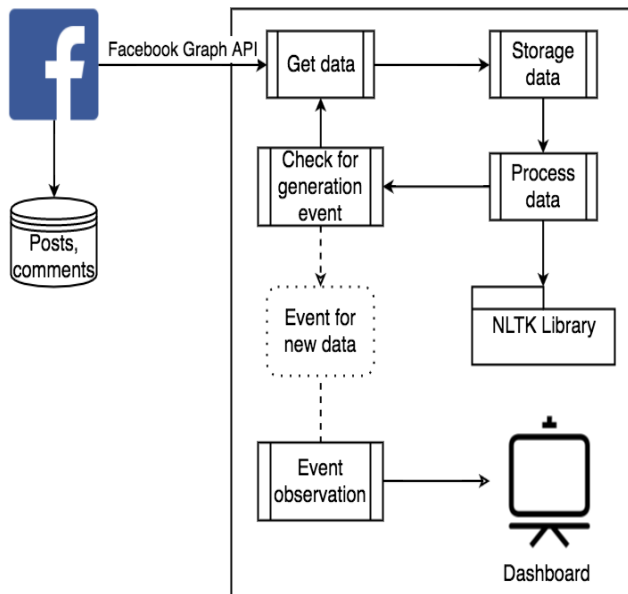
Figure 2: Processing of Facebook data using NLP

REQUIREMENTS

This work requires following packages to be installed:

- NLTK – The Natural Language Toolkit provides support to Python programs to work with human language data
- facebook-sdk - Python SDK for Facebook's Graph API
- Matplotlib - Matplotlib is a part of Python library for plotting of graphs
- scikit-learn – A Machine Learning library in Python
- pandas – is an open source library which provides high-performance, data analysis tools for the Python programming

First user need to download the post from a Facebook using the Facebook Graph API .The comparison of both the python Textblob API and Vader is done. Sentiment analysis (SA) is one of them. In many sentiment analysis systems this API is used. There are different tools and algorithms available for sentiment calculations on various social media platform.

TEXTBLOB

TextBlob is a part of Python library for the processing of text based data. The textblob provides the Application Programming Interface for the task of natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification translation etc.The Textblob has two properties for Sentiment Analysis.

1. polarity and
2. the subjectivity

Textblob identifies subjectivity and the polarity of the sentence expressed on social media.

POLARITY The polarity means emotions expressed in the sentence is analysed. That means that how much emotion contained in the sentence whether the sentence is positive, negative or it is neutral. It ranges from -1 ,0 and 1.
If polarity is -1That means that the sentence is negative.
If polarity is +1 That means that the sentence is positive.
And if polarity is 0 then it's a neutral sentence. That means its neither positive nor negative.

SUBJECTIVITY People used to express their ideas, views, thoughts and emotions on social media platform. The subjectivity means whether the sentence is having emotion or not. it is also important to know that whether the sentence is a fact (objective) or only an opinion (subjective). This is represented by the value 0 or 1 means
If the Value is 0 then it is objective
And if the value is 1 then it is Opinionated
from text blob import Textblob

test = Textblob(" The movie was quite good but the actors were ok")

print (test. Sentiment)

Table 1: Textblob Polarity

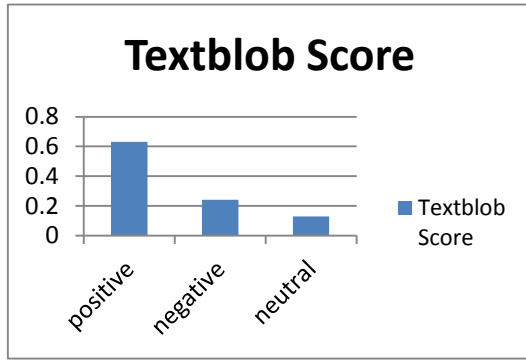| Polarity	Textblob Score
positive	0.632
negative	0.24
neutral	0.128

Figure 3: Textblob Score as per polarity

VADER

VADER stands for Valence Aware Dictionary and Sentiment Reasoner. The VADER is a tool which is based on lexicon for the sentiment analysis which is specially developed and designed for the analysis of sentiments on social media platform .The vader uses a features of lexical example. word which are labelled to find out for textual sentiments as negative or positive according to the orientation of semantic. After analysis Vader returns that the probability of a sentiments, negativity positivity . The Vader gives polarity as well as intensity.

