DALHOUSIE UNIVERSITY

A HYBRID CLASSIFICATION ALGORITHM TO CLASSIFY ENGINEERING STUDENTS SENTIMENTS

CS 6405: Data Mining Project

April 11, 2017

Submitted to:
Dr. Qigang Gao

Submitted by:
Keshav Mittal, B00751567
Keerthana S.Mohan, B00741574

Contents

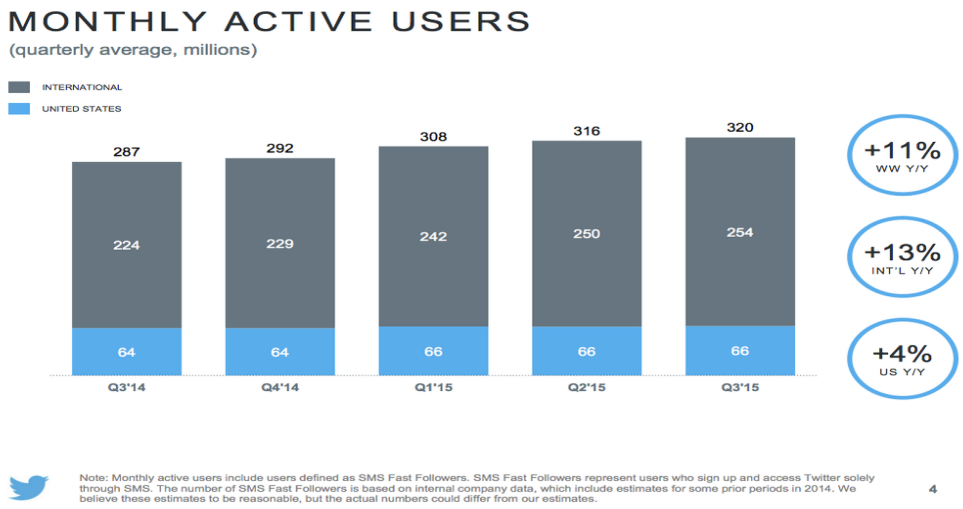
ABSTRACT

The social networking sites have now become the most easily used and accessible way for expressing views and opinions of individuals. They also provide medium to students to share their sentiments including struggles and joy during the journey. Such informal information has a great venue for decision making. The large and growing scale of information needs automatic classification techniques. Sentiment analysis is one of the automated techniques to classify large data. The existing predictive sentiment analysis techniques are highly used to classify reviews on E-commerce sites to provide business intelligence. But the conventional sentiment analysis techniques are not much useful to draw decisions in education system as they classify the sentiments into mostly three pre-set categories: positive, negative and neutral. Moreover, classifying the students' sentiments into positive or negative category does not provide deeper insight into their problems and perks. In this paper, we are utilizing linguistic features for detecting the sentiment of Twitter messages. We evaluate the usefulness of existing lexical resources as well as features that capture information about the informal and creative language. We take a supervised approach to the problem, but leverage existing hash tags in the Twitter data for building training data. Here we propose a novel Hybrid Classification Algorithm to classify engineering students' sentiments. Unlike conventional predictive sentiment analysis techniques, the proposed algorithm makes sentiment analysis process descriptive. It classifies student's perks in addition to problems into several categories to help future students and education system in decision making.

1. INTRODUCTION

We are living in an era where the advancement in technology is happening at a tremendous rate. There is much advancement done in the field of computer science. A major advancement in technology is emergence of Social Media applications. Social media applications have become so interesting that people spends hours communicating with friends and family on applications like Twitter, Facebook, and LinkedIn etc. on daily basis.

In the past few years, there has been a huge growth in the use of micro blogging platforms such as Twitter. Spurred by that growth, companies and organizations are increasingly seeking ways to mine Twitter for information about what people think and feel about their Products and Services.



[1.1]

Social media has also emerged as a personal communication technology which provides a large platform for individuals to share their feeling, emotions, and opinions in an informal and casual way to seek social support. The student's informal talks on social media platforms such as Twitter, Facebook, and YouTube shed insight into their present or past learning experiences.

