



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



A Machine Learning Based Emotional Analysis Behaviour Using Questionnaires and Twitter

Shivapriyadharshini A¹, Priyadarshini K¹, Kalainilavu²

U.G. Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India¹

Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India²

ABSTRACT: How various societies respond constantly given an emergency is transcendent in a general public's standards what's more, political will to battle the circumstance. Frequently, the choices made are required by occasions, social pressure, or the need of great importance, which may not address the country's will. While some are satisfied with it, others may show disdain. COVID-19 brought a blend of comparative feelings from the choices taken by their individuals especially students. Online media was barraged with posts containing both positive and negative slants on the COVID-19, pandemic, lockdown, and hashtags recent months. Regardless of geologically close, many people responded diversely to one another. The motivation behind this examination is to dissect response of residents from various students to the novel pandemic or any related issue and individuals' notion about ensuing activities taken by individuals during such prevailing conditions. An optimal SVM algorithm and Naive Bayes Classifier is used for text classification to determine whether a given tweet belongs to a set of predefined categories because it is very fast approach and gives high accuracy. The bag-of-words method is a representation used in natural language processing, it used as a training data to build the predictive models. Classifying the tweets this way helps us to determine different emotions of the people and we can provide support if needed.

KEYWORDS: Machine Learning, Python, Support Vector Machine, Naive Bayes, Twitter

I. INTRODUCTION

Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets. These procedures operate by construction of a model from example inputs in order to make data-driven predictions or choices rather than following firm static program instructions. "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." So, if we want our program to foresee, for example, traffic forms at a busy node (task T), we can run it through a machine learning process with data about previous traffic patterns (experience E) and, if it has successfully "learned", it will then do better at predicting upcoming traffic patterns (performance measure P).

We need machine learning in the following cases:

- Human expertise is absent. E. g. Navigating on Mars.
- Humans are unable to explain their expertise. E. g. Speech

Consider the recognition of spoken speech, where an acoustic speech signal is converted to ASCII text. The pronunciation of a word may vary from person to person due to differences in age, gender or pronunciation, so in machine learning, the approach is to collect a large collection of sample utterances from diverse people and learn to plot these two words. As another example, consider routing packets over a computer grid. The trail maximizing the quality of service from source to destination changes regularly as the system traffic changes. A learning routing procedure is able to adapt to the best path by monitoring the network traffic.

II. RELATED WORK

In the year 2018, Dunkel, Alexander; Andrienko, Gennady [1]. proposed Events are a core concept of spatial information, but location-based social media (LBSM) provide information on reactions to events. Individuals have varied degrees of agency in initiating, reacting to or modifying the course of events, and reactions include observations of occurrence, expressions containing sentiment or emotions, or a call to action. Key characteristics of reactions include referent events and information about who reacted, when, where and how. Collective reactions are composed of multiple individual reactions sharing common referents. They can be characterized according to the following dimensions: spatial, temporal, social, thematic and interlinkage. We present a conceptual framework, which allows characterization and comparison of collective reactions.

In the year 2016, A. Jenifer Jothi Mary, S. Santiago [2] proclaimed Now-a-days, students write feedback of the courses in every semester and discuss with their friends in chat rooms, mails and social media sites. This type of discussions on social media which is called as opinions, may be in the form of feedback, remarks, reviews, comments, clarifications, observations and explanations irrespective of their course. It is evident that these opinions may help the teaching community to identify students’ feelings and attitudes towards teaching learning. The feedback taken during the course will help to address the issues of teaching-learning rather than taking at the end of every semester or a course. So, the opinions collected from the social sites, need to be analyzed using any of the sentiment analysis techniques for better understanding of the students.

In the year 2020, Richard J. Medford, Sameh N. Saleh, MD [3] experimented, Twitter has been used to track trends and disseminate health information during viral epidemics. The CDC activated its Emergency Operations Centre and the WHO released its first situation report about Coronavirus disease 2019 (COVID-19), sparking significant media attention. How Twitter content and sentiment has evolved in the early stages of any outbreak, including the COVID-19 epidemic, has not been described.

III. PROPOSED METHODOLOGY AND DISCUSSIONS

Information and emotions towards public issues could spread widely through online social networks. Although aggregate metrics on the volume of information diffusion are available, we know little about how information spreads on online social networks. Using them various models have been built. The existing system uses sentiment analysis technique to identify the sentiment of the tweet. However, this is not sufficient to identify the anti-social messages. The sentiment analysis might give the indication of the emotion related to positive or negative. But it is important to classify whether the tweets considered the student activities.