Import nltk

Nltk.download('vader_lexicon')

From nltk.sentiment.vader import

SentimentIntensityAnalyzer

Si= SentimentIntensityAnalyzer( )

sentiment_analyzer_scores("The movie was quite good but actors were ok "){'neg': 0.0, 'neu': 0.335, 'pos': 0.754}

Table 2: Vader Polarity

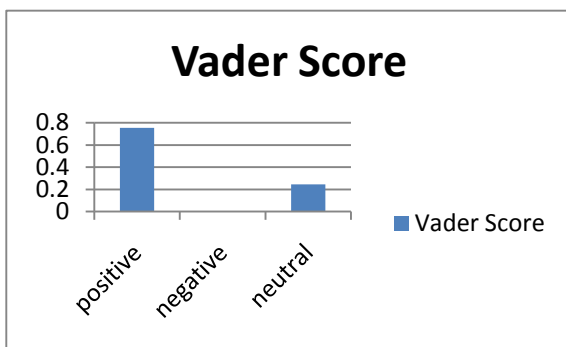| Polarity	Vader Score
positive	0.754
negative	0.0
neutral	0.335

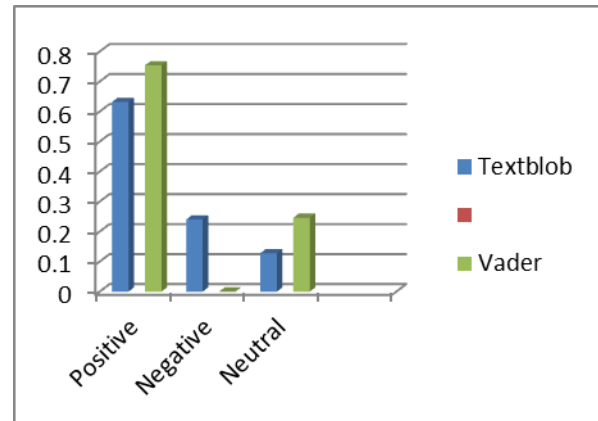**Figure 4:** Vader score as per Polarity

Figure 5: Overall Polarity of Textblob vs Vader

IV. CONCLUSION

With the development in the field of Artificial intelligence algorithms, it is now easy to understand and manage text based data. Moreover, this algorithm shows higher accuracy rates for the analysis of sentiments with the data. Both Textblob and Vader are host of many features and functionalities—Plotting the graph for the same for the comparison and found that Vader is mainly designed for social media platform and can give better results along with intensity when used with data from social media like Facebook, and twitter etc.

Sometimes the lacking is seen with the rule-based approach for analysis of sentiments is that the method only emphasizes about individual words and completely ignores the context in which it is used. The result found that VADER performs well than Text blob in sentiment analysis of tools.

REFERENCES

1. O. Araque, I. Corcuera-Platas, J. Fernando, S. Rada and A. Carlos. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", Expert Systems with Applications-Elsevier, Vol 17, No 9, pp 236–246, 2017.
2. Y. Zhao, "Analysing twitter data with text mining and social network analysis," in Proceedings of the 11th Australasian Data Mining and Analytics Conference, Vol 5, No 3, pp. 23, 2013.
3. R. Khan , P. Shrivastava , A. Kapoor , A. Tiwari , A. Mittal, "Social Media analysis with AI: Sentiment analysis techniques for the analysis of Twitter COVID-19" in Journal of Critical Reviews, Vol 7, No 9, pp 539-46, 2020.
4. B. Venkateswarlu , N. Kumaresh , N. Janardhan," A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis " in Asian Journal of Computer Science and Technology ,Vol.8 No.2, pp. 1-6, 2019.
5. S. Mrityunjay, A. Jakhar, S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model" In proceedings of IEEE MTT-S, Springer, pp1-11, 2021.
6. C. J. Hutto and E. Gilbert, "VADER : A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text", Proceedings of the Eighth International AAAI

Conference on Weblogs and Social Media, Vol. 10,No 7, pp. 216–225, 2014.

7. C. Kaur, A.Sharma," Twitter Sentiment Analysis on Coronavirus using Textblob" in Proceedings of research gate, pp 1-12, 2020.

8. S. Kiritchenko, X. Zhu, S. Mohammad, "Sentiment analysis of short informal Texts" in Journal of Artificial Intelligence Research, Vol 50 No 11, pp 723-762, 2014.

9. S. Tiwar and A. Sinha,"2 Sentiment Analysis of Facebook Data using Machine Learning " in International Journal of Innovative Research in Applied Sciences and Engineering, Vol. 4, no 4, pp 735-742, 2020.