

# Sentiment Analysis by Using Social Media Post

RANJANA D SARGAR<sup>1</sup>, DR VIVEK KUMAR SINGH<sup>2</sup>, RESHMA R GULWANI<sup>3</sup>

<sup>1, 2, 3</sup> Department of Information Technology Engineering, Ramrao Adik Institute of Technology, Nerul, India

**Abstract-** *The Internet and online social networks have increased connection between people. Such information is often spread to too many of us in seconds. During this survey the results of variety of machine classifiers like Support vector machine, K nearest Neighbor and Decision tree are used. Aim of classification of text concerning psychological state of user through comments on twitter also as Facebook. These algorithms distinguish between the more worrying content, like sad, happiness, anxiety, anger etc. Feeling examination might be a preparing to sort the look of the person who might be tweets are often named positive, negative or impartial. For case, the tweet "motion picture impressive" could also be a positive substance and therefore the tweet "motion pictures might be most noticeably awful" might be a negative substance. During this research they firstly explained about factors of depression. There are several signs and symptoms from which we will identify depression level of social media user*

**Indexed Terms-** *classification, KNN, SVM, Social media post, sentiment analysis, facebook, twitter*

## I. INTRODUCTION

Social Arrange Destinations (SNS) may be source of information and screening apparatus to classify the clients concurring to client created substance.. By utilizing machine learning algorithms such as Naive Bayes depression level of client is classified into diverse levels and gives doctor’s area close to client area. It appears to be major issue which is the reason it persuade us to worked on it. Prior determination of discouraged understanding were done on premise of surveys and But the result was not so subjective and precise. In differentiate with that, social media is effective instrument for anticipating sadness levels of a person[1].

Social media has given web clients a setting for communicating and sharing their considerations on diverse occasions .Estimation examination may be a strategy of computing and fulfilling a view of a individual given in a chunk of a content, to distinguish people considering approximately any subject Opinion investigation alludes to the utilize of characteristic dialect handling .

## Literature review

In this literature review of machine learning classifiers, first focused is on four sorts of variables such as passionate prepare, transient prepare, linguistic fashion and all (passionate, worldly, etymological fashion) of the features of data collectively for the discovery and editing of depressive information which was received as social media posts.. Diverse execution measurements are utilized to assess distinctive Machine Learning Algorithms. Until further notice, we will concentrate on the ones utilized for Classification issues. Example of metric evaluation of machine learning algorithm is recall and precision. The survey was conducted using MATLAB2016b.This survey based on four major classifiers such as Decision tree, support vector machine, K Nearest Neighbor Classifier. Each of this classifier has sub classifier such as for decision tree simple decision tree, complex decision tree and medium decision tree classifier is used.

KNN classifier has sub classifier as fine, medium, cubic and weighted. And for SVM sub classifier such as linear, quadratic, cubic and fine used. By using above classification technology examined performance of social media comments.

SNS	Posts	Depressed post	Not depressed post
Live journal	2132	758	1374

Twitter	2354	489	1865
Facebook	2287	826	1461
	6773	2073	4700

Table: 1 Post in the supervised dataset for each SNS

Analysis is done on emotional process, linguistic process, temporal process and all features. From classification Decision tree given highest result as compared to KNN. But KNN gives highest precision as compared to Decision tree. Similarly linguistic style decision gives highest result for precision. Evaluation of matrices parameters has used to implement above classifiers and that are based on four different ways.

Those are given below:

- True positive value (TP) = positive value and expected to be positive
- True negative value (NP) = Negative value and anticipated to be negative
- False negative (FN) = values which are positive but expected result is negative.
- False positive (FP) value = values which are negative but expected to be as positive

All of the evaluation of results of all the classifiers precision and recall values are calculated by using following formulas:

Precision is the value which are not more correct but gives the result which approximately similar to the result. Precision can be calculated as the ratio of true positive values of the cases which are anticipated as positive. By this method chosen cases are right.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

Recall is the percentage of relevant data results into correctly classified by using machine algorithm.

$$\text{Recall (R)} = \frac{\text{True positive}}{\text{True positive} + \text{false positive}}$$

