

MOOD UPLIFTER

A Project Work Report

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

ARTIFICIAL INTELLIGENCE

AND MACHINE LEARNING

Submitted by:

Kirti Rattan(18BCS6013)

Under the Supervision of:

Mr. Digvijay Puri



**CHANDIGARH
UNIVERSITY**

Discover. Learn. Empower.

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI -
140413, PUNJAB**

March 2020

DECLARATION

I, **Kirti Rattan** student of '**Bachelor of Engineering in Computer Science and Engineering, session:2018-2022**', Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled '**Mood Uplifter**' is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Kirti Rattan
(18BCS6013)

Date: March 8th,2021

Table of Contents

1	Title Page	1
2	Abstract	3
3	Introduction	5
4	Literature Review	6
5	Problem Formulation	7
6	Research Objectives	8
7	Methodology	9-10
8	Result	11
9	Conclusion and Future Scope	12
10	References	13

ABSTRACT

In social media, data is present in huge amounts. Extracting information from social media helps us in many ways and in different fields. Twitter is an online micro-blogging and social-networking platform on which users can write short status updates which are also known as tweets, the length of those tweets is of maximum length of 140 characters. It has a user base of over 330 million active users and half of them log in on twitter on a daily basis, posting nearly 250 million tweets per day. Due to the data present in such a huge amount usage we, in this project have hoped to achieve a reflection of the sentiment of the people around the globe by analyzing their sentiments being expressed in the tweets. Sentiment Analysis is important for many applications such as firms which try to find out the real response of their products in the market, how much value their product actually has in people's lives, we can also predict political elections and stock exchange too.

This report briefly explains my project “Mood Uplifter” using machine Learning. This project is at its ground level right now, but serves our current objective of which is to try to analyze the emotions of people through social media and recommend them accordingly.

The aim of this project is to develop a classification model which can predict the sentiments of a user by analyzing the tweet This project analyzes the tweets of the users and then recommends them to calm down their bad mood with a decaf and celebrate the good mood with some good meal in a nearby restaurant, which will be suggested by the model too.

1. INTRODUCTION:

This project deals with analyzing the tweets of different users and applying various models on it to classify the state of tweet, whether it is positive or negative and then based on the state of their tweet, it will recommend them, i.e., if the state of a user's tweet is negative, it will recommend the user to calm down and have a decaf, and, if the state of the user's tweet is positive, it will recommend the user to have his/her favorite meal, by doing this, my aim is to develop a healthy environment for the people across the globe, that is why the name "**Mood Uplifter**", because it will be used to motivate the ones who are doing good and cheer the ones who are not feeling good and appreciating them for the efforts they put in and the courage with which they have said it to the whole world, as twitter is a global platform with 310 million active users.

1.1 Problem definition:

Sentiment analysis is basically a technique of natural language processing to classify words, analyzing them, train our model to predict the sentiments of the people through the text. Sentiment analysis has been introduced a long time ago, it has done wonders in various fields and is still evolving. In the field of social media, predicting the emotions of people through text can only be done by using twitter, as it is the sole platform with a user active base of 310 million active users, and it is a platform for the words like a micro-blogging site, which makes it easier to be worked upon by sentiment analysis as other social platform are not that much text oriented, lot of research has been done in this field too, but there is still a lot of room for further research in this field, not only on twitter this work of sentiment analysis has been done on blogs too, but blogs are different from twitter as you can write more

than a tweet in a blog, i.e., a twitter has a word limit of 140 words per tweet which forces the user to express his/her opinion compressed in a very short and precise manner. The best results which have been achieved for sentiment analysis are by using various classification models such as supervised learning models which include Naïve Bayes and Support Vector Machines (SVM), but the manual labelling required for the supervised learning approach is quite expensive. A lot of work has been done in this field by using various machine learning models such as unsupervised and semi-supervised models, but still there is a lot of room for more research. The testing and training of various new features and classification techniques sometimes just compare their results to satisfy the base-line performance. What actually has been done is analysis of data for the use of the firms and other companies to improve their product either, or the predication of stock market or other things, means the data has been taken from people but the people aren't returned with anything.

