

Real-Time Twitter Sentiment Analysis

Guided By- Prof.Shubhangi Chaware
Assistant Professor Department of IT

1. Miss. Prerana Dahiwade
preranadahiwade9@gmail.com

2. Miss. Disha Diyewar
dishadiyewar28@gmail.com

3. Mr.Shubham Kothe
shubhamkoth73@gmail.com

4. Mr.Ishan Kshirsagar
ishankshirsagar4@gmail.com

Abstract: In this paper, we present a system for Real-time sentiment analysis of Twitter data. The proposed system relies on feature extraction from tweets, using both morphological features and semantic information. For the sentiment analysis task, we adopt a supervised learning approach, where we train various classifiers based on the extracted features. Finally, we present the design and implementation of a real-time system architecture in Storm, which contains the feature extraction and classification tasks, and scales well with respect to input data size and data arrival rate. Using an experimental evaluation, we demonstrate the merits of the proposed system, both in terms of classification accuracy as well as scalability and performance. Detection of depression through messages sent by a user on social media can be a complex task due to their popularity and trends in them. In recent years, messages and social media have ended up being a very close representation of a person's life and mental state. This is a huge stockpile of data about a person's behavior and can be used for the detection of various mental illnesses (depression in our case) using Natural Language Processing and Deep Learning.

Keywords:

Machine Learning, Python, Web Development, HTML, CSS, JavaScript, Django, Jupyter Note Book.

Introduction:

As the internet is growing larger, its reach to the masses is becoming wider. Social Media and Microblogging platforms like Twitter, Facebook, and Tumblr dominate in spreading encapsulated news and trending topics across the globe at a rapid pace. A topic or news becomes trending if many users are contributing their opinion and judgments, thereby making it a valuable source of online perception on that particular topic. These topics are generally intended to spread awareness or to promote political campaigns, public figures during elections, product endorsements, and entertainment like award shows, and movies. Large organizations and firms take advantage of people's feedback on these platforms to improve their products and services which further help in enhancing marketing strategies.

Literature review:

[1] Michel M Tadesse, Hongfej Lin, Bo Xu, And Liang Yang 2019 detection of depression related posts in reddit social media forum we can significantly improve performance accuracy .the best single feature is bigram with the support vector machine (SVM) classifier to detect depression with 80% accuracy and 0.80 FA scores.the strength and effectiveness of the combined features (L1WC +LDA+bigram)are not successfully demonstrate with the multilayer with the multilayer perception (MLP) classifier.

[2] P.V.Rajaraman Asian Nath ,Akshaya P.R,Chatur Bhoja.G 2020 Depression detection of tweets and a comparative test message and social media has on ded up being a very close representation of a person life and his mental state .this is a huge itockpile of data about a person behaviour and can be used for detection of various mental illness using natural language processing and deep learning.

[3] **Neethu M S, Rajasree R**, the author have used the machine learning techniques in this survey paper to explore the twitter data related to electronic product. They have used feature vector for the tweets classification . they have used three types of classifier i.e. SVM, naïve bayes, maximum entropy, and these classifier were tested using Matlab simulator. SVM and naïve bayes classifier are implemented using built in function. Whereas MaxEnt classifier is used by MaxEnt software. So basically the all classifier have nearly the same performance.[7]

[4] **Pulkit et al.** built and proposed a model which extract tweet from twitter based on the post terror activities. they made their study on terrorist attack which was occurred in uri on 18 september 2016. They considered 59,988 tweet which had taken after the attack. They consider only those tweets which has #UriAttack, #uriattack. #uriattacks. They have used the naïve bayes and SVM to extract the last re-tweet time and number of re-tweet.[8]

[5] **Sudarshan Sirsat et al.** proposed a technique in sentiment analysis on twitter data where they have collected reviews of the product. They have used naïve Bayes algorithm which perform better in term of accuracy and efficiency. They have extracted 200 tweets where the average length of tweet was 70.105. the aim of this research is to identify the characteristic of tweet like how many times the tweet was liked and how many times they have re-tweet the tweet.[9]

[6] **Hetu et al.** proposed a model in sentiment analysis on twitter data based on anaconda python. They extract the dataset from kaggle in which they classify the people emotions based on positive and negative reviews. This model gives high accuracy on large dataset.[10]

[7] **Ali hasan et al.** proposed a model using the hybrid approach that comprise sentiment analyzer machine learning. They took only those tweet that is followed by the hashtag(#) and contain the current political trends. Basically this model converts the urdu tweet into English tweet. They took 1690 tweet for training data and 400 for testing the data. They have used the naïve bayes and SVM classifier for training the dataset in weka and building a model. They have used three different libraries to calculate the subjectivity and polarity.[11]

[8] **Feddah Alhumaidi Al Otaibi et al.** proposed a model by using the supervised and unsupervised algorithm. They wanted to know that which restaurant has more popularity between mcdonald and kfc by using the sentiment analysis. Moreover , they extracted 7000 tweets of both the restaurant by twitter API. The tweet was in English and they used R programming language.