Mining such homogeneous data extracts useful pattern from the large volume of students' generated raw data to support decision making in education system. While there has been a fair amount of research on how sentiments are expressed in genres such as online reviews and news articles, how sentiments are expressed given the informal language and message-length constraints of micro blogging has been much less studied.

Social media can be used for research purpose as it provides vast amount of data both in quantity and variety. That is why data collection is done from social media websites for analysis purpose. However, variety of platforms, informal language, slangs, timing of different posts, and different languages make the analysis process difficult to gain the hidden emotions in educational decision making. Hence, we require an efficient analysis and classification technique that incorporates qualitative analysis and automated technique to get deeper insight into the students' sentiments.

In this paper, we explore one method for building such data: using Twitter hash tags (e.g., #engineeringproblems, #engineeringperks) to identify student's emotion such as happy, sad, depressed, excited, etc unlike the conventional sentiment analysis techniques that only predict emotions such as positive, negative or neutral. As student's emotion cannot be just classified into three emotions, we need more descriptive way to classify their sentiments.

We propose a novel Hybrid Classification Algorithm (HCA) for descriptive sentiment analysis to understand students' problems and perks. The proposed algorithm classifies engineering students' sentiments into several categories (fig 1) that help future students and education system in decision making. We are using a Twitter corpus that is list of English words rated with an integer that in turn are used to calculate polarity and classify that into emotions. The dynamic process eliminates the requirement of changing the algorithm for newly added data. In our study, we have collected data from Twitter that has witnessed a tremendous growth in the number of users recently. The Tweets of students on Twitter consist of hash tags such as #engineeringProblem. Tweets are mostly public and limited to 140 characters that simplify identification of emotions in text.

Emotions
Neutral
Glad
Frown
Chill
Disappointed
Cheerful
Happy
Excited
Anxious
Sad
Dead

Fig 1.2

2. RELATED WORK

In recent years, a voluminous amount of research has been done in the field of sentiment analysis. In [12], authors have presented the logical approach for extraction of the sentiment on widely used social networking sites. They have analyzed the sentiments of the text using combinatory categorical grammar, lexicon acquisition and annotation, and semantic networks for analyzing. Classification accuracy of the feature vector is tested for electronic products domain International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016 75 using different classifiers such as Nave Bayes, SVM, Maximum Entropy, and Ensemble classifiers in [14]. In [15], authors have introduced a hybrid method that combines usage of sentiment lexicons along with a machine learning classifier for polarity detection of opinionated texts in the domain of consumer-products. In [16], authors have proposed a set of techniques of machine learning with semantic analysis for classifying the sentence and product reviews based on twitter data using WordNet for better accuracy. In [17], authors have examined the performance of different classifiers such as Naive Bayesian, SMO, SVM and Random Forest to classify Twitter data. It is observed that the existing techniques typically classify the text data into pre-set categories: positive, negative and neutral. Classifying the students' sentiments into positive or negative category does not provide deeper insight to their problems. Our proposed algorithm classifies sentiments apart from just pre-set categories of

positive, negative and neutral and helps in predicting student's sentiments with a deeper understanding.

3. APPLICATION SCENARIO

The conventional sentimental analysis techniques classify emotions only based on positive, negative or neutral. Our application focuses on classifying student's emotions like engineering students which can't be just described as positive or negative and hence need more descriptive way to classify these emotions. Hence, we cannot use the conventional techniques and need to introduce more emotions so prediction can be more descriptive and deep. This can further help institutions to understand student's problems and help in more reliable decision making that can positively affect learning process for students.

Here we are using supervised machine learning algorithm specifically Support Vector Machines. SVM is a classification algorithm. We are using text based classification.
Below is the roadmap of the steps performed in application scenario:

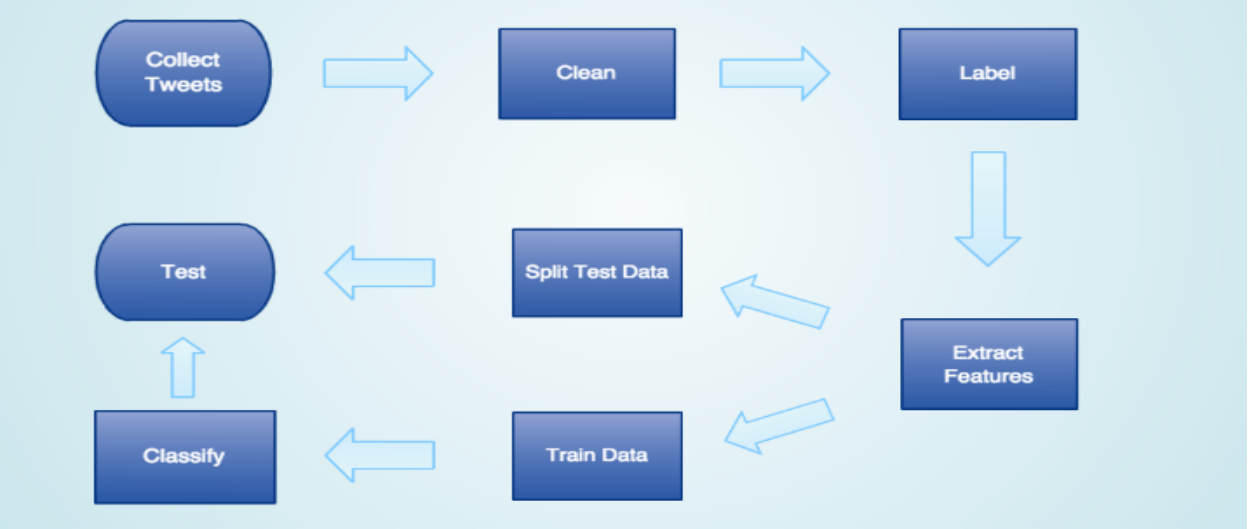


Fig [1.3]

First we would like to give a quick overview of the whole process will work before moving on to the algorithm implementation. For our application scenario we have performed the following steps for classification:

Step 1:
Data collection:
We are using Twitter API to collect data on the basis of keyword given by the user. After user give the keyword, the API will fetch 200 tweets based on that keyword and saving it in a CSV file.

Step 2:
Data Preprocessing:
After collecting raw data in form of tweets we need to pre-process it as it contains many useless keywords, stop words, hash tags, non uniform casing, account ids, quotations, and emoticons.
So to pre-process data and use clean data for prediction we are going to remove all the useless keywords or symbols by creating functions.

Step 3:
Training:
For training first we use TEXT BLOB SENTIMENT ANALYSIS then we extract features vector of the tweets and we match them with twitter corpus that contains English keywords with their impact value to calculate the polarity of the words to predict the emotion of the tweet. This in turn is trained using SVM

Step 4:
Labeling and Classification:
Dataset is labeled to classify the categorical classes they belong to and labeling is based on the polarity calculated.

Step 5:
Testing:
Apply SVM classifier function to predict the classes of the new tweets.

4. PROJECT OBJECTIVE

In this paper, Hybrid Support Vector Machine approach of Machine learning algorithm is implemented in such a way that it can classify the sentiments in more descriptive way.
It classifies engineering student's perks in addition to problems into several categories to help future students and education system in decision making.
Here we can classify the emotions in more than 3 categories like angry, funny, confused apart from the traditional categories like positive, negative and neutral.

