Roshini

Ramakrishna Prasad Koganti

2202-CSE-5301-003

05/09/2020

Estimation of the depression level using social media posts

ABSTRACT

The utilization of Social Network Sites (SNS) is expanding these days, particularly by the younger generation. Using the SNS the people post their feelings, inclinations, interests, and emotions on a daily basis. Many kinds of research have found that using the user-generated content (UGC), it is possible to estimate the mental condition of a person. Depression is the most serious illness which affects sleep, work, eat and having fun. However from the person's social media account by analyzing all the posts we can estimate the person's mood, negativism and depression levels. In this project, our main aim is to analyze the SNS posts of the users and estimate one's mental levels and state of mind. In this project we are taking the dataset(tweets) from the twitter account and estimate the depression levels of the user. We use the techniques like the Naïve Bayesian, DTree, SVM for text mining and plot the confusion matrix, accuracy of each technique.

Keywords: Depression, SNS, word clustering, mining

INTRODUCTION

Mental illness is the most common problem in this world, this is due to the psychological problem like depression. 60% of the world population experiences depression once in their life time. Untreated depression may cause the chance for dangerous behavior. The depression levels may vary based on the factors like age, sex etc. In order to obtain the accurate data we get the data from the social media. Nowadays, social media is a great platform where people can express their opinions. Social media generates millions of data from so many active users and community in entire community, it changes the interaction between the human beings. The social media data is more wide, rich and easily accessible than the normal traditional data. However, in order to investigate this fast growing data we need to use the advanced technologies. The advanced technologies include social media analysis, Data mining, machine learning, Natural Language Processing (NLP). The goal of our project is to find the depressive symptom of the people by analyzing the data from the twitter by using the data mining techniques like Naïve Bayes, Decision Tree, Support Vector Machine. The next section of this report will be taking of Problem, Simulation, or Testing, Results, Conclusions, References.