F measure calculated as mean of precision and Recall. In this both false positives and false negatives into results.

$$\text{Performance Measures (F)} = 2 \frac{PR}{P+R}$$

feature	Emotional process			All features		
	Pt	Rec all	Fmeasure	Pt	Rec all	Fmeasure
Complex tree	0.57	0.9	0.70	0.57	0.84	0.69
Medium tree	0.56	0.96	0.71	0.57	0.91	0.71
Simple tree	0.56	0.98	0.73	0.57	0.98	0.73
K nearest neighbor Classifier	0.56	0.98	0.83	0.57	0.99	0.68
Fine knn	0.56	0.57	0.58	0.59	0.57	0.57
Coarse KNN	0.59	0.56	0.58	0.59	0.53	0.57
Cosine KNN	0.59	0.88	0.59	0.59	0.60	0.68
Cubic KNN	0.59	0.57	0.60	0.60	0.52	0.60
Weighted KNN	0.57	0.56	0.59	0.58	0.64	0.54
Support vector machine Classifier	0.55	0.86	0.55	0.77	0.65	0.73
Linear SVM	0.57	0.18	0.58	0.57	1	0.72
Quadratic SVM	0.57	0.75	0.59	0.57	0.99	0.72
Cubic SVM	0.56	0.80	0.59	0.57	0.70	0.62
Fine Gaussi	0.57	0.87	0.59	0.58	0.94	0.71

an SVM						
Medium Gaussian SVM	0.57	0.9	0.57	0.57	0.98	0.72
Coarse Gaussian SVM	0.57	1	0.57	0.57	1	0.72

Table 2: performance metrics of emotional process and linguistic styles

feature	Temporal process			All features		
	Pt	Re	Fm	Pt	Re	Fm
Decision tree						
Complex tree	0.57	0.90	0.71	0.58	0.84	0.69
Medium tree	0.57	0.96	0.72	0.57	0.98	0.71
Simple tree	0.57	0.98	0.73	0.57	0.98	0.72
K nearest neighbor	0.73	0.56	0.66	0.57	0.58	0.73
Fine knn	0.56	0.57	0.58	0.58	0.56	0.58
Coarse KNN	0.57	0.56	0.58	0.58	0.53	0.56
Cosine KNN	0.58	0.88	0.58	0.58	0.60	0.66
Cubic KNN	0.58	0.57	0.60	0.60	0.53	0.61
Weighted KNN	0.59	0.56	0.58	0.59	0.63	0.56
Support vector machine	0.57	0.1	0.57	0.60	0.97	0.73
Linear SVM	0.57	0.1	0.57	0.59	1	0.75
Cubic SVM	0.57	0.80	0.57	0.59	0.71	0.63

Fine Gaussian SVM	0.57	0.87	0.58	0.59	0.93	0.73
Medium Gaussian SVM	0.57	0.98	0.58	0.56	0.99	0.72
Coarse Gaussian SVM	0.57	1	0.56	0.59	1	0.72

Table.3: Performance metrics of temporal process and all features

In above results of classifier of SVM, KNN and Decision tree is noted. The above result shows that best performing model is decision tree. For KNN and SVM gives same high precision but the decision tree gives the highest result of f measures and recall.

## II. RELATED WORK

No	Author s	Year	Methodology Used	Results
1	A.M.R amadh ani and H.S Goo[7]	2012	Twitter Sentiment Analysis Using Deep Learning Methods	In this paper research done on sentiment analysis. In this research machine learning algorithm such as deep learning in terms of neural network is used. The result of this research was 75%
2	Milica ciric, Aleksa ndar Stanim irovic, Nikola Petrovi c [6]	2013	Comparison of Different Algorithms for Sentiment Classification	In this research they taken useful data from twitter that is also called tweets. By applying several classification algorithm extracted meaningful data from it.As well as they tried to reduce noise from output. Their research is on tweet from user.