5. ALGORITHM

HYBRID CLASSIFICATION ALGORITHM

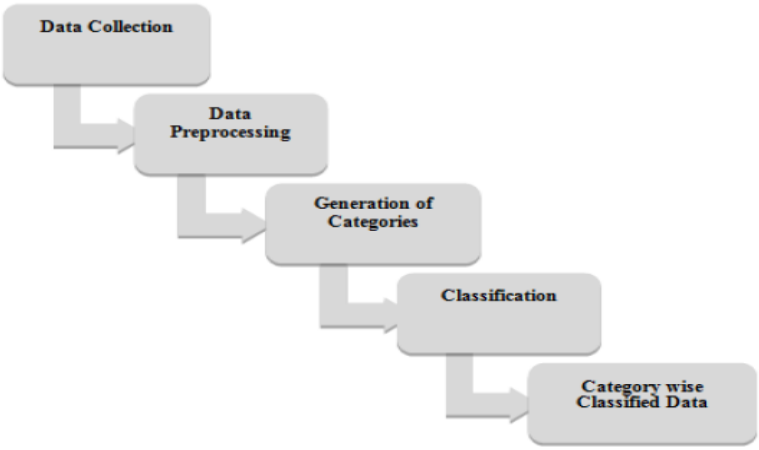


Fig 1.4

The sentiment analysis of student data is an emerging field that needs much more attention. In addition, classifying students' emotions into just positive or negative opinions do not shed any light into the actual problems in the existing learning process.
The proposed hybrid classification algorithm classifies students' problems and perks shared on Twitter into various belonging categories rather than merely positive or negative. Thus, the proposed algorithm makes the sentiment analysis process descriptive. The descriptive technique is more useful to get deeper insight into students' problems and perks than the traditional predictive techniques. [0]

Support Vector Machine

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
SVM provides higher theoretical accuracy in terms of correctly classifying the input data into valid categories.

Pseudo code for Algorithm:

Input: a dataset T = {t_1, t_2, t_3… t_n} of n Tweets that consist of #engineeringProblems and #engineeringPerks hashtags

Output: classification of n Tweets into m categories c_1, c_2, c_3, …., c_m

Hybrid Classification Algorithm (T, n)

BEGIN
1. for i = 1 to n Tweets
2. Convert t_i into uniform case
3. Remove Twitter notations, emoticons, URLs, and stop words from t_i
4. Compress the elongated words in t_i
5. Decompress the slang word in t_i
6. Generate c_1, c_2, …, c_m categories.
7. Apply SVM to classify n Tweets into their respective belonging class categories

END

6. DATA SOURCE AND PREPARATION

6.1 Data Source

First step in the data preparation is to fetch data. Here we are using Twitter API to fetch data from twitter based on certain keywords.

Data Source
https://dev.twitter.com/overview/api/tweets

Data Corpus

The data corpus for the polarity calculation is obtained from: https://github.com/ujjwalkarn/Twitter-Sentiment-Analysis/blob/master/AFINN-111.txt

Here keywords play a crucial role in fetching data as there may be some keywords that will fetch fewer tweets while some will fetch high number of tweets. However, twitter API is limited to fetch 200 tweets. As we can see in the below image how keywords fetch different number of tweets.

