**Self Harm Prevention Using Naive Bayes Classifier**

**on Real Users Timeline in Social Platform.**

*Rinku Yadav Dr.Varun Gupta*

*BE(CSE) 4th yr Student Associate professor*

*CSE Department ,CCET CSE Department ,CCET*

**ABSTRACT**

The World Wide Web, and online social networks in particular, have increased connectivity between people such that information can spread to millions of people in a matter of minutes. This form of online collective contagion has provided many benefits to society, such as providing reassurance and emergency management in the immediate aftermath of natural disasters. However, it also poses a potential risk to vulnerable Web users who receive this information and could subsequently come to harm. One example of this would be the spread of suicidal ideation in online social networks, about which concerns have been raised.

In this paper we report the results of a number of machine classifiers built with the aim of classifying text relating to suicide on Twitter. The classifier distinguishes between the more worrying content, such as suicidal ideation, and other suicide-related topics such as reporting of a suicide, memorial, campaigning and support.

***Keywords****:  twitter api ,positive words, negative words, suicidal, classification.*

**1. INTRODUCTION**

In this paper we are showing how to classify the user tweets using sentiment analysis of twitter

Opinion mining (sometimes known as sentiment analysis or emotion AI) refers to the use of [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [text analysis](https://en.wikipedia.org/wiki/Text_analytics), [computational linguistics](https://en.wikipedia.org/wiki/Computational_linguistics), and [biometrics](https://en.wikipedia.org/wiki/Biometrics)to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to [voice of the customer](https://en.wikipedia.org/wiki/Voice_of_the_customer) materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from [marketing](https://en.wikipedia.org/wiki/Marketing) to [customer service](https://en.wikipedia.org/wiki/Customer_relationship_management) to clinical medicine.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation (see [appraisal theory](https://en.wikipedia.org/wiki/Appraisal_theory)), affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

**2. RELATED WORK**

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

**Knowledge-based techniques** classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored.Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions .

**Statistical methods** leverage on elements from [machine learning](https://en.wikipedia.org/wiki/Machine_learning) such as [latent semantic analysis](https://en.wikipedia.org/wiki/Latent_semantic_analysis), [support vector machines](https://en.wikipedia.org/wiki/Support_vector_machines), "[bag of words](https://en.wikipedia.org/wiki/Bag_of_words)" . More sophisticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the effect is felt).To mine the opinion in context and get the feature about which the speaker has opined, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text.

**Hybrid approaches** leverage on both machine learning and elements from [knowledge representation](https://en.wikipedia.org/wiki/Knowledge_representation) such as [ontologies](https://en.wikipedia.org/wiki/Ontologies) and [semantic networks](https://en.wikipedia.org/wiki/Semantic_network) in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so.

Open source software tools deploy [machine learning](https://en.wikipedia.org/wiki/Machine_learning), statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media.Knowledge-based systems, on the other hand, make use of publicly available resources, to extract the semantic and affective information associated with natural language concepts. Sentiment analysis can also be performed on visual content, i.e., images and videos. One of the first approach in this direction is SentiBank utilizing an adjective noun pair representation of visual content. In addition, the vast majority of sentiment classification approaches rely on the bag-of-words model, which disregards context, grammar and even word order. Approaches that analyses the sentiment based on how words compose the meaning of longer phrases have shown better result, but they incur an additional annotation overhead.

**3. DATA**

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 user name which is publicly known in social platform ,tweet as text column, label as the positive ,negative and neutral.

Re-tweet are labelled as neutral else text are labelled according to the sentiment.

JSON FORMAT

[

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

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

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

]

CSV FORMAT

I love this sandwich.,pos

This is an amazing place!,pos

I do not like this restaurant,neg

**4. MACHINE CLASSIFICATION METHOD**

**4. 1  Methodology**

We use the collection of positive and negative words as JSON or CSV file with word as text and response as label(pos/neg),based on that we classify the given tweet by the user in his timeline.

Actually we use the twitter api for fetching the real time tweet and classify them using naive bayes classifier in a textblob library,which train on the following words.

**4.1.1 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, adequate, admirably, admiration, 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, adequate, admirably, admiration, direction, delicate, decent, desirable,...

**4.1.2 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 Training :**

Train :variable contain the text with label contains 100s of word 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 string 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:**

It is a classification technique based on [Bayes’ Theorem](https://en.wikipedia.org/wiki/Bayes%27_theorem) with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as ‘Naive’.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

**4.4.1 Pros:**

* It is easy and fast to predict class of test data set. It also perform well in multi class prediction
* When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.
* It perform well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

**4.4.2 Cons:**

* If categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as “Zero Frequency”. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
* On the other side naive Bayes is also known as a bad estimator, so the probability outputs from predict\_proba are not to be taken too seriously.
* Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

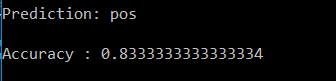
**5. RESULT AND EVALUATION**

Prediction done by the naive bayes classifier is using the classifier in textblob and accuracy is evaluated based on the prediction.The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by variant measures based on [precision and recall over](https://en.wikipedia.org/wiki/Precision_and_recall) the two target categories of negative and positive texts. However, according to research human raters typically only agree about 80% of the time (see [Inter-rater reliability](https://en.wikipedia.org/wiki/Inter-rater_reliability)). Thus, a program which achieves 70% accuracy in classifying sentiment is doing nearly as well as humans, even though such accuracy may not sound impressive.

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

print(prediction)

‘neg’



If a program were "right" 100% of the time, humans would still disagree with it about 20% of the time, since they disagree that much about *any* answer.

On the other hand, computer systems will make very different errors than human assessors, and thus the figures are not entirely comparable.

**6.  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, volume 8455 of Lecture Notes in Computer Science, pages 250–253. Springer, 2014.

[2] D. Baker and S. Fortune. Understanding self-harm and suicide websites. 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. Association for Computational Linguistics, 2010.

[4] K. Becker and M. H. Schmidt. When kids seek help on-line: Internet chat rooms and suicide. reclaiming children and youth, 13(4):229–230, 2005.

[5] L. Biddle, J. Donovan, K. Hawton, N. Kapur, and D. Gunnell. Suicide and the internet. Bmj, 336(7648):800–802, 2008.

[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 re created 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 postpartum