We propose a machine learning based classification algorithm to analyse the live tweet. In this system we are using a classifier to identify the words used in the tweet. The algorithm takes specific keywords list as input and the tweet to be identified as another input. The results were plotted in a graph. In today’s digital world lot of data is available for sentiment analysis of both image and text through social media networks like Face book, twitter etc., but using data from some social media networks raises concern of privacy, thus, twitter is the most suitable social media network for getting enough data and also to protect us from privacy laws. This proposition aims to apply natural language processing on Twitter feeds for conducting emotion analysis focusing on depression. Individual tweets are classified as positive or negative, based on a curated word-list to detect depression tendencies. Numerous machine learning algorithms are available for the classification of the tweets.

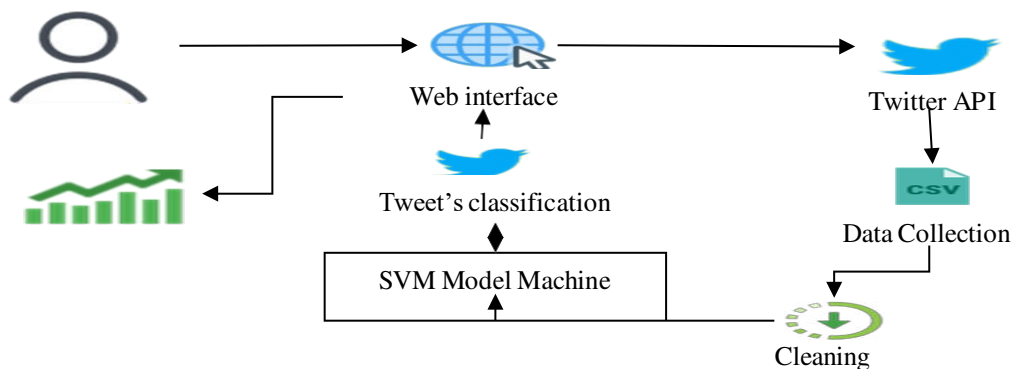


Fig 1 Architecture diagram



DATA COLLECTION

A Python API is used to access Twitter server for searching tweets. Generating Access Key and consumer key from Twitter website to access tweets. Initiating OAuth (Open Authorization) with the access key to login to twitter server for tweet search. OAuth allows the python to access the twitter server without exposing the username and password.

DATA PREPROCESSING

The majority of the tweets usually can be divided into 3 parts, not specifically in the same order. The first part contains the people whom the tweet is intended to or who is it referencing. This is usually denoted by "@". Example: @Hickman. The second part contains the actual message. This is what the user actually wants to convey. The third part contains the hash tag denoted by "#". This is usually to categorize the tweets. This can also be used by others to find tweets related to the particular content. Example: #SupportWHO. The tweets are pre-processed to filter the first part and third part since they do not hold very less to zero significance in sentiment analysis, later, the remaining message part of the tweet is pre-processed further for obtaining useful keywords which accounts much significance in identifying the emotions

MACHINE LEARNING ALGORITHMS

Naive Bayes Classifier Naive Bayes Classifier implements Bayes theorem with a solid independence assumption, particularly, independent feature model. Bayes Theorem works on conditional probability which finds out the probability of an event given that some other event has already occurred. It predicts the conditional probability of a class given the set of evidences and finds the most likely class based on the highest one. A naive Bayes classifier is a famous and popular technique because it is very fast approach and gives a high accuracy.

Support Vector Machine Support Vector Machine (SVM) is a supervised learning algorithm that analyses the data and recognizes patterns. Given an input set, SVM classifies them as one or the other of two categories. SVM can deal with both linear and non-linear classification. With kernel trick, it can efficiently perform non-linear classification. It does so by mapping the input set to high dimensional feature space. The types of kernel include polynomial, Gaussian radial basis function (RBF), Laplace RBF kernel, Hyperbolic tangent kernel and Sigmoidal kernel. Construction of hyper plane is employed by SVM for the classification. The goal of a text classification system in this paper is to determine whether a given tweet belongs to a set of predefined categories. An optimal SVM algorithm for text classification does this via multiple optimal strategies.