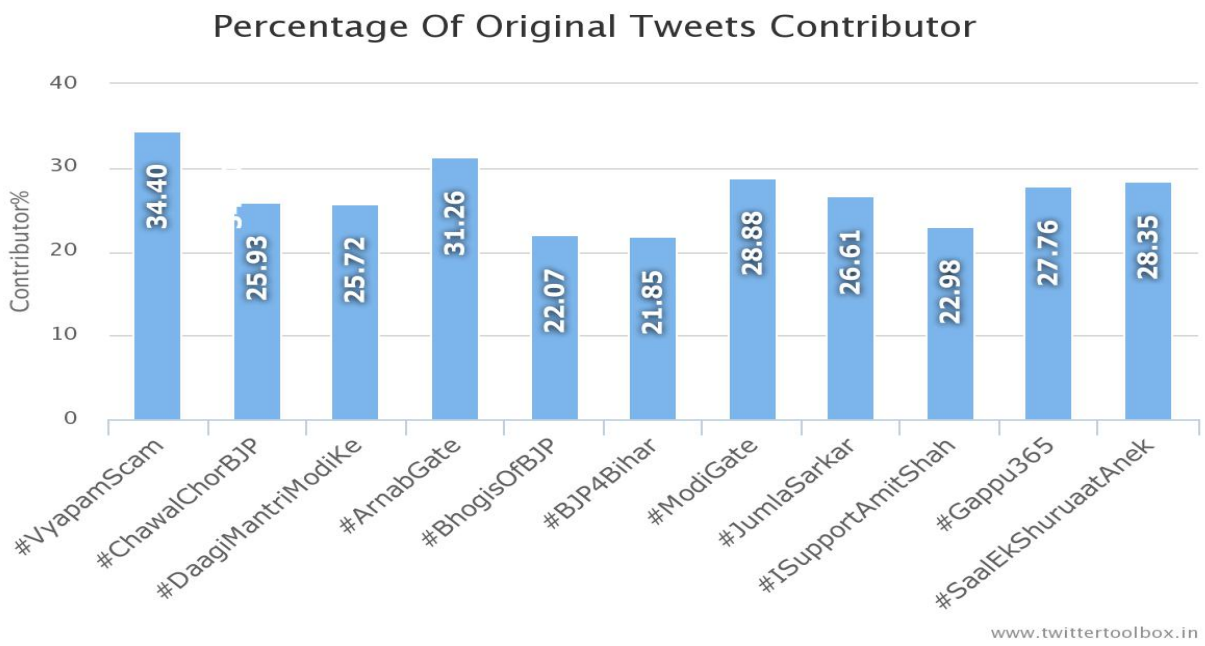


Fig 1.5

6.2 Data Preparation

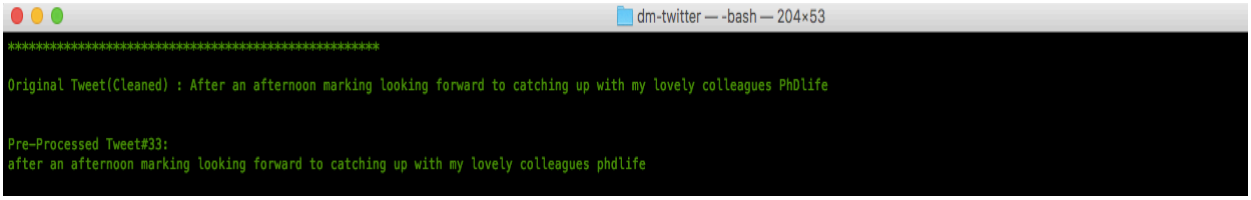
We are following the below steps to prepare our data:

Data Collection:
We extract data from twitter API. Then we pass it through TEXT BLOB sentiment analysis method. TEXT BLOB derives sentiment value like positive, negative and neutral of each of the tweet. Then we save that tweets with sentiment in a csv file. Below image contains all the details we are getting from tweet.

