

How to Cite:

Mittal, N., Khan, I., & Sharma, P. (2022). Analyzing sentiments of human through twitter comments. *International Journal of Health Sciences*, 6(S5), 3854–3859.
<https://doi.org/10.53730/ijhs.v6nS5.9454>

Analyzing sentiments of human through twitter comments

Dr Neha Mittal

Associate Professor, Echelon Institute of Technology, Faridabad
Email: eit.nehamittal@gmail.com

Dr Irfanullah Khan

Professor, Echelon Institute of Technology, Faridabad
Corresponding author email: irfandotin@gmail.com

Pinky Sharma

Assistant Professor, Echelon Institute of Technology, Faridabad
Email: pinkysharma056@gmail.com

Abstract--This disquisition is to find the people behaviour through their comments or post on twitter what actually like or dislike about a particular product or topic. In this we use Hybrid algorithm to build more well grounded and sturdy process. In this growing epoch of online community users. A huge numeral of people every day pin (up) their point of view in the form of tweets. This approach is useful to organization in many different ways who get introduced or marked in the tweet. By classifying those tweets into positive or negative according to the sentences by converting the unstructured tweet in structured format .tweets are going through cleaning phase first of all and after that the retrieve of tweets have been adept through libraries by using the API of Twitter.

Keywords---Sentiment Analysis, Twitter API, KNN, Hybrid algo, NLP, Naïve Bayes.

I. Introduction

API Offers by twitter is very powerful for programing language python and java which we are using currently. That's why extricate information from Twitter post or comments is relatively easy. People can point out their opinions about their feelings in the form of "tweets" as we know that twitter is an online platform. All these tweets express the opinions of people about different topics. This process of analyzing sentiment is prophecy of feelings into word .this is deliberated to present an application so that we can analyze those emotions, outlook and

judgment written in those online comments or post. Accurately it is archetype of designate dialogue into positive and negative or unbiased labels.

In this disquisition we put out effort to build a system that will give sentiment rating to different piece of tweets that have been extricate on the basis of their sentiments from their sentences for this we created a java application during decision making we always look for the opinions of others. Single people ask for it while organizations do conduct polls or surveys. Now sentiment analysis is term to use to determine the attitude emotion or judgment of people about some specific topics.

The topic or document we want to culling are written in natural language which we do say unstructured data. so it is extremely difficult to detect sarcasm or noisy data. So we have to classify sentence according to the negative or positive by classification so this analysis is a classification problem.

II. Sentiment Analysis

The concept of analyzing the opinion of any specific point against a given content, To come on the outcome. This process is known as conclusion mining which or assumption mining or has several other terms to be called influence examination, conclusion extraction, and subjectivity examination, so on. Twitter is the micro-blogging portal that is the source of enormous information for the research to get knows sentiments of users on specific content or any product. There is a tone of data that is easily available and free to get hands on the concept of sentiments for research. The data can be access by fetching the Twitter API then we can apply the practical & theoretical concepts of machine learning, dictionary based approach, which generally categorizes the outcomes in the class of Positive, Negative, and Neutral which is the actual opinion of any product or it can be any opinion.

Tools available for the sentiments recognitions:

1. Tweepy: Its a python library to fetch Native Api of Twitter. Which provides all references for its methods which is sumptuous for the generations of bots and automation. The Stream Listener object oversees the tweets in real-time and catches them.
2. Text blob: Text blob is predefine NLP (Natural language processing) library. It is setup on the NLTK (Natural Language Toolkit) for using to perform various tasks pass from opinion mining to part-of- speech marking, tokenization, Classification, spell check parsing.
3. Most Efficient data frames which is an upgrade of NumPy in python. Pandas library offers groundbreaking, demonstrative, and accommodating information to get most meaningful benchmark, observation among various distinct objects.

III. Twitter

Twitter is considered as micro-blogging which is the most popular social media networking site, among others. It came to effect on 21st March 2006. It follows the concept of Tweets as the posts which are the short opinion of users it is the up-gradation of Blogs into the Micro-Blog. The Posts have character block of posts have fixed size. A tweet is just not a post, it contains Meta information related to the tweet. This makes Twitter is a more interesting website. The Tweets contain the meta-information which is used to discover the area of a tweet. further, we use Hashtags(#) trailed by a word which is a collection of related tweets, URLs, Twitter User ID which is unique, RT(Retweets), (@) is used to get client IDs.

A. Examine the Twitter Data.

Main purpose of examine the data is to figure out the different tweets to put in the Sentiments category. There are several ways in the area related to Twitter data While instructing the model, check its effectiveness. Tweets are not effortless to figure out to put where some reasons are as follows:

1) Distinct linguistics origins:-

Varying language due to different cultures with tweets using their native language, these words could be slang or native.

2) Character Limitation size of Block:-

It has 280 characters in scope, the number of words in the tweet should be recognizable and very limited.

