

Crowd based Live Emotion Response Decoder using Social Media Platforms- Twitter

Shravani Sanjyot Shete

School of Computer Science and Engineering, Vellore Institute of Technology Vellore, India

Abstract:- Here in this project the main goal is to gain access to the social media platform posts, here we use Twitter. The most important asset here is the tweets that users post using their account. Using the Tweepy library in Python we extract these tweets using a get request and perform sentiment analysis on them. This is a classic example of using artificial intelligence for data analysis. Theory understands whether a tweet is categorized into a positive or negative category. The main intention here is to identify the crowd emotion using this method. This solution can be used in various fields and applications to gain a large scale input on an ongoing topic. We are using social media platforms such as it is one of the most used platforms.

Using sentimental analysis and hashtags we are able to gauge the sentiment behind the tweet. To analyze the tweets we are using the following libraries, TextBlob - this is used to keep processed the textual data gain from this live data set from Twitter.

For the implementation of the project we have gone through various approaches with different types of dataset such as hashtagged dataset, Emoticon dataset, Positive and Negative dataset. After the pre-processing of these datasets we calculate the accuracy and check which among the following gives us the maximum accuracy and we continue to use those methods for the further classification of our extracted features.

Keywords:- Artificial Intelligence, Twitter API, Sentimental Analysis, Textblob Library, Data Set, Vader Sentiment, Accuracy.

I. INTRODUCTION

Social media is a rich asset of data indicating the real world activity on a digitalized platform and twitter is one such social media platform.. It is the most mainstream miniature contributing to a blog website which permits their client to share data and short messages which are called tweets where a large number of individuals tweet each day. Twitter exchanges a wide variety of local and real-world events. Twitter aims to limit itself to two following features:

Here we intend the shortness of tweets, that is it cannot go beyond 140 characters.. It is easy to share messages to a large number of users within a short amount of time. During the global pandemic this platform was used to generate crowd resources on the availability of beds and other necessities in hospitals.

Tweets follow a generic pattern, they include information as follows:

User Mentions (@): Represented using the @ symbol indicate that a user has mentioned a friend user in their tweets/posts.

Retweet: Used to reshare the information that may have been originally tweeted by someone else.

Replies: This indicates that a reply has been made to a tweet or a retweet. Here the name of the replier along with the message is visible to all the users.

Hashtags (#): A hashtag is a component of the tweet that follows the hashtag symbol # that indicates the terms along with a reversed algorithm can be run in order to find all the tweets with the same hashtag. This denotes a trend or a supporter of the topic mentioned after the hashtag.

While there has been a considerable measure of exploration on how suppositions are communicated in sports, for example, online surveys and news stories, how conclusions are communicated given the casual language and message-length requirements of microblogging has been significantly less contemplated. Highlights, for example, programmed grammatical form labels and assets, for example, estimation dictionaries have demonstrated valuable for slant investigation in different spaces, however will they additionally demonstrate helpful for conclusion examination in Twitter? We will have a look at this feature through the implementation of this project

II. LITERATURE OVERVIEW

A number of research papers that were related to Movie preferences, personalities based on adjustment among youth were referred for this project. The research papers have also been reviewed to get ideas on how to make our project more relevant in the current scenarios.

A. Hutto, Clayton, 'and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the international AAAI conference on web and social media*. Vol. 8. No. 1. 2014.'

Authors present a basic model for showing the contrast in eleven common practice benchmarks that include common and most used algorithms and models such as LIWC, ANEW, the General Inquirer, Senti WordNet, and AI arranged procedures depending on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) calculations. Using these techniques they find out the required output and the standards of lexical highlights which are present in this microblog setting. Later the utilization of

another model for this survey, VADER gives us a F1 classification accuracy higher compared to the others.

B. Lee, Hooyeon. *Using 'Twitter to estimate 'and predict the trends and opinions. 'Diss. Master's thesis, Stanford University, 2012.'*

As we relate to charaIn this paper, we endeavor to foresee individuals' suppositions and patterns by investigating the Web information, for example, Twitter. There are a few fascinating inquiries we could pose, given billions of Tweets: how individuals feel about the economy; how individuals rate Obama; people's opinion of the best five contenders left in the TV show, American Idol. Gathering surveys by asking individuals face to face or by telephone is exorbitant, however in the event that we could get similar outcomes from the openly accessible Web information, it could enhance or override the ordinary method of gathering survey

C. Kouloumpis, 'Efthymios, Theresa Wilson, and Johanna Moore. *"Twitter sentiment analysis: The good the bad and the omg!."* *Proceedings of the international AAAI conference on web 'and social media. Vol. 5. No. 1. 2011.'*

In the following paper they have adapted the strategy to use the existing hashtags in on Twitter for building and preprocessing the information. They show the utility of etymological highlights in order to recognise different conclusions made from the tweets.