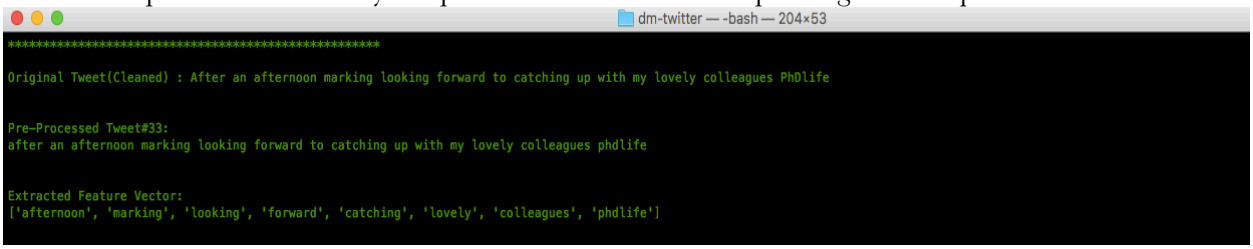
A1 | fx | text

	A	B	C
1	text	sentiment	
2	RT How to write a PhD literature review amp develop a conceptual framework phdchat phdadvice phdfo	neutral	
3	Crushing a term paper and lemon wafer cookies like it s my job Well the former IS my job but I m definitely better at the latter phdlife	positive	
4	Tip If you want your research assistant to be able to speak to materials you developed you must pay them to review the materials PhDlife	positive	
5	productivity tip review your day amp plan priorities for the next day No more than 3 5 per day acwri phdlife	negative	
6	Essentially my to do list for Easter shutupandwrite phdlife	neutral	
7	Advice in piece is all good but also highlights huge demands on PhD students today phdlife phdchat	positive	
8	TFW you find a book whose title amp description appear super relevant only to open it after getting it on ILL and realize nope phdlife	positive	
9	RT Stay productive during your PhD withaphd phdlife phdchat phdlife From	neutral	
10	Successful proposal defense phdlife	positive	
11	I was shortlisted in Doctoral Life images competition Me juggling PhD life note the look of exhaust	neutral	
12	Writing is a challenging medium in which to communicate about writing by PhDLife	positive	
13	That PhDLife	neutral	
14	The River Teme riverteme fieldwork phdlife Bransford Worcestershire United Kingdom	neutral	
15	After over one year of PhD research I can confirm that the figure really depicts reality phdchat phdlife	positive	
16	Policy analyst software dev cheesemonger TRaCE project tracks what PhDs do after highered	neutral	
17	RT Quand tu termines tous les livres que t as emmen s en mission de recherche et il n y a aucune librairie qu 25 km de chez	neutral	
18	My body has decided today is gonna be an eye twitchy kind of day phdlife	positive	
19	Step 3 going live in 10 minutes at phdchat PhD phdlife academicmamas	positive	
20	Quand tu termines tous les livres que t as emmen s en mission de recherche et il n y a aucune librairie qu 25 k	neutral	
21	Now this is a PhD I could get behind PhD Scholarship phdlife	negative	
22	How to write a PhD literature review amp develop a conceptual framework phdchat phdadvice	neutral	
23	and then I say to myself you can eat the muffin only after you ve done 2 more hours of transcription phdlife try it	positive	
24	Some of antidoping s next generation enjoying a first look at newresearch at PCC2017 womeninSTEM phdlife	positive	
25	RT So what is OMGitsscience all about Find out and subscribe to see new content first PhDlife UHasselt	positive	
26	RT You shall now call me Dr Lacy PhinisheD phd phdlife phdwoman phdstudent	neutral	
27	When you get official notice of your advancement to candidacy phdchat phdlife	neutral	
28	You shall now call me Dr Lacy PhinisheD phd phdlife phdwoman phdstudent	neutral	
29	evergreentweets phdlife	neutral	
30	RT So proud of my mentor Not only does she excel in research but her mentorship is second to none phdchat	positive	
31	Picasso summing up academia phdlife phdchat picasso	neutral	
32	Beards and books So excited to dive into these bad boys phdlife studentlife	negative	
33	RT News 7 advantages PhDs have over other job candidates phdlife phdadvice Via	negative	
34	Don t make fun of grad students they just made a terrible life choice phdlife	negative	
35	When you actually plan on a vending machine breakfast ES please phdlife	neutral	
36	Easter is my big chance to catch up on my work while sitting in the garden phdlife PhD	neutral	

Data Pre-processing:
Data collected in a csv file is then fetched and cleaned to remove all unwanted things like stop words, symbols, RT, uniform casing etc. Now we just have the useful words in the tweets.

