**Self-Harm PreventionBased On Social Platforms User Data Using Naive Bayes Classifier**

|  |  |
| --- | --- |
| Rinku Yadav  *Department of Computer Science*  *Chandigarh College of*  *Engineering and Technology*  *rinkuccet@gmail.com* | Varun Gupta\*  *Department of Computer Science*  *Chandigarh College of*  *Engineering and Technology*  *varun3dec@gmail.com* |

*\*Corresponding Author*

**Abstract**

With the increase of the Internet i.e. World Wide Web ,the social networking sites such as facebook , twitter , instagram , google+ are also in bloom, this social networking sites not only used by the youth but also been used by the analyst to analysis the need ,emotion, feeling ,comments of the user over the network where the user comments over the feed of other people or post something that directly states there state of mind and are widely used in emotion AI but actually the analysis of the comments are used by analyst to post advertisement over the homepage of user or make suggestion of the products, by analysis the like , dislikes of the person the mental health of the person can be predicted , but recently studied shows that analyst are using the post analysis of user to make benefit of the political parties in the U.S. to predict the favourable political party by the majority

In this paper, we use naive Bayes classifier for detecting user tweets related to self-harm on Twitter for prevention of self-harm tendencies of the user .The results obtained from the work are promising and can be quite helpful in the development of a system that can be used for prevention of self-harm tendencies of the mankind using data retrieved from social platforms.

***Keywords****: Self-harm prevention, Opinion mining, sentiment analysis, emotional AI, naive Bayes classifier ,machine learning, and artificial intelligence*

**1. INTRODUCTION**

Emotion AI can be divided into facial detection, speech recognition and text analysis, where the user is involved in making some actions and these actions are analysed to make use of machine intelligence and perform some automated task i.e, prediction of like , dislike of the people or make prediction of mental health of the people ,as the human make non-verbal cues such as facial expressions, gesture, body language and tone of voice, to communicate their emotions, but in these paper our intention is to use the text sentimental analysis of the tweets made by the person over the social platform twitter.

Also the sentimental analysis is used to make analysis of the attitude of writer with respect to some topic or the polarity and subjectivity towards some topic

In this paper, we are using sentiment analysis of twitter user data for detection of self-harm tendencies being shown by the users in their tweets. For the purpose, naive Byes classifier is being used for the classification of tweets into words showing positive attitude of the user and negative attitude of the user. The detection of words in user tweets depicting negative attitude or self-harm attitude can be quite helpful in timely counselling of the concerned person which can lead us to the prevention of self-harm tendencies of the person.

**2. RELATED WORK**

Existing approaches [1-10] to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches.

**Knowledge-based techniques** or simply the fact based which make prediction on the given facts using some logical operation on the given test text ,which includes the prior knowledge of the text used in various moods of mental health like in positive mood generally used words are happy, rejoiced and other synonyms while in negative mood generally used words are sad , afraid and bored though these words are generally used to train the model and the model is simply a mapped model which model the given text into the data stored to make prediction. .

**Statistical methods**  or simply the distance metrics used by the models such as SVM or the probabilistic classifier such as Bayes theorem where the features are extracted from the given text in training phase and prediction is done testing phase and this includes the “bags of words” plus the training model through which the various statistical metric such as word vector and various measure are used to define the accuracy of the prediction.

**Hybrid approaches** which include both the knowledge based method for manual training of the new words existing or the abbreviations or Emoji’s used by the people recently over some action and express there feelings .

Also the word order and grammatical mistakes can be taken into account for the user’s mind set mental health, recently studied shows that various abbreviation are also used by the user for the post in their timelines of social networks.

**3. DATA USED IN THE WORK**

A dataset of 200 X 3 plus entries are fetched from the real-time user timelines and home timelines using the Twitter API and saved as username which is publicly known in the social platform,tweet as text column, label as the positive,negative and neutral.

Re-tweets are labelled as neutral else text are labelled according to the sentiment. Formats of the dataset used are JSON and CSV as shown below:

**JSON FORMAT**

[

{"text": "I love sandwich.", "label": "pos"},

{"text": "This is an Exciting!", "label": "pos"},

{"text": "I do not like anyone", "label": "neg"}

]

**CSV FORMAT**

(I love to eat apple, pos)

(This is going to be good direction, pos)

(All in vein not helpful ,neg)

(A bad step , neg)

(No wait at all , neg)

**4. MACHINE CLASSIFICATION METHOD**

**4. 1 Methodology**

We use the collection of positive and negative words as JSON or CSV file with the word as text and response as the label(pos/neg) based on that we classify the given tweet by the user in his timeline. We use the Twitter API for fetching the real-time tweets and classify them using naive Bayes classifier in a textblob library, which is trained on the following words.