D. Wang, 'Hao, et al. *"A system 'for real-time twitter sentiment analysis 'of 2012 US presidential' election cycle."* *Proceedings of the 'ACL 2012 system demonstrations. 2012.'*

This paper depicts a framework for ongoing investigation of public opinion toward official up-and-comers in the 2012 U.S. political decision as communicated on Twitter, a microblogging administration. Twitter has become one of the most important websites where individuals feel free to express their voice, perspectives and views amongst various groups and competitors. Rising functions or news are frequently followed in a flash by a burst in Twitter volume, giving an exceptional occasion to check the connection between communicated public assumption and discretionary functions. Likewise, assessment examinations can help investigate how these functions influence popular feeling. While customary substance investigation takes days or weeks to finish, the framework shown here dissects conclusions in the whole Twitter traffic about the political race, conveying results immediately and ceaselessly. It offers the general population, the media, legislators and researchers another and opportune viewpoint on the elements of the constituent cycle and popular feeling.

E. Sharma, 'Nitesh, et al. *"Web-based 'application for sentiment analysis 'of live tweets."* *Proceedings 'of the 19th Annual International Conference on 'Digital government research: Governance' in the data Age. 2018.'*

Through this paper the authors have presented a web application that showcases the visualization of sentiments with respect to that keyword or phrase. This paper acts as an

inspiration as we also plan on proposing an application which provides a user with his service so that they can understand the crowd sentiment merely with a press of a button. Your users are able to see the tweet sentiment not only for the given keyword but also with respect to the specific geographical area as well. Hence providing the user with a platform that provides advantage of two features in one.

F. Punjabi, V. D', et al. *"A survey on trend' analysis on Twitter 'for predicting public 'opinion on ongoing 'events."* *International Journal 'of Computer Applications' 180.26 (2019): 13-17.'*

We wish to predict which topics will trend. And apply algorithms to find out what public opinion about that topic is used to predict mood. In this paper, we propose a model which uses a machine learning algorithm and classifies sentiment of twitter messages. For that we collect tweets, preprocess that tweet, find trending topics and apply multi classifier algorithms which predict public mood. We are going to use different measures such as precision, recall, and F-measure. We will be going to achieve better accuracy. Through this method we are finding out the keywords in the tweet and predict whether it is having weightage positive or negative weightage by applying a machine learning algorithm.

G. Phand, Shital Anil, and 'Jeevan Anil Phand. *"Twitter 'sentiment classification using stanford NLP."* *2017 1st international conference on intelligent systems and information management (ICISIM). IEEE, 2017.'*

In the Facebook and Twitter data, NLP is used in the following paper to extract the features. The tools follow a maximum entropy of machine learning algorithms. Stanford is an NLP methodology that is based on the maximum entropy. Here the twitter accounts for API contact us page authenticates the user credentials and does a word search on the fetched tweets.

H. Nguyen, Thien Hai, and Kiyooki Shirai. *"Topic modeling based sentiment analysis on social media for stock market prediction."* *Proceedings of 'the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.'*

Through this paper the main aim of the author is to arrange for a model that analyzes the market conditions and stock value using the available web based content. They have improved their accuracy from the previously developed models by around 6%. The results obtained thus tell us that the use of artificial intelligence in stock market prediction can prove helpful in individual and organizational gains. They also propose the topic based sentiment latent dirichlet allocation (TSLDA) in order to capture the key value and sentiments side by side. Solution if implemented on a large scale can be of great advantage

I. Pandarachalil, Rafeeqe, Selvaraju Sendhilkumar, and G. S. Mahalakshmi. "Twitter sentiment analysis for large-scale data: an unsupervised approach." *Cognitive computation* 7.2 (2015): 254-262.

Through this paper they are proposing a solution for diversity in the content demand based on domain independent solutions in order to identify and analyze sentiments on social media platform Twitter). The polarity value of the tweets here is being calculated using the following, SenticNet, SentiWordNet and SentislangNet. The F1 score obtained in this paper is good and could be further improved.