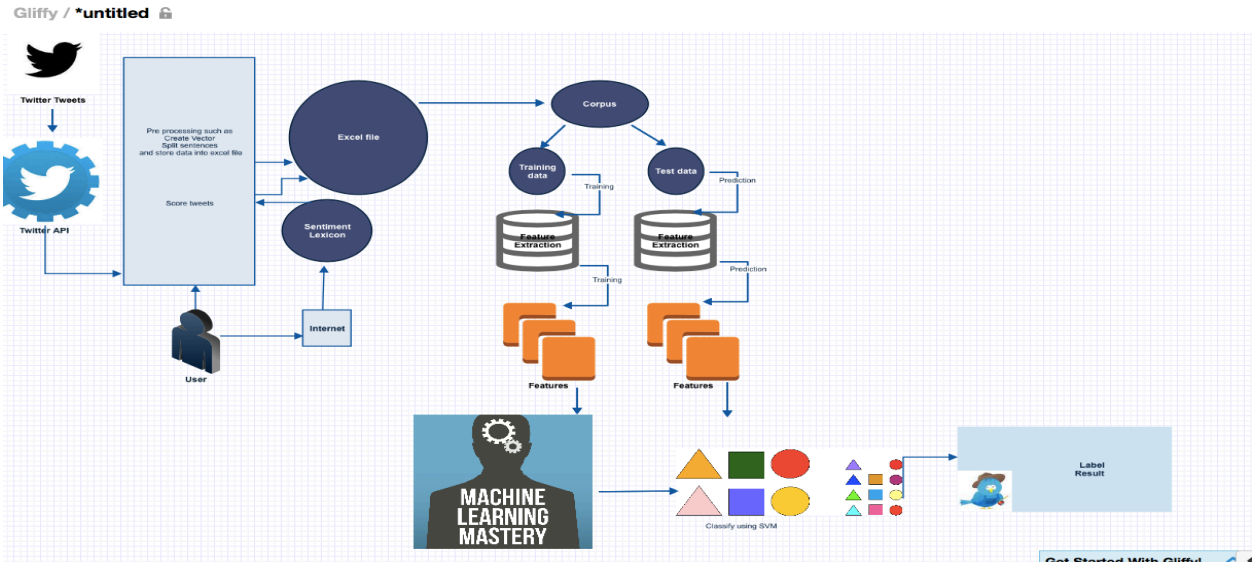


Feature extraction:
Then we are extracting features in the sense that we are extracting the words from each tweet to further compare it to see if they are present in the Twitter corpus to get the impact value.



Unigrams:
To get the impact value from corpus we need to compare the features extracted from tweets and the words present in the corpus.
To compare it we are using Unigrams. Unigrams means we are taking single word instead of combination of words for comparison. E.g.: "Most" this is a unigram while "most of" is a bi-gram.

7. PROJECT ARCHITECTURE



High level architecture of program. (Created using Gliffy)

Program architecture of our application contains several python files that are explained below with their functionalities in the table below.

gettwitterinput.py	Contains the main (). Prompts user to input the hash tag for which the data is to be fetched and contains the authentication for twitter API and function calls to other parts of the program
preprocessing.py	Contains the functions involved with stripping the words in each tweet and pre-processing them
polarity.py	Calculates the polarity of each tweet to assign it to multi-class emotions
svmdem.py	SVM training and testing is performed

stopwords.txt	Contains the stop words which are to be removed
feeds.csv	The extracted tweets from the API are saved to feeds file
AFINN-111.txt	Contains the adjectives and their impact values

7.1 Building the Program

Open the python console and change to the directory of the source code folder.
Install libraries like nltk, tweepy, tabulate etc.
Run the file gettwitterinput.py

Usage:

python gettwitterinput.py

7.2 Running a Demo

Input filename: gettwitterinput.py
Tweet file generated: feeds.csv
Skipping header column of feeds.csv

To build the program executables, navigate to the folder containing files in the terminal and type

Python gettwitterinput.py

Upon the program starting, you'll be prompted to enter a hashtag for which the data is to be collected. Since our aim is to do multi class classification for student tweets, some of the best hash tags to use are #phdlife, #gradlife etc., However feel free to try any hashtag.

```
Enter a hashtag for which tweets are to be obtained: gradlife
```

The program will fetch data for the hashtag, save it to feeds.csv file and use those tweets to pre-process and come up with feature vectors for each tweet. The polarity and sentiment is then computed and the SVM classifier is trained with this data.

```
**********************************************************

Training SVM Classifier....

Positive tweets percentage: 33 %
Negative tweets percentage: 16 %
Neutral tweets percentage: 50 %
Percentage of emotions detected in the training set :
---------------------------------------------------------------
Disturbed tweets percentage: 44 %
```

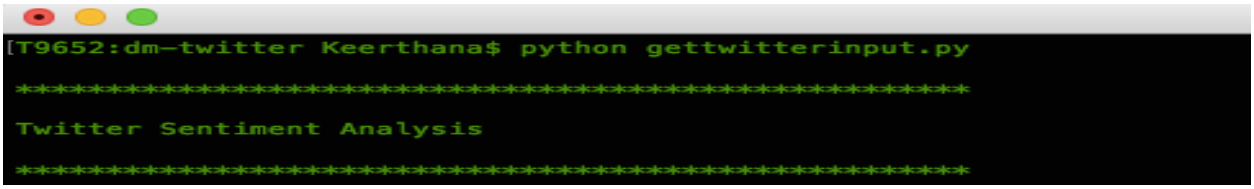
```
cheerful tweets percentage: 16 %
anxious tweets percentage: 8 %
over the clouds tweets percentage: 5 %
Needs help tweets percentage: 4 %
Excited tweets percentage: 4 %
happy tweets percentage: 4 %
frown tweets percentage: 3 %
sad tweets percentage: 2 %
deeply depressed tweets percentage: 1 %
Glad tweets percentage: 2 %
Frown tweets percentage: 3 %
Chill tweets percentage: 1 %
Disappointed tweets percentage: 0 %
Neutral tweets percentage: 0 %
Training done.
```

After the training is done, you'll be prompted to enter any tweet to test the classifier. Feel free to enter a tweet of your choice to check the sentiment generated.