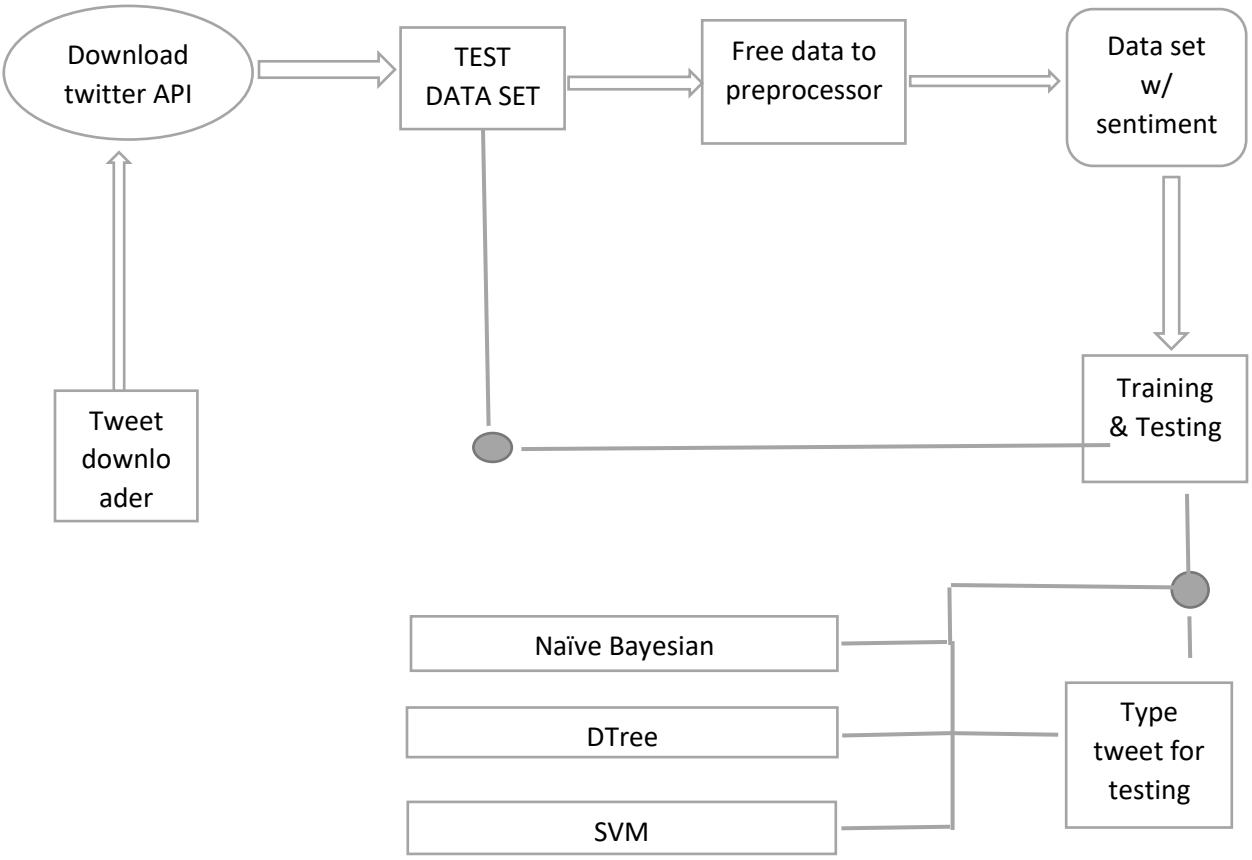
PROBLEM

Twitter produces on average 500 million tweets per day by all the active users in twitter. Twitter helps us to reach the broad audience and connect with people. On the other end, it is harder to detect the negative content, it might end up in unwanted chaos. This is one of the reasons why social listening i.e., monitoring conversation and the feedbacks or comments in social media has become a crucial process in almost all the social media platforms.

The thoughts and views of people are monitored by the help of tweets. And also detect the depression level of the person, based on the tweets made by that person.

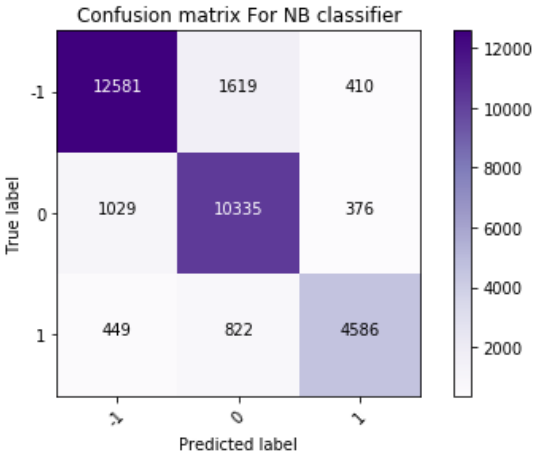
SIMULATION AND TESTING

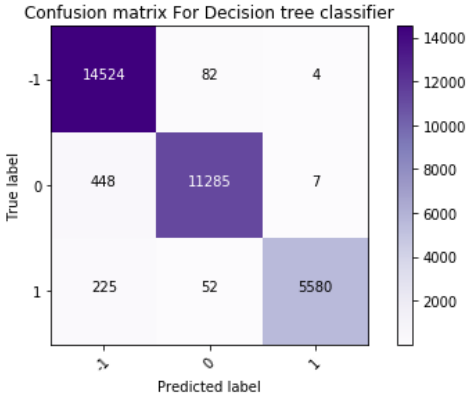
First, we should create a twitter developers account we need four key parameters for it like consumer_key, consumer_secret, access_token, access_secret. After establishing the developer account we then download the stream of tweets of a user by applying the negative keywords filters to the tweet, so we can obtain the required dataset and proceed to the preprocessing stage. In this stage, we will find the polarity of the tweet using the dictionary file. In the dictionary file, we will have the classification of the words into either positive or negative. So in the preprocessing stage for every tweet, we will be adding the polarity of each word divided by the total number of words, which helps us to find out the polarity of the tweet. After the preprocessing is done we are going to insert the sentiment of each tweet in our file estimating whether the tweet is positive or negative. Now we have the proper twitter dataset and its corresponding sentiment filtered by the positive, negative or neutral words. Now we are going to perform depression sentimental analysis on our obtained dataset using the techniques mentioned above. Now we briefly describe each technique we are using in our project. Naive Bayes works on the assumption that the features of the dataset are independent of each other—hence called Naively. It works well for bag-of-words models as in text documents since words in a text document are independent of each other and the location of one word doesn't depend on the location of another word. Hence, it satisfies the independence assumption of the Naive Bayes model. Thus Naïve Bayes is the most commonly used model for text classification, sentiment analysis, spam filtering & recommendation systems. A decision tree is a selection support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only consists of conditional control statements. Decision trees are commonly used in operations research, mainly in decision analysis, to help identify an approach most likely to reach a goal, however, they are additionally a famous tool in machine learning. The paths from the root to the leaf symbolize classification rules. A decision tree consists of three sorts of nodes Decision nodes – commonly represented by squares, Chance nodes – normally represented via circles, End nodes – commonly represented via triangles. We will be using the other techniques like SVM for the depression sentimental analysis and get the outcome whether the user is making positive tweets or negative tweets and determining the mental condition of the user.