III. METHODOLOGY

For the implementation of this project, we follow these below steps

- Using the key credentials access the twitter API
- Access the datasets used for this project, that is 7
- Analyze the IMDB-dataset as it has various types of comments giving us a vast coverage on both short and long positive and negative comments- done using knn accuracy model.
- Twitter Stanford dataset is analyzed for its emoticons and the positive and negative data using a naïve bias model.
- Next we use tweets dataset of airlines and parsed preprocessed tweets and analyses them using textblob and Vader sentiment we observe a steep rise in its accuracy hence we go ahead and propose the following model for our project.
- Give a GET request to twitter API to get the tweets.
- Preprocessing of the textual dataset. This step includes the stemming, lemmatizing, tokenizing and reduction of unnecessary redundancy.
- Making functions to classify the texts
- Parsing and then classifying all tweets into categories, positive, neutral and negative.
- Then we use data visualization to represent the tasks using interactive graphs
- We then make a frontend using HTML, Bootstrap and js to push the visualized data on the portal developed for the viewers to see and analyze it.
- To display these interactive results as graphs we use canvas js.

A. Preprocessing and Working:

You're after getting all the required data from the tweets to be pre-process by arranging it in a readable format and making sense of the raw data.

In this process of Preprocessing the information we remove the unwanted and the noisy data from the chunk of data retrieved from Twitter. As we see in the fig1. below it is difficult to interpret any data in this obtained format. This process removes the unwanted lexicons, punctuation marks, slang and emoticons that are not included in the process of sentiment analysis. We then use a pre acquired list of words (a list of positive words and a list of negative words) used to classify the overall sentence into the appropriate sentiment. We can further improve the efficiency of this process by calculating the overall score of a sentence using the various words from the lists to get the better accuracy with respect to

the sentiments. Here the main task is to understand the tweet sentiment for associating them to a present situation. The trained dataset is labeled with 1, 0, and -1 to classify them into positive, neutral and negative respectively.

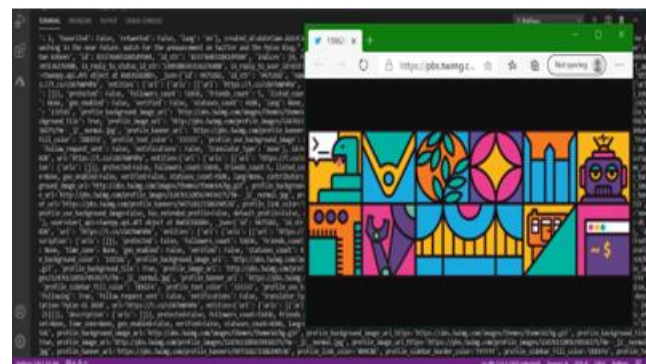


Fig 1. (a) Live tweets being parsed from twitter using the Tweepy API



Fig 1. (a)Raw data being fetched using the Tweepy API

```

RangeIndex: 20 entries, 0 to 19
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Tweets      20 non-null     object
 1   id           20 non-null     int64
 2   len         20 non-null     int32
 3   date        20 non-null     datetime64[ns]
 4   source      20 non-null     object
 5   likes       20 non-null     int32
 6   retweets    20 non-null     int32
dtypes: datetime64[ns](1), int32(3), int64(1), object(2)
memory usage: 784.0+ bytes

```

Fig 2. We extract the above attributes of twitter and convert them into the following dataset for further use

B. Results and Visualization of the Analyzed Tweets:

Exploring and visualizing data, no matter whether its text or any other data, is an essential step in gaining insights. We produce the following graph in fig 3. depicting the positive and negative sentiments until today. Also, we construct a pie chart to indicate the total number by percentage of positive, negative, and neutral tweets obtained from our analysis.

After completion of this project, the applications should be able to fetch and accurately classify a social media post as positive, negative, or neutral. Using the various methods available for generating the results we compare and rank them in accordance with their performance, accuracy and F1 score.

Hello, twitter enthusiasts!

The following website is developed identifying the need my mass media to serve their needs of knowing the positive and negative estimation of an agenda. And Twitter being one of the majorly used applications in the society to again importance or throw light on major concerns we have developed a simple analyser for the same. For this analysis we have used the VaderSentiment and Textblob method for analysis as they have a comparatively higher rate of accuracy. The form option displayed below has the default date set to today's date once you click on the submit option it generate a Data Visualization graph to show the presentation of positive and negative happening LIVE!

User Email address

Default Date

Set Graph image size

Fig 3. Application used to provide this feature as a service and obtain the required results

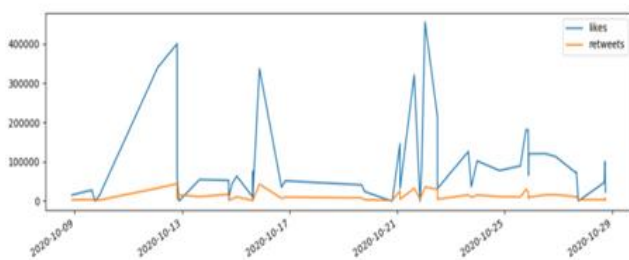


Fig 4. Tweets in the graph are within the time span mentioned during the extraction of tweets.