1.2 Project Overview:

Sentiment analysis, which is also known as opinion mining, is basically a type of classification system of natural language processing in which classification of words is performed by determining the state of the given text. Using the machine learning models and natural language processing techniques, we can extract the information from the document by determining its state whether it is positive or negative. It is very useful analysis as we can determine the overall opinion from the text, it can be of any type, about a product selling, advertisement, general blogging, stock market prediction or anything like that and we can determine the opinion of various people, i.e., if they think positive about it, increase in stock market and the list goes on.

Sentiment analysis is actually far to be solved 100% accurately as human language and humans are very complex (objectivity/subjectivity, negation, vocabulary, grammar,) but this is also one of the reasons why it is very interesting to work on. In this project I have choose to try to classify the tweets from Twitter into two states “positive” and “negative” sentiment by building a model based on probabilities. Twitter is a platform for the words like a micro-blogging site, which makes it easier to be worked upon by sentiment analysis as other social platform are not that much text oriented by restricting the user to a limit of 140 characters. A registered user on twitter can tag another registered user by using the sign '@' directly and participate in a topic by using the “#” symbol in the tweet, due to this huge usage of Twitter, and it being a text-oriented platform makes it a perfect source for data extraction and data analysis to determine the opinion of the people across the globe about anything.

This project works upon the idea of Sentimental analysis in which the tweets tweeted by the users will be analyzed, processed and classified on the basis of the sentiments they show, i.e., sad or happy. The testing and training of various new features and classification techniques sometimes just compare their results in order to satisfy the base-line performance and select the most efficient classification techniques to be used for the particular applications and then a recommendation system will be built to recommend the user to have a decaf if he/she is sad or to have a good meal if he/she is happy.

1.3 Hardware Specifications:

A computer with:

- Windows 7 or 10 OS, or
- Mac OS X 10.00 or higher, 64-bit OS, or
- Linux: RHEL 6/7, 64-bit OS, or
- X86 64-bit CPU (Intel/AMD Arch.)
- Minimum 8 GB RAM
- 8 GB Free Disk Space

1.3 Software Specifications:

- Python environment 3.6.1 or higher
- Jupyter Notebook
- Web Browser
- Text viewer
- Notepad
- Spreadsheet
- Anaconda
- Twitter Data set or Twitter Developer Account
- Various python modules

2. LITERATURE REVIEW

Existing systems have used Naive Bayes theorem, features like Ngram and POS-tags, Maximum Entropy model, SVM, Maxent, bag of words and many other models to increase the classification accuracy of a tweet being treated as positive or negative. The basic problem is the accuracy of these models is good but not at par and fully reliable. The project aims at solving this problem by taking into account five models and comparing their accuracy. The goal of this project is to feed a processed dataset which is well labelled and runs a series of functions to identify the tweet. Another aspect of the project is to classify the state of a tweet as based on the labelled dataset and keywords as -1,1.

2.1 Existing System:

In recent years a lot of work has been done in the field of “Sentiment Analysis on Twitter” by number of researchers.

Pak and Paroubek (2010) [1] proposed a model to classify the tweets as objective, positive and negative. They created a twitter corpus by collecting tweets using Twitter API and automatically annotating those tweets using emoticons. Using that corpus, they developed a sentiment classifier based on the multinomial Naive Bayes method that uses features like Ngram and POS-tags. The training set they used was less efficient since it contains only tweets having emoticons.

Parikh and Movassate (2009) [2] implemented two models, a Naive Bayes bigram model and a Maximum Entropy model to classify tweets. They found that the Naive Bayes classifiers worked much better than the Maximum Entropy model.