3) Usage of Hashtags:-

Twitter provides hashtags to express the emotions, events, etc. which required to processed the tweet in a way to understand the given tweets. opinion mining is mainly collected of stages as shown in fig 1:

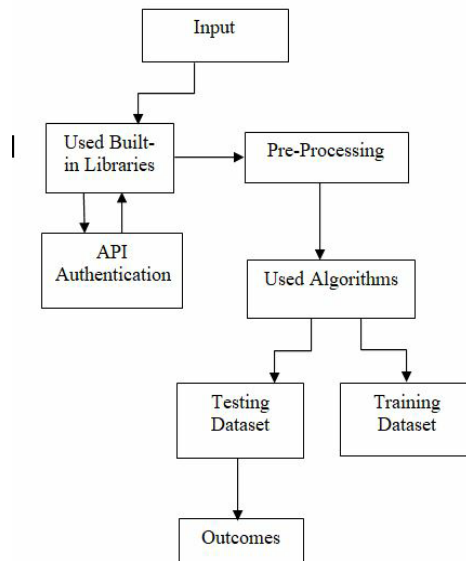


Fig-1: Working Flowchart

Input

The tweet is a combination of hashtags, tags, and many emotions, we can take any tweet as input.

Estimation Through Libraries

Libraries like Tweepy used the approach to authenticate the Twitter API and generate the token regarding them to evaluate the methods, sentiments of tweet.

Data Preprocessing

This step plays a crucial role in cleaning the noisy data from the tweets like as User-Id, White-Spaces, unit, hash tags(difficult ones), etc. Next to this stage, at most the major twitted data becomes present for mining.

Algorithm

Several algorithms are:- 1.KNN (K-nearest neighbors) 2.Naive Bayes 3.SVM (Support Vector Machine), etc. Used to edify data models and evaluate the produces outcome in the preprocessing stage. Explanation of algorithms is below.

KNN Algorithm

The Fundamental & efficient algorithm in AI for calculations. Used for both sequence and degenerate issues. The estimation depends upon the parallelism approach. The structure of the model can determine from the given dataset. It will be supportive in practice to do not follow the mathematical assumption on a real dataset. All training datasets were used in the trial period. This makes the training of the dataset is fast-moving and the testing period slower-moving and expensive. The expensive term in the testing period means that how much memory and time will use for the best and worst cases for training and testing of datasets. In a worst-case scenario, the algorithm (KNN) takes additional time & memory to examine all data nodes, scanning requires extra memory & time for organized training data.

This algorithm is basic and easy to get. Learning gather based model is KNN, the calculations are the torpid learning strategies. The algorithm scans the collection of the K-questions in the closest put together data to objects into the novel ones or data testing. The Euclidean Distance equation helps to characterize between the data points to put where in that classification. KNN techniques as follows

a) K initializes:

The approximation of the K implicit that clamor will affect the output and a vast value make it computation high. Generally we take an odd number if it is not. and the classes is 2 then we can take another easy method to deal with a selection of K. For which are to set the $K = \text{sqrt}(N)$.

b) Calculated gap between input data and trained datasets:-

Find the quantity of the closest neighbors using K with the Euclidean Distance equation for whole inputs data and each trained set of neighbors, then we are minimizing the gap in such a manner we can produce the closest gap between specimen and neighbors.

c) Closest specimen:-

Depending upon the distance evaluated between the inputs & trained datasets, then select the closest specimen (K - Neighbor).

d) Majority:-

Put in the majority of the dummy inputs with minimum gaps to get the most out of the KNN algorithms.

VI. Conclusion

Twitter is a social media networking site (micro blogging site) that contains a lot of unstructured datasets and sentiments of the users and a ton of information there. In Twitter the sentiment analysis digs out every tweet, this paper describes step by step that how the sentiment analysis on Twitter can be done using the various tools and algorithms. We use the KNN algorithm to get most out of it. The simple one is Naive Bayes, which is one to not just complicate things that categorizes the tweets in positive, neutral, negative. The tweets have been given as input to the sentiment analysis phase & it goes with the whole process of all algorithms and finally gives the outcome. Analyzing the tweets is to know to get the actual reviews of any emotions, product, etc. Every step in each phases of this concept are elaborated sequentially.

References

- [1] Sahar A. El_Rahman, "Sentiment Analysis of Twitter Data," Computer and Information sciences College Princess Nourah Bint Abdulrahman University,
- [2] Anurag P. Jain, "Sentiments Analysis Of Twitter Data Using Data Mining," Dept. of Information Technology Pimpri Chinchwad College of Engineering Pune, India, 2015 ICIP.
- [3] Rasika Wagh & Payal Punde, "Survey on Sentiment Analysis using Twitter Dataset," (MTech): Department of Computer Science and IT, Dr. BAMU Aurangabad, India, ICECA 2018.
- [4] Adyan Marendra Ramadhani & Hong Soon Goo, "Twitter Sentiment Analysis using Deep Learning Methods," Department of Management Information Systems Dong-A University Busan South Korea, 2017.
- [5] Rinantha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In 2017 5th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-4). IEEE.
- [6] Bing Liu, Sentiment Analysis and Opinion Mining Morgan and Claypool Publishers, May 2012.
- [7] V. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, vol. 139, p. 11, 2016.
- [8] Huma Parveen & Prof. Shikha Pandey "Sentiment Analysis on Twitter Dataset using Naive Bayes Algorithm," Dept. of Computer Science and Engineering Rungta College of Engineering and Technology Bhilai, India, 2016.
- [9] Herliah, A., Baso, Y. S., Hidayanty, H., Syarif, S., Aminuddin, A., & Bahar, B. (2022). Effect of web-based she smart education models on adolescent

girl's knowledge, attitudes, and practice about obesity. International Journal of Health & Medical Sciences, 5(1), 50-55. <https://doi.org/10.21744/ijhms.v5n1.1832>

- [10] Rincy Jose & Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach" ,Department of Computer Science and Engineering Rajagiri School of Engineering and technology Ernakulum, India , 201