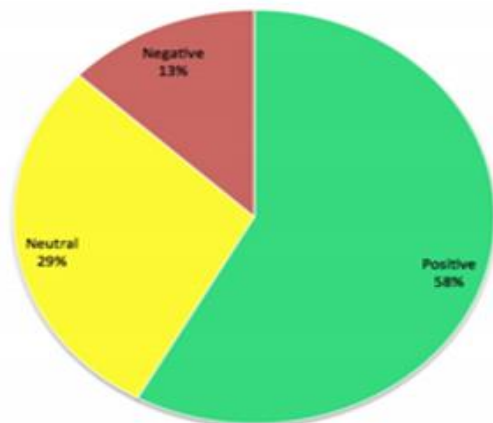


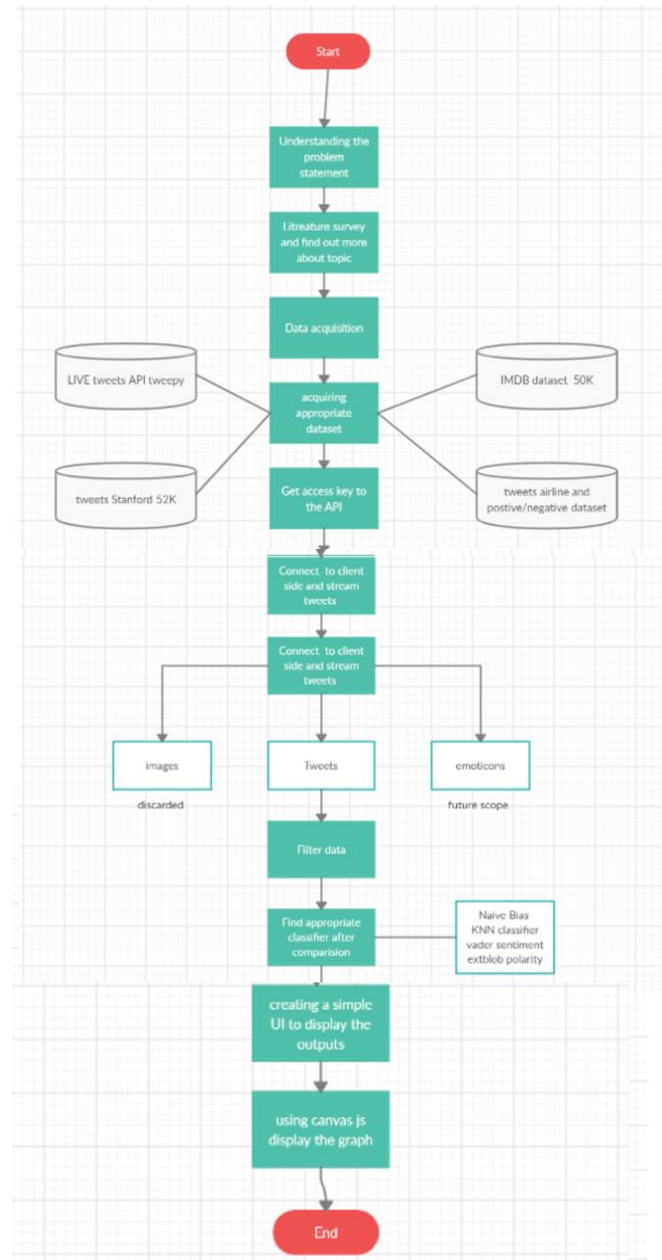
Fig 5. Sentiment analysis is performed on the tweets and the results are shown as the graph above

IV. DATASETS USED AND THEIR SPECIFICATIONS

Dataset	Size
Twitter live dataset using Tweepy API	50K live reviews
IMDB dataset	50K reviews
Tweets dataset with emoticons	13K tweets
Twitter airlines dataset	10K reviews

Stanford tweets dataset	52K tweets
Positive/ negative sentiments	1 lakh combined

V. FLOWCHART AND METHODOLOGY ADAPTED



VI. METHODS OF PREDICTION USED

We go over the following methods of prediction in our project with their respective predicted accuracy as below. The reason for going over so many methods is to analyze and evaluate the most suitable one.

KNN method- 69.98 %

Naïve Bias method- 82.3%

Vader sentiment method- ranges between 40 to 60%

Textblob polarity method- ranges between **98 to 100%** this method gives us a better accuracy compared to the rest.