Go and L. Huang (2009) [3] proposed a solution for sentiment analysis for twitter data by using distant supervision, in which their training data consisted of tweets with emoticons which served as noisy labels. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM). Their feature space consisted of unigrams, bigrams and POS. They concluded that SVM outperformed other models and that unigram were more effective as features.

Turney et al [4] used bag-of-words method for sentiment analysis in which the relationships between words were not at all considered and a document is represented as just a collection of words. To determine the sentiment for the

whole document, sentiments of every word was determined and those values are united with some aggregation functions.

2.2 Proposed System:

This project works upon the idea of Sentimental analysis in which the tweets tweeted by the users will be analyzed, processed and differentiated on the basis of the sentiments they show, i.e., sad or happy by classifying the state of a tweet as based on the labelled dataset ad keywords as -1,1 by using proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications and then a recommendation system will be built to recommend the user to have a decaf if he/she is sad or to have a good meal if he/she is happy.

3. PROBLEM FORMULATION:

In the existing systems, we can say that there is one common thing that all of them are majorly based upon one algorithm although the most efficient one but still not perfect. As we all know that the decisions are more efficient when taken by a hundred minds rather than one. Similarly, there may be algorithms which can make the identification correct but also many other can make it wrong and we are not sure of which one is correct and which one is wrong. So, it would be more efficient to make the decisions out of different algorithms and make the decision out of their mean results. This could not only make it more efficient but also a dynamic method for identification out of every aspect and covering up the limitations of each other.

This project works upon this idea of Sentimental analysis in which the tweets tweeted by the users will be analyzed, processed and differentiated on the basis of the sentiments they show, i.e., sad or happy by classifying the state of a tweet as based on the labelled dataset and keywords as -1,1 by using proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications and then a recommendation system will be built to recommend the user to have a decaf if he/she is sad or to have a good meal if he/she is happy.

4. RESEARCH OBJECTIVES:

The proposed research is aimed to carry out work leading to the development of the analysis of sentiments of users through their tweets, by analyzing them and recommending them depending upon their tweets. The proposed aim will be achieved using following objectives:

- Analyzing the updated dataset of tweets from Twitter.
- Using the traditional methods and the proposed method to check the effectiveness.
- Compare them.
- Proposed method will be able to give more precise results.
- The analyzed tweets will be processed to recommend the user a few things with a thought to uplift his/her mood.

Advantages:

- Output generated on the basis of updated twitter API.
- Use of various up to date modules in python.
- Relatively the accuracy is at par.

Limitations:

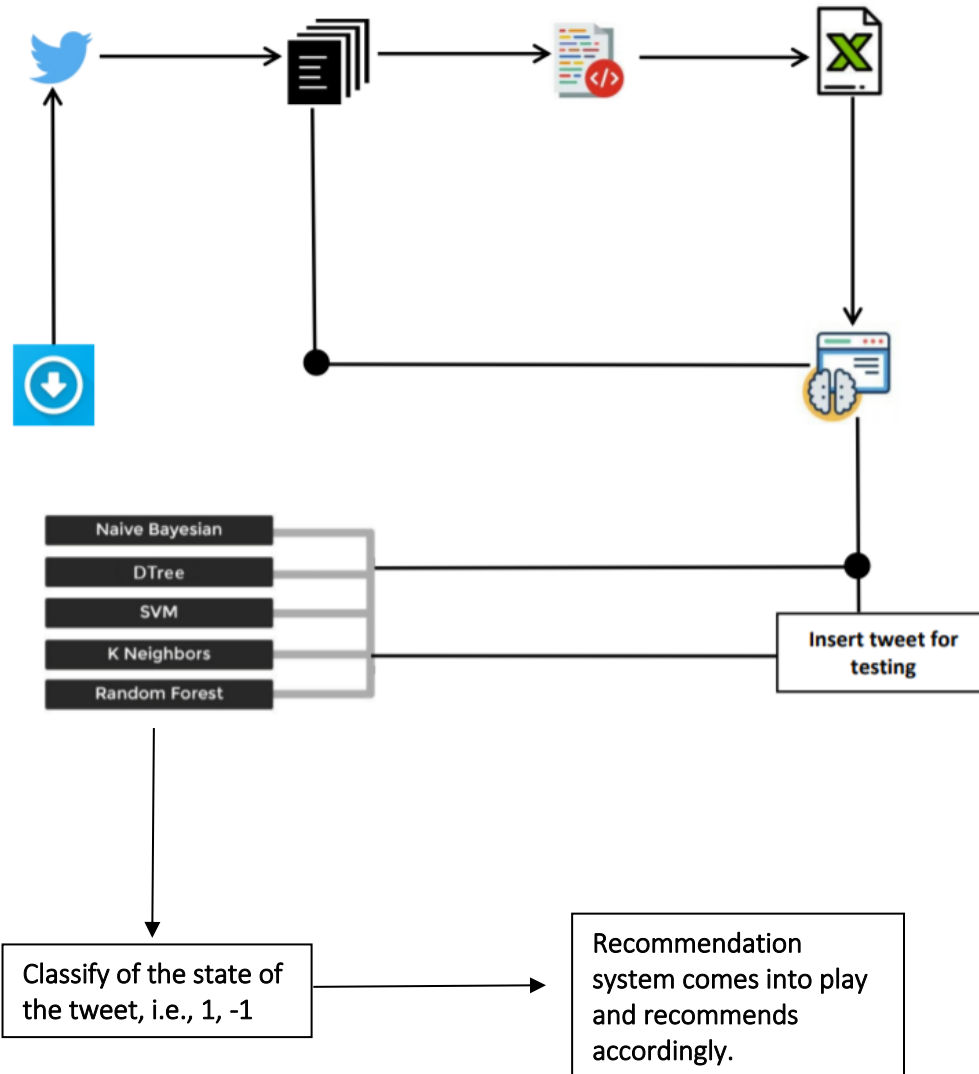
- The computation time is large.
- Language is still a subjective matter, what might be positive to one may be negative to the other.

Comparison done on the basis of 5 important models:

- Naive Bayes.
- Decision tree.
- Support vector machine.
- K-neighboursclassifier.
- Random Forest.

5.METHODOLOGY:

The Flowchart:



The following methodology will be followed to achieve the objectives defined for proposed research work:

- Detailed study of datasets and algorithms.
- Installation and hand on experience on existing approaches of depression sentiment analysis.
- Various parameters such as emotions, accuracy, stop words will be identified to evaluate the proposed system.

- The first step will be to create a Twitter development account and extract the following:
 consumer key = "
 consumer secret = "
 access token = "
 access secret = "
- Download the dataset of tweets from twitter with the help of specific keywords such as 'happiness', 'excitement', 'sad', 'gloomy', etc.
- Then process the data to make it fit better in the model.
- Pre-process the data from the dictionary and data set, the dictionary contains the polarity of the words, each word is tokenized and given a label, -1,1.
- Each tweet consists of the summarization of all the polarities of each word and divided by the number of words in that particular tweet.
- Insert tweet for testing.
- It shall look like this after extracting the data and processing:

```
runall()

#print("Input your tweet : ")
#inputtweet = input()
#datreeINPUT(inputtweet)
```

Naive Bayes Accuracy :
 93.79406648429645 %
 Completion Speed 0.35348

Decision tree Accuracy :
 98.55668748040587 %
 Completion Speed 1.6326

Support vector machine Accuracy :
 93.62738823407057 %
 Completion Speed 15.40661

Kneighborsclassifier Accuracy :
 81.464022923447 %
 Completion Speed 4.132

Random Forest Accuracy :
 46.92481190883222 %
 Completion Speed 0.76296

- After pre-processing the file is saved in the directory 'output.xlsx'.