TRAINING & TESTING PROCESS

Creation of bag of words model: The bag-of-words method is a simplifying representation used in natural language processing and information retrieval. In this model, a text like tweets is represented as the bag of its words, disregarding grammar and even word order but keeping multiplicity. The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each word is used as a feature for training a classifier. Ex. Joey likes food Joey likes games. The bag of words will be ["Joey":2," likes":2," food":1," games":1]. The Bag-of-words model is mainly used as a tool of feature generation. After transforming the text into a "bag of words", we can calculate various measures to characterize the text. The most common type of characteristics, or features calculated from the Bag-of-words model is term frequency, namely, the number of times a term appears in the text. Creation of Predictive model: Two Predictive models are created using naïve Bayes and support vector machine algorithms which were discussed afore. The data file will have two columns with one column representing either the tweet categories and other will contain the actual tweet. The column with tweets is pre-processed. The pre-processed data is used to create bag words model which will be used as training data to build the predictive models.

Testing Process: In testing phase, the 30% of data which was split randomly from the dataset is tested on the predictive model. The test data is pre-processed and classified either.

DATA VISUALIZATION:

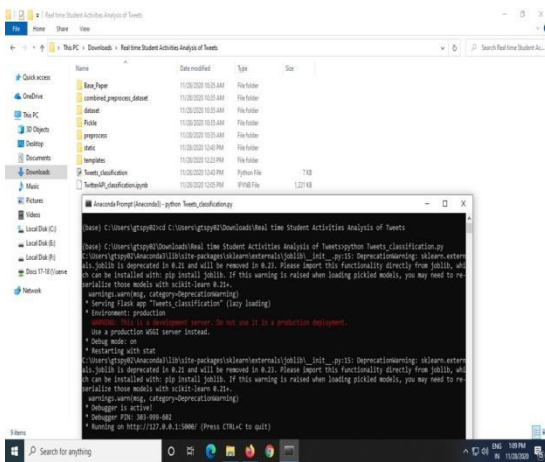
The trained model can be stored as a Python object so that one need not train each and every time we need to predict. The accuracy of the trained model is calculated before prediction. We use data visualization to present the result in a visualized manner.



IV. EXPERIMENTAL RESULTS KEYWORDS EXTRACTED FROM DATA SET REPRESENTED IN WORD-CLOUD



Fig 2 Keywords Extracted From Data set Represented In Word-cloud



(a)

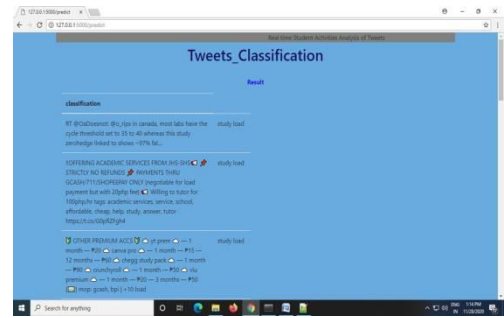


(b)

Fig (a) Starting The Server (b) Login/Register



(c)



(d)

Fig (c) Predict The Tweets That Talk About Study-Load

(d) Results Of The Tweets With The Keyword And Their Predicted Emotions

IV. CONCLUSION AND FUTURE ENHANCEMENTS

CONCLUSIONS

Based on the above observations, we can conclude that for the context of depression detection Multinomial Naive Bayes algorithm has performed well since the precision score is higher for Multinomial Naive Bayes. We also observed anomalies like conjoined words, separation of which will increase the efficiency of the model. Though better results have been achieved in the previous proposals, those results are over a limited number of tweets and since data is dynamic, it requires dynamic methods to tackle all the obstacles that arise in the sentiment analysis in the future.

FUTURE ENHANCEMENTS

Till to date, the pandemic is still rising in other parts of the world, including Brazil and Russia. It would be interesting to observe more extended patterns of tweets across more countries to detect and assert people's behavior dealing with such calamities. Especially, tweets and other social media platforms' post in local languages like Urdu, Hindi and Swedish etc. may reveal even more interesting patterns related to pandemic. We hope and believe that this study will provide a new perspective to readers and the scientific community interested in exploring cultural similarities and differences from public opinions given a crisis, and that it could influence decision makers in transforming and developing efficient policies to better tackle the situation, safe-guarding people's interest and needs of the society.