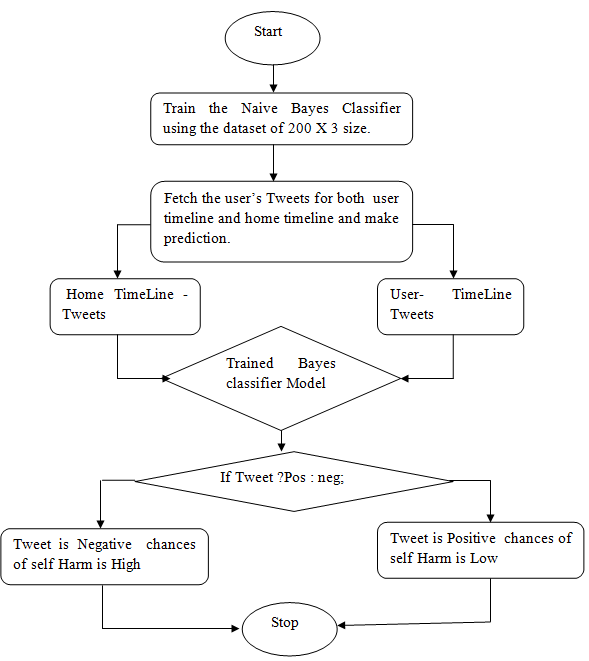
**Positive words:-**

alacrity, altrucause, amiable, astounding, attractive, alive – aliveness, acclaim, abundant gratification, acclamation, accomplished, accomplishments, accurate, accurately, achievable, achievements, action for happiness, active and constructive steps, acts of kindness, adaptable, adaptive, admiration, direction, delicate, decent, desirable,...

**Negative words:**

I’m just tired. I just want to be done.I just want to sleep.I can’t keep doing this.I just want to be alone.I want to go home.If anything happens to me, promise to take care of…I’m just stressed out.I’m having a hard time.No one cares.I don’t care.What will heaven be like?I should just kill myself.I can’t imagine living the rest of my life like this.I feel so much better.You know I love you, right?I want to disappear.I want to tell you something. Oh, never mind.I don’t know.I’m not feeling good.I don’t think I’ll be at school next week,Why, Why, Woe, Women, Wrists, Wrong…..suicidal; suicide; kill myself; my suicide note; my suicide letter; end my life; never wake up; can't go on; not worth living; ready to jump; sleep forever; want to die; be dead; better off without me; better off dead; suicide plan; suicide pact; tired of living; don't want to be here; die alone; go to sleep forever

**4.2 Flow Chart**



**4.2 Training**

Train variable contains the text with the label containing 100 of words for both positive and negative labels.

Cl : be the classifier variable.

Prediction : variable for text analysis

*from textblob.classifiers import NaiveBayesClassifier*

*cl  = NaiveBayesClassifier(train)*

**4.3 Testing**

Given the example strings which are used to classify based on the training done on the given positive and negative words.

*prediction   = cl.classify("all things are waste!!!")*

**4.4 Classifier**

In this paper we used the naive bayes classifer to make prediction of the sentiment of the Advantages of Naive Bayes model are as follows:-

* Naive bayes classifier are able to perform best when the case of independence of features come into picture as compare to logistic regression.
* Naive bayes classifier are used for multi-classification and are easy to perform.
* This classifier is able to extract the independent features of an objects
* Performance of the naive Bayes classifier is best in case of categorical variable but in case of numerical variable normalization is needed.

Disadvantages of Naive Bayes model are as follows:-

* If the category of the categorical variable is different from the training dataset then it is called “Zero Frequency” as it need more training of the dataset for this we use various smoothing techniques such as Laplace estimation.
* Naive bayes classifier is considered as bad estimator in case of logistic regression problems.
* Naive Bayes works on the independent features as a working rule but in real life it is not possible as all the features are interconnected to some context.

**5. RESULTS AND EVALUATION**

Prediction done by the naive Bayes classifier is by using the classifier in text-blob and accuracy is evaluated based on the prediction.The human judgement need to done via his action and this is true after doing the sentimental analysis we perform the tweeter analysis.Tweepy python library is used to fetch the user’s both hometimeline and user-timeline tweets which further analysed.