[9] Because R programming language can perform big computational task. They have used several machine learning techniques but they found MaxEnt has performed better result compare to other technique. Moreover they have also found KFC have many neutral tweet. And McDonald have more positive and negative tweet.[12]

Problem Statement:

we have given a large collection of tweet that contain multiple types of features and opinions. Our task is to extract the opinion from the dataset that describes the target feature and distribute it as

positive or negative, neutral. Sentiment analysis deals with the process of extracting the features from people's views, opinions, thoughts and feelings, which they used to post on social websites. The outcome of SA is the classification of human language into the classes such as positive negative and neutral. Huge amount of data is generated from social networking sites. these data can be in any form whether unstructured or structured, generally, these type of data are unstructured. unstructured data do not convey any meaning until it is not analyzed. Therefore to utilize these unstructured data, there is a need of performing the sentiment analysis on this data i.e. take the valuable feature from these data and classify them. Nowadays, performing sentiment analysis on data is very important because the data on the internet is growing with very high rate and people are pretty much affected with the opinion of other people and they are distracted by this unstructured data. In today's world if someone wants to buy anything or he wants to watch a movie or he wants to sell anything on the internet then before doing any of the above activities, that person will go through the thoughts of other people which that means what other people think about that activity so it is very important. therefore we can conclude that we should generate that type of system which could automatically perform sentiment analysis on this big amount of data.

Methodology:

As mentioned earlier we aim to do sentiment analysis for Twitter data. By using various kinds of machine learning classifiers, we will build a classifier. once it gets trained then we will follow the different step to sentiment analysis mentioned in the below diagram:

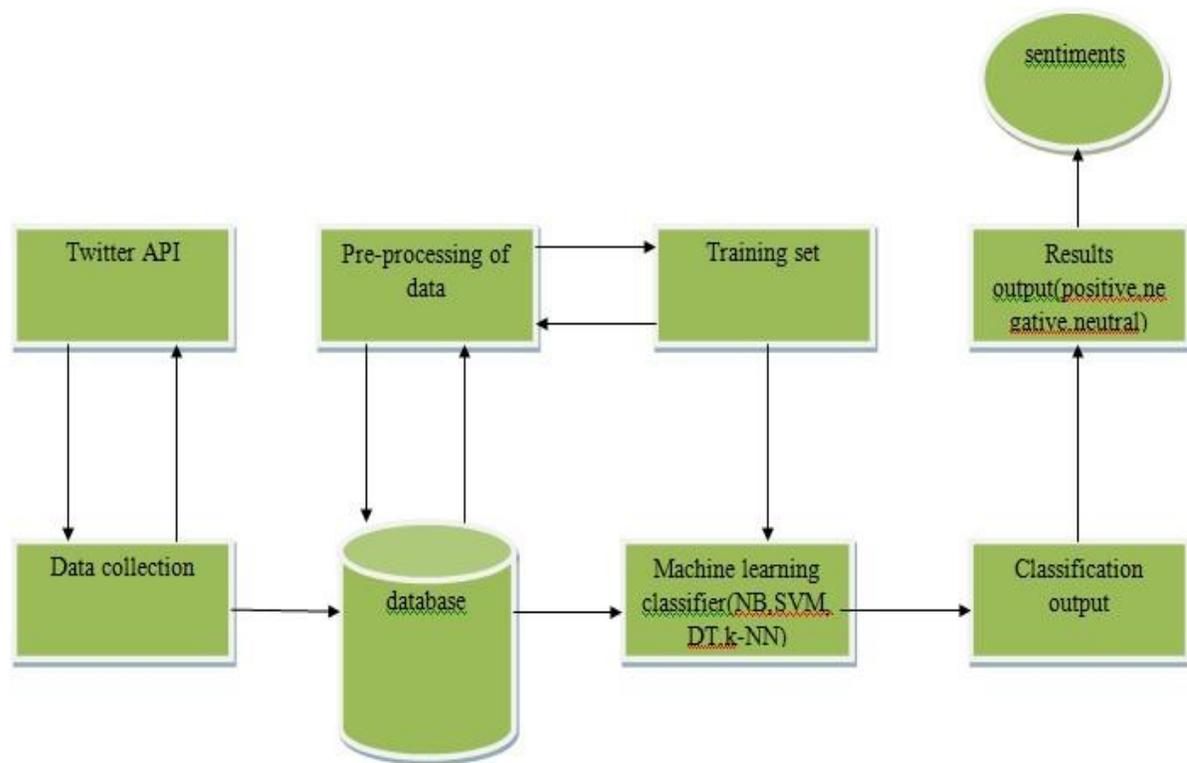


figure 4.1 flow diagram of Twitter sentiment analysis

Hardware/Software requirements:

1. Operating System:- Windows 7, and above
2. Language :- Python ,HTML ,CSS, JavaScript
3. IDE :- Jupiter Notebook , Django
4. RUN :- Chrome Browser ,Google etc.