REFERENCES

1. Praseon Gupta, Sanjay Kumar, R. R. Suman, and Vinay Kumar, "Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter" IEEE July 2020
2. Ali Shariq Imran (Member Ieee), Sher Muhammad Daudpota, Zenun Kastrati, And Rakhi Batra, "Cross-Cultural Polarity And Emotion Detection Using Sentiment Analysis And Deep Learning On Covid-19 Related Tweets" IEEE Received September 21, 2020, accepted September 24, 2020, date of publication September 28, 2020
3. T. Beck, Depression: Clinical, experimental, and theoretical aspects. University of Pennsylvania Press, 1967.
4. Benton, M. Mitchell, and D. Hovy, "Multi-task learning for mental health using social media text," CoRR, vol. abs/1712.03538, 2017.
5. R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, "Natural language processing in mental health applications using non-clinical texts," Natural Language Engineering, vol. 23, no. 5, pp. 649–685, 2017.
6. G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "From adhd to sad: Analyzing the language of mental health on twitter through selfreported diagnoses," in Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015, pp. 1–10.
7. E. Durkheim and A. Suicide, A study in sociology. Routledge & K. Paul London, 1952.
8. J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preo, tiuc-Pietro, D. A. Asch, and H. Schwartz, "Facebook language predicts depression in medical records," Proceedings of the National Academy of Sciences, vol. 115, no. 44, pp. 11 203–11 208, 2018.
10. S. Freud, "The psychopathology of everyday life. se, 6," London: Hogarth, 1901.
11. M. Friedrich, "Depression is the Leading Cause of Disability Around the World Depression Leading Cause of Disability Globally Global Health," JAMA, vol. 317, no. 15, pp. 1517–1517, 2017.
12. S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," Current Opinion in Behavioral Sciences, vol. 18, pp. 43–49, 2017.
13. Á. Hernández-Castañeda and H. Calvo, "Deceptive text detection using continuous semantic space models," Intelligent Data Analysis, vol. 21, no. 3, pp. 679–695, 2017.
14. D. Maupomé and M.-J. Meurs, "Using topic extraction on social media content for the early detection of depression," in Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018., 2018.
15. T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 217–226, 2014.
16. M. Nadeem, "Identifying depression on twitter," CoRR, vol. abs/ 1607.07384, 2016.
17. S. Paul, S. K. Jandhyala, and T. Basu, "Early detection of signs of anorexia and depression over social media using



- effective machine learning frameworks,” in CLEF, 2018.
18. D. Preotiuc-Pietro, J. C. Eichstaedt, G. J. Park, M. Sap, L. Smith, V. Tobolsky, H. A. Schwartz, and L. H. Ungar, “The role of personality, age, and gender in tweeting about mental illness,” in CLPsych@HLT-NAACL, 2015.
 19. T. Pyszczynski and J. Greenberg, “Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression.” *Psychological bulletin*, vol. 102, no. 1, p. 122, 1987.
 20. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with twitter data,” *Scientific reports*, vol. 7, no. 1, p. 13006, 2017.
 21. P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, “Beyond lda: Exploring supervised topic modeling for depression-related language in twitter,” in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 99–107.
 22. S. W. Stirman and J. W. Pennebaker, “Word use in the poetry of suicidal and nonsuicidal poets,” *Psychosomatic medicine*, vol. 63, no. 4, pp. 517–522, 2001.
 23. H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, “Towards assessing changes in degree of depression through facebook,” in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 118–125.
 24. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing depression from twitter activity,” in CHI, 2015.
 25. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing depression from twitter activity,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 2015, pp. 3187–3196.
 26. Y. Tyshchenko, “Depression and anxiety detection from blog posts data.” University of Tartu Institute of Computer Science Computer Science Curriculum, 2018.
 27. J. Wolohan, M. Hiraga, A. Mukherjee, Z. A. Sayyed, and M. Millard, “Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp,” in *Proceedings of the First International Workshop on Language Cognition and Computational Models*, 2018, pp. 11–21.
 28. W. H. Organization, “Depression and other common mental disorders: Global health estimates. geneva: World health organization; 2017. licence: Cc by-nc-sa 3.0 igo.” <http://www.who.int/en/news-room/factsheets/detail/depression>, 2017.



**Impact Factor:
7.512**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details