VII. CONCLUSION AND FUTURE WORK

Through this project we understand that the raw data on twitter can be used to our advantage by categorizing it into useful information. Using the Tweepy library we can get live tweets from the social platform and understand the sentiments using the sentiment analysis of text. This process in a similar way can be used for producing and analyzing similar results for other social media platforms as well. In the end we can show the results obtained on our website prototype. The goal being to avail this solution to users using which they can get the information as necessary.

This process can then be used to specify a crowd emotion over a certain topic. A few applications of this project could be in the current pandemic situation, or for small scale business owners to market their projects using a trend amongst the crowd without the use of paid data analysts. It can also be used on mass media platforms to predict the results over the topics for going over debates or predictions of success of socially supported events such as elections or sports games (who will win this season of the IPL or the football league) or even used by radio jockeys to decide the music of the day.

ACKNOWLEDGEMENT

It is my pleasure to express a deep sense of gratitude towards my Artificial Intelligence Professor for her guidance in the subject. I would also like to express my gratitude to SCOPE School, Vellore Institute of Technology, Dr. G. Viswanathan, Dr. Sekar Viswanathan, Mr. G.V. Selvam, Dr. Rambabu Kodali, Dr. S. Narayanan and Dr. Anthony Xavier M, Vellore Institute of Technology, for providing such an environment to work and gain knowledge through the course.

REFERENCES

- [1]. Hutto, Clayton, 'and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Proceedings of the international AAAI conference on web and social media. 'Vol. 8. No. 1. 2014.'
- [2]. Wang, 'Hao, et al. "A system 'for' real-time twitter sentiment analysis 'of 2012 US presidential' election cycle.'" Proceedings of the 'ACL 2012 system demonstrations. 2012.
- [3]. 'Ceron, Andrea, et al. "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France." New media & society 16.2 (2014): 340-358.
- [4]. Paltoglou, Georgios, and Mike Thelwall. "Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media." ACM Transactions on Intelligent Systems and Technology (TIST) 3.4 (2012): 1-19.
- [5]. Sharma, 'Nitesh, et al. "Web-based 'application for sentiment analysis 'of live tweets." Proceedings 'of the 19th Annual International Conference on 'Digital government research: Governance' in the data Age. 2018.'
- [6]. "Spencer, James, ' and Gulden Uchyigit. "Sentimentor: Sentiment analysis of twitter data." 'SDAD@ ECML/PKDD. 2012.'
- [7]. 'Punjabi, V. D', et al. "A survey on trend' analysis on Twitter 'for predicting public 'opinion on ongoing 'events." International Journal 'of Computer Applications' 180.26 (2019): 13-17.
- [8]. 'Song, Junseok, et al. "A novel classification approach based on Naïve Bayes for Twitter sentiment analysis." KSII Transactions on Internet' and Information Systems (TIIS) 11.6 (2017): 2996-3011.'
- [9]. 'Abdul-Mageed, Muhammad, Mona' Diab, and Sandra Kübler. "SAMAR: Subjectivity and sentiment' analysis for Arabic social media." Computer Speech & Language 28.1 (2014): 20-37.'
- [10]. 'Phand, Shital Anil, and 'Jeevan Anil Phand. "Twitter 'sentiment classification' using stanford NLP.'" 2017 1st international conference on intelligent systems and information management (ICISIM). IEEE, 2017.
- [11]. "Rao, Tushar, and Saket Srivastava. "Analyzing stock market movements using twitter sentiment analysis." (2012): 119-123.'
- [12]. 'Bravo-Marquez, Felipe, Marcelo Mendoza, and Barbara Poblete. ' "Combining strengths, emotions' and polarities for boosting twitter 'sentiment analysis." Proceedings of 'the second international workshop on issues of sentiment discovery 'and opinion mining. 2013.'
- [13]. Jianqiang, Zhao, and ' Gui Xiaolin. "Comparison research on text 'pre-processing methods' on twitter sentiment analysis." *IEEE access* 5 (2017): 2870-2879.'
- [14]. 'Nguyen, Thien Hai, and Kiyooki Shirai. "Topic modeling based sentiment analysis on social media for stock market 'prediction.'" Proceedings' of 'the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th 'International Joint 'Conference on Natural' Language Processing (Volume 1: Long Papers). 2015.'
- [15]. Pandarachalil, Rafeeqe, ' Selvaraju Sendhilkumar, and G. S. Mahalakshmi. "Twitter sentiment analysis for large-scale data: an unsupervised approach." *Cognitive computation* 7.2 (2015): 254-262'.'