Applications:

- 1) This real-time end-to-end Twitter monitoring system is designed for the enterprise to evaluate Twitter data to inform their negative or positive emotions.
- 2) In practice, keep tracking all relevant Twitter content about a brand in real-time, perform analysis as topics or issues emerge, and detect the anomaly with alerting.

Conclusion and Future Scope:

In today's world, a spacious amount of data is generated by various communication such as social media, organizations, etc. these data may or may not be in a structured form. Therefore to understand the polarity of data first we need to do the sentiment analysis of data. Opinion mining can be performed in various fields such as marketing and customer feedback. A large number of organizations are taking the valuable feedback of people and performing opinion mining on those data so that they could provide better services to the customer and this data helps the organizations to enhance their future services. Furthermore, there are various scopes where we can perform opinion mining such as sentence, paragraph, documents, and sub sentences levels. In addition to this, we took some sentiment classifiers such as support vector machine, naïve Bayes, decision tree, and K-nearest neighbor which performs best in terms of accuracy, precision, and recall. Out of these classifiers, we conclude that DT performs best in finding the accuracy of the twitter dataset. It is the best classifier on this dataset.

Basically, our goal in this thesis is to find the public opinion and perform the opinion mining. Generally what happens, through tweets people express their thoughts, feelings, etc. but we could not able to find the people's thoughts and feelings. So by performing sentiment analysis on those tweets finally, we can conclude how many people are in favor of this mission and how many people are against this mission.

Future scope includes, we can make a web application for our work. In addition to this, we can improve our classifier system such that it could deal with sentences that convey multiple meaning. Furthermore, we can add more classification categories so that we could get better results. We can also design a system such that it can detect the images in tweets with the help of image processing.

References:

- 1) Efthymios Kouloumpis and Johanna Moore, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.
- 2) S. Batra and D. Rao, Entity Based Sentiment Analysis on Twitter Stanford University, 2010.
- 3) Saif M. Mohammad and Xiaodan Zhu, Sentiment Analysis on of social media texts, 2014.

- 4)Ekaterina Kochmar, University of Cambridge, at the Cambridge coding Academy Data Science.2016.
- 5)Manju Venugopalan and Deepa Gupta, Exploring Sentiment Analysis on Twitter Data, IEEE 2015.
- 6)Brett Duncan and Yanqing Zhang, Neural Networks for Sentiment Analysis on Twitter.2017.
- 7)Afroze Ibrahim Baqapuri, Twitter Sentiment Analysis: The Good the Bad and the OMG! Proceedings of the Fifth International AAI Conference on Weblogs and Social Media.2011.
- 8)Kishori K. Pawar, Pukhraj P Shrishrimal, R. R. Deshmukh, Twitter Sentiment Analysis: A Review International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015.
- 9) Prof. SudarshanSirsat, Dr.Sujata Rao, Dr.BhartiWukkadada”Sentiment Analysis onTwitter Data forproduct evaluation” IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719PP 22-25.(2019)
- 10)Hetu Bhavsar, Richa Manglani” Sentiment Analysis of Twitter Data using Python”International Research Journal of Engineering and Technology (IRJET) Mar 2019e-ISSN: 2395-0056 p-ISSN: 2395-0072
- 9) Ali Hasan, Sana Moin, Ahmad Karim and ShahaboddinShamshirband” Machine Learning-Based Sentiment Analysis forTwitter Accounts” 2018 by the authors. Licensee MDPI, Basel, Switzerland.
- 10)Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi ,Wejdan Abdullah AlShehri “ Sentiment Analysis of Twitter Data”.
- 11) "India announces first manned space mission". Bangalore: BBC News. ^ Press Trust of India (25 April 2012). "Spaceflight stuck due to budget: CAG". *Times of India*. New Delhi. Retrieved 11 June2013.
- 12)Press Trust of India. "Human space flight mission off ISRO priority list". Retrieved 18 August 2013.
- 13)<https://indianexpress.com/article/what-is/what-is-gaganyaan/>
- 14)Priyadarshi, Siddhanta (23 February 2009). "Planning Commission Okays ISRO Manned Space Flight Program". *Indian Express*. p. 2.
- 15)Beary, Habib (27 January 2010). "India announces first manned space mission". Bangalore: BBC News.

- 16) E. Loper and S. Bird, "NLTK: the Natural Language Toolkit", Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics ,vol. 1,pp. 63-70, 2002.
- 17) P. Pang and L. Lee, "Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval", vol. 2(1-2), pp.1-135, 2008.
- 18) A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification", Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998
- 19) G. Kontaxis, I. Polakis, S. Ioannidis, and E.P. Markatos. Detecting social network profile cloning. In Pervasive Computing and Communications Work- shops (PERCOM Workshops), 2011 IEEE International Conference on, pages 295300. IEEE, 2011.
- 20) Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal (2014).
- 21) H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B.Y. Zhao. Detecting and characterizing social spam campaigns. In Proceedings of the 10th annual conference on Internet measurement, pages 3547. ACM, 2010.