- This file contains: ID (tweet) and Sentiment of each tweet is separated into 2 columns.
- The twitter dataset along with the sentiments is now ready.
- Now for training and Predicting.
- The code will run through the output.xlsx file and at the same time recover the tweet corresponding to the id of each sentiment.
- Using this we use the original data and feed them to our classifiers. When everything is done you should have all the AUC of each classifier listed in the console. We can now even type a sample tweet like this:

```
print("Input your tweet : ")
inputtweet = input()
```

```
datreeINPUT(inputtweet)
```

```
Input your tweet :
```

```
I am feeling sad
```

```
Hey, bud! You alright? Try having a decaf and take a deep breath. It will be alright. :)
```

```
*****
```

```
Negative
```

```
*****
```

- Negative means the person is likely to be sad or not in a good mood, hence, the recommendation system recommended the user to have a decaf and cheered the user up by telling him that whatever bothering him will be alright.

6.RESULTS:

The proposed model yields the following accuracy with the given computation time:

Naive Bayes Accuracy :
93.79406648429645 %
Completion Speed 0.35348

Decision tree Accuracy :
98.55668748040587 %
Completion Speed 1.6326

Support vector machine Accuracy :
93.62738823407057 %
Completion Speed 15.40661

Kneighborsclassifier Accuracy :
81.464022923447 %
Completion Speed 4.132

Random Forest Accuracy :
46.92481190883222 %
Completion Speed 0.76296

- The model, predicts the sentiment correctly and overcomes the limitation of existing state of the art.
- The model pre-processes the data efficiently.


```
=====
Retrieving TXT File
=====
Retrieving Successfull
=====
```

```
=====
Recovering Data Teets
=====
Data Tweets Recovered
=====
```

```
=====
Reading Dictionary
=====
Dictionary Preparation Done
=====
```

```
=====
Processing please wait...
=====
```

Processing time: 78.16853857 Seconds

```
=====
Processing Finish
=====
Data Saved!
=====
```

- The model labels data and specifies sentiment of the tweet efficiently.

```
In [14]: print("Input your tweet : ")
inputtweet = input()

datreeINPUT(inputtweet)

Input your tweet :
I am feeling sad
Hey, bud! You alright? Try having a decaf and take a deep breath. It will be alright. :)

*****
Negative
*****
```

```
In [15]: print("\nInput your tweet : ")
inputtweet = input()

datreeINPUT(inputtweet)

Input your tweet :
I am feeling good
Hey, You seem to be in a glee. Why not have your favourite meal? :)

*****
Positive
*****
```

7. CONCLUSION AND FUTURE SCOPE:

Although, the proposed model has given great accuracy and more than the previous models as an output result, but still the computation time being used is very large.

8. FUTURE SCOPE OF THE PROJECT:

As of now the model can successfully, retrieve the tweets from the dataset, analyze them and classify them according to their state, with a very higher accuracy, and can successfully recommended the user to try a decaf or a meal based on the state of his/her tweet.

For the future scope of this project, I have two things to do:

- 1. The classification of the tweets:**

I will classify the tweets into more than two states such as excited, neutral, grey, feeling-low, and many more states

- 2. The recommendation system:**

The vocabulary of the recommendation system will be made wider and location of the nearby place of a café or restaurant will be added and then, the user will be recommended with the location of his/her nearby place to relax too.

REFERENCES

- [1] A. Pak and P. Paroubek,” Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326
- 2] R. Parikh and M. Movassate, “Sentiment Analysis of User- Generated Twitter Updates using Various Classification Techniques”, CS224N Final Report, 2009
- [3] Go, R. Bhayani, L. Huang. “Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper,2009
- [4] P. D. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.
- <https://vgpena.github.io/classifying-tweets-with-keras-and-tensorflow/>
- <https://medium.com/@sonicboom8/sentiment-analysis-torchtext-55fb57b1fab8>
- <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2857/3251>
- <https://monkeylearn.com/blog/sentiment-analysis-of-twitter/>