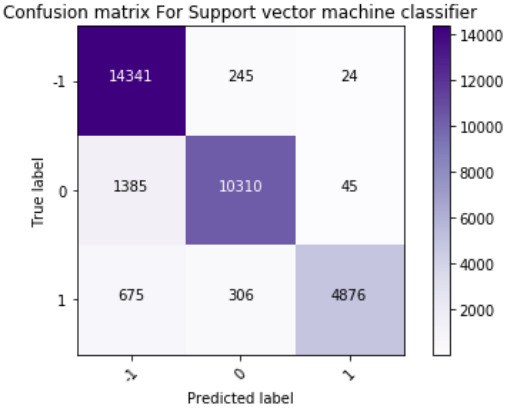


RESULTS

The results we obtained are the confusion matric and accuracy, completion time of Naïve Bayes, Decision Tree and SVM








```
In [7]: runfile('C:/Users/tadir/Downloads/Classification-of-
depression_sentiment_analysis.py', wdir='C:/Users/tadir/Down

Naive Bayes  Accuracy :
 89.54384429147692 %
 Completion Speed 2.06055

Decision tree Accuracy :
 97.03807021164444 %
 Completion Speed 19.23094

Support vector machine Accuracy :
 90.74528896965275 %
 Completion Speed 618.94205

Input your tweet :

Iam so sad

*****************
Negative
*****************

In [8]:
```

CONCLUSION

In this paper, I present the Analysis of twitter data using 3 models: Naïve bayes, Decision tree, Support Vector Machine. The performance factors like accuracy, completion time have been tested by running the resultant data set obtained from the preprocessing the tweets and the results from the models are presented using the confusion matrix.

Among the three models we got Decision tree results more accurate than the Naïve Bayes and Support Vector Machine. The completion time is less for the Naïve Bayes model when compared to the Decision tree and Support Vector Machine. We are able judge the polarity of the tweet as positive, negative, neutral.

REFERENCES

[1] Knigge, Marlene & Kohl, Christopher & Koleva, Galina & Kienegger, Harald & Krcmar, Helmut. (2018). Sentiment Analysis on Twitter Data Using R Algorithms

[2] H. Kaur, V. Mangat and Nidhi, "A survey of sentiment analysis techniques," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 921-925.

[3] S. A. El Rahman, F. A. AlOtaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-4.

[4] Griffin, J. M., Fuhrer, R., Stansfeld, S. A., Marmot, M.: The importance of low control at work and home on depression and anxiety: do these effects vary by gender and social class?. Social science and medicine, vol. 54(5), pp. 783--798. (2002)

[5] Ma, L., Zhang, Y.: Using Word2Vec to process big text data. In Big Data (Big Data), 2015 IEEE International Conference on, pp. 2895--2897. (2015)

[6] Moin Nadeem1 , Mike Horn. 1 , Glen Coppersmith2 , Johns Hopkins University and Dr. Sandip Sen3 , PhD, University of Tulsa, Identifying depression on twitter