				But from machine learning algorithm we can also classify customers' feedback to company as well as government's policies and decision to citizens.						
3	Rupinder Kaur, Hardeep Singh, Gaurav Gupta[5]	2019	Sentiment analysis on facebook comments using data mining	In this paper the used data mining techniques for sentiment analysis on facebook comments. Large set data set is taken and by using data mining techniques they converted into understandable numerical form						classification error rate. The decision tree gives more precision values than the other machine learning algorithm. Machine learning algorithm techniques identify solutions related to mental state of among facebook users. By using machine learning algorithms detections of major depression among facebook user. In this survey Support vector machine gives excellent result. But for larger data Naive Bayes classifier used.
4	Rastogi et al[6]	2014	Sentiment analysis based on facebook recommendation	A review discussed about offering new friend who have different opinions but similar interest. In this paper developed a Techniques which is user recommended technique based on novel weighting function is implemented. Which not only includes user interest but also their sentiments are considered.						
5	Md. Rafiqul Islam, Ashir Ahmed	2018		In this survey they taken psycholinguistic features. In this they improved accuracy and						

### III. METHODS USED FOR CLASSIFICATION

Classification can be done by using different algorithm which are listed below:

- Support Vector Machine (SVM)
- K-Nearest Neighbor (KNN)
- Decision Tree (DT)

All the algorithms follow some common step for the classification.

- Data collection: In this method raw data is collected from social media and it is analyzed by using software LIWC. It analyses each and each line of data.
- Data preparation: In this survey they taken 21 columns data in which 13 columns are of linguistic styles such as conjunction, nouns, verbs and negation etc. Other 5 columns represents

emotional information and remaining 3 lines represents temporal process.

- **Building Data Set:** In this data from social media comments that were divided into positive, negative and neutral comments are labeled as for positive(a) and for negative comment(b) and for neutral comment(C). In this total 7145 comments were taken for classification out of that 58% comments were positive. And 42% were negative.
- This research gives the results of emotional post, negative post as well as positive post of social media user, And LIWC (Linguistic Inquiry and word count) vocabulary which contains vocabulary based on mental state of user. This first classifies the comments or tweets of user in to numerical value. After the conversion of comments or tweets different classifier are used to compare the result
- **Extraction of features:** To represent the depressive or non-depressive comments of user gives the different views in psycholinguistic processes. This table gives conversion of comments of user into numbers. Which represents the how users are depressed and non-depressed.

Data set	Numerical value
No of comments of Facebook user	7145
Depressed user	4149
Non depressed user	2996

Table 4: Classification of depression-based data

A. Review on classifier:

Today is the universe of new innovations and new technologies. For the most part the work is finished by using the web. Web applications is the new reason for the beginning of learning, shopping and training. The aim of web is to collect the information from different online site. By using different sites reviews of user can be identified easily. And company rating or product of company is increased. Such information gathering is called sentiments or opinion mining.

By the application of different classifier the output gives 70 which highlights the higher level of mental illness of the person. The process of psycholinguistic processes are classified as following:

- Psychological process or internal mental state process — learning abilities such as knowledge, emotion, perception, motivational thoughts, skills, past, present and future, feelings.
- Linguistic process— Punctuation, Clauses, Voice of Speech,
- Grammar—Degree comparison, figure of speech, Adjectives, comparisons, numerical value, quantifiers, Tenses.

In this survey out of 70 factors only 23 and that converted this positive and negative comments into numbers. Given below table explain different classes mental state.

- **Measurement of depressive behavior:** In this five emotional behaviors were considered i.e. positive, negative, sad, anger and anxiety. By using LIWCScale 2015 calculated their values in numerical form [1].

LIWC derived words	Example word
Emotional process Positive negative sadness anger anxiety	Happiness, hope, good Lose, hurt Worried, sadness Stop, hating worry
Temporal process Present Past Future	Today, now Ago, did May, will
Linguistic styles Articles Prepositions Auxiliary verbs Adverb Conjunction negation	a, an For, in, of Do, have etc Quickly, slowly etc And, but etc

A distant better understanding of the common intuition behind discouragement, in this paper, connection between Decision Tree classifier, K nearest neighbor classifier and SVM classifier is shown.

In this survey used classification methods are explained on the basis of phonetic fashion, enthusiastic handle, transient handle and all (Phonetic, enthusiastic and worldly) highlights are able to effectively extricate.