*prediction=cl.classify("all things are waste!!!")*

*print(prediction)*

*Output:- ‘neg’*

Figure 1 shows the user profile data fetched using the twitter API, secret key of the user as follows.

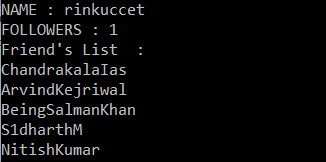


Figure 1 User Profile Data of twitter

Figure 2 shows the user timeline tweets which are classified as negative tweet by the model as follows

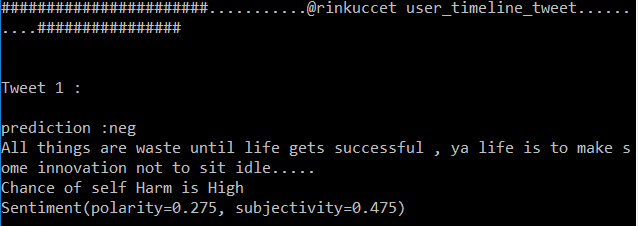


Figure 2 User TimeLine Negative Tweets

Figure 3 shows the user timeline tweets which are classified as Positive tweet by the model as follows.

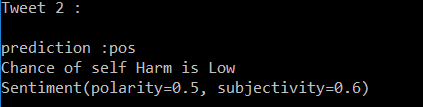


Figure 3 User TimeLine Positive Tweets

Figure 4 shows the user ‘s home timeline tweets are classified as negative by the model as follows

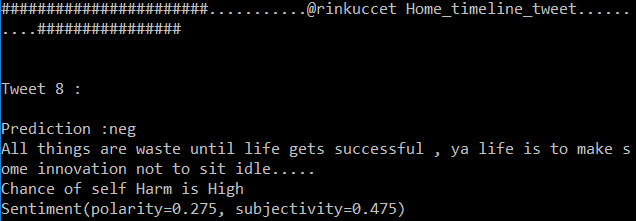


Figure 4 Home- TimeLine Negative Tweets

Figure 5 shows the accuracy of the naive Bayes model classifier and shows the both training accuracy and testing accuracy of the model.

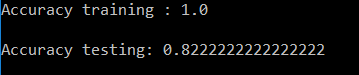


Figure 5 Accuracy of the Naive Baye’s model

We obtained the perfect accuracy on training the labelled dataset which is quiet often as all the dataset is labelled while training but in testing we used to hide the labels so as to make prediction of labels based on the given text we are able to sentimental analysis of the tweets of the user in both user timeline and home timeline

Labelling of the dataset are as ‘pos’ for positive tweets and ‘neg’ for negative tweets as shown in figures above. Both the user timeline and home timelines are predicted and analysed via this method.

**REFERENCES**

[1] A. Abboute, Y. Boudjeriou, G. Entringer, J. Aze, S. Bringay, and P. Poncelet. Mining Twitter for suicide prevention. In Natural Language Processing and Information Systems, Lecture Notes in Computer Science, Vol. 8455,pages 250–253. 2014.

[2] D. Baker and S. Fortune. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 29(3):118–122, 2008.

[3] L. Barbosa and J. Feng. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44. 2010.

[4] K. Becker and M. H. Schmidt. When kids seek help online: Internet chat rooms and suicide. Reclaim Child Youth.13(4):229–230, 2005.

[5] Biddle, Lucy & Donovan, Jenny & Hawton, Keith & Kapur, Navneet & Gunnell, David. (2008). Suicide and the Internet. BMJ (Clinical research ed.). 336. 800-2. 10.1136/bmj.39525.442674.AD.

[6] L. Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996.

[7] P. Burnap, O. F. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, J. Morgan, and L. Sloan. Detecting tension in online communities with computational twitter analysis. Technological Forecasting and Social Change, 2013.

[8] M. D. C. S. Counts and M. Gamon. Not all moods recreated equal!a exploring human emotional states in social media. 2012.

[9] K. Daine, K. Hawton, V. Singaravelu, A. Stewart, S. Simkin, and P. Montgomery. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PloS one, 8(10):e77555, 2013.

[10] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff. Characterizing and predicting postpartumdepression from shared facebook data.In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, Pages 626-638, Baltimore, Maryland, USA. February 15 - 19, 2014.