This survey has put the ground for future inventions about on deductions and determination of other data based on circumstances such as discovery of certain feeling or cause, as well as expectation of open supposition based on situation etc. It can be observed that Decision Tree algorithm gives the better output as compared to other classification algorithm. Additionally, in this paper they connected 21 types of traits of LIWC program for recognizing sadness, but they are able to apply more than 54 attributes. Through the various researches and application of different classifier the accuracy was between ranges of 60 to 80. It is difficult to understand who is suffering by depression by only seeing. person. But it becomes easier through the post of social media [2].

### CONCLUSION

In this term paper they have spoken to capability of utilizing Facebook as apparatus to degree or identify the discouragement level among clients. In the investigation they performed investigation on chosen information set. In this investigate they firstly clarified approximately components of misery. There are a few signs and symptoms from which we will distinguish sadness level of client. In this inquire about it dissected 7146 depressive demonstrative comments. Out of that 54,77% depressive clients communicate with their companions from midnight to late morning and 45,22% from late morning to early afternoon To demonstrate this classifier given result between 60% to 80%.LIWC has 50 properties but out of that only 21 are taken for overview for detecting depression level of client. The foremost critical property are found to be within the shape of feelings [2].

### ACKNOWLEDGMENT

In order to explain my sincere appreciation and profound respect for my project guide and PG coordinator, this study is accompanied by excellent direction, tracking and relentless encouragement during the execution of this article. I greatly admire his attempts to develop my professional writing ability. His blessing, aid and encouragement from time to time will take me a long way on the path of life that I am close to embarking on.

### REFERENCES

- [1] Islam, Md Rafiqul & Kabir, Ashad & Ahmed, Ashir & Kamal, Abu & Wang, Hua & Ulhaq, Anwaar. (2018). Depression detection from social network data using machine learning techniques. Health Information Science and Systems. 6. 8. 10.1007/s13755-018-0046-0.
- [2] Arvind and Deva Prasad “Text Recognizer and Document Reader in Various Languages and Accents” 2017I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [3] M M Aldarwish and H.F.Ahmed,” Predicating Depression Levels Using Social Media Post,”2017, IEEE 13<sup>th</sup> International Symposium on Autonomus Decentralized System (ISADS), Bangkok 2017, pp.277-280
- [4] Gamon, Michael & Choudhury, Munmun & Counts, Scott & Horvitz, Eric. (2013). Predicting Depression via Social Media. Association for the Advancement of Artificial Intelligence. Christo Troussa, Maria, Kurt Junshean Espinosa, Kevin liaguno,Jasme cargo sentiment analysis of Facebook statuses using naïve Bayes classifier for language learning 2013.
- [5] Rupinder Kaur , Dr. Harman deep Singh , Dr. Gaurav Gupta, 2019, Sentimental Analysis on Facebook comments using Data Mining Technique, International Journal Of Engineering Research & Technology (IJERT) Volume 08, Issue 08 (August 2019).
- [6] M.Ciric , A.Stanimirovic , N.Petrovic and L.Stoimeniv , “ Comparison of different algorithms for sentiment classification,”2013

11<sup>th</sup> International Conference on Telecommunication in Modern Satellite, cable and Broadcasting Services (TELSIKS), Nis, 2013, pp. 567-570. doi: 10.1109/TELSEKS.2013.6704442

- [7] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," *2017 7th International Annual Engineering Seminar (InAES)*, Yogyakarta, 2017, pp. 14. doi: 10.1109/INAES.2017.8068556
- [8] Varsha Sahayak, Vijaya Shete, Apashabi Pathan, "Sentiment Analysis on Twitter Data", *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, Issue 1, Volume 2, January 2015, pp. 178-183.
- [9] Purva Mestry, Shruti Joshi, Sonal Mehta and Ashwini Save. Article: A Survey on Twitter Sentiment Analysis with Various Algorithms. *IJCA Proceedings on National Conference on Role of Engineers in National Building NCRENB2016* (1): 20-24, July 2016. Pp. 35-38.