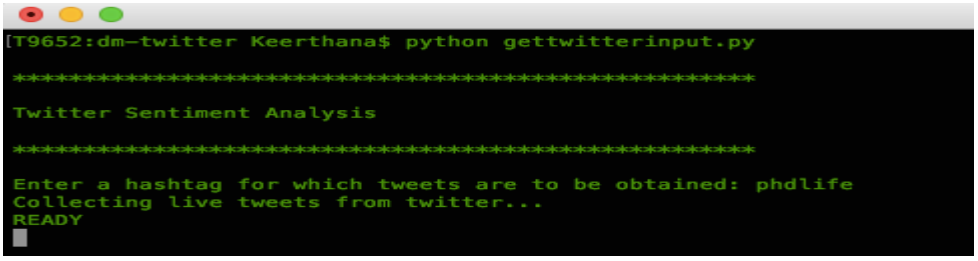
```
Enter a tweet message to find its sentiment polarity: i have no idea what im doing with my life
#gradlifesucks
Calculating Polarity of your tweet....
Emotion Analysed:
['disturbed']
```

7.3 User Interface

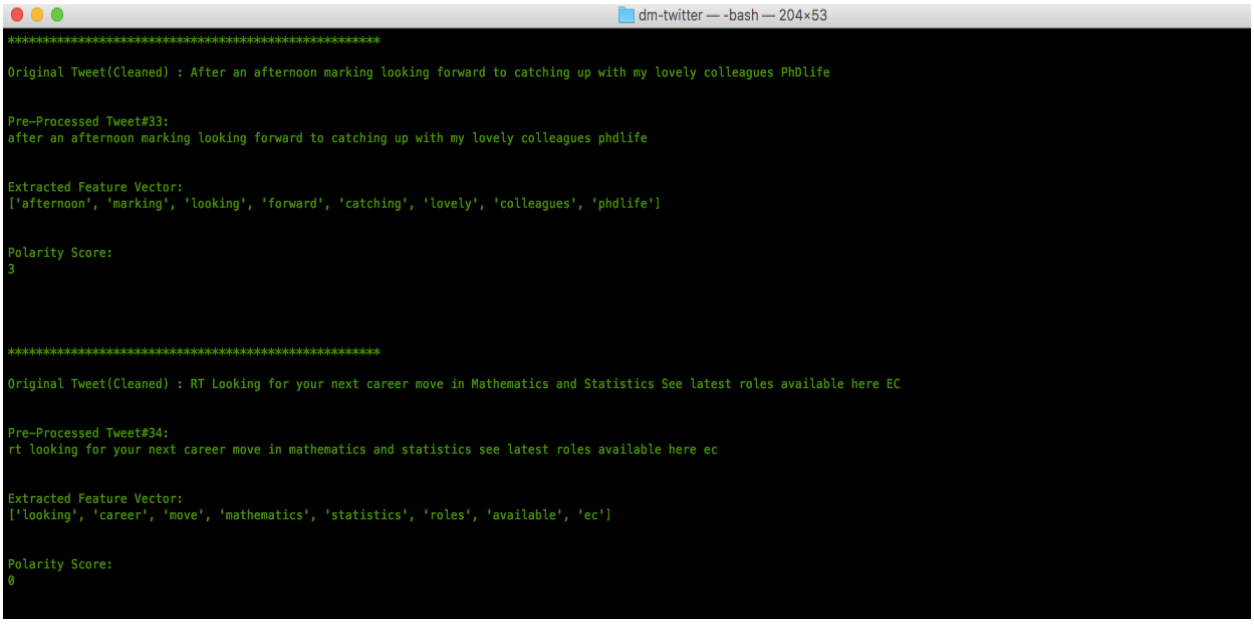
Our application is a console application so everything needs to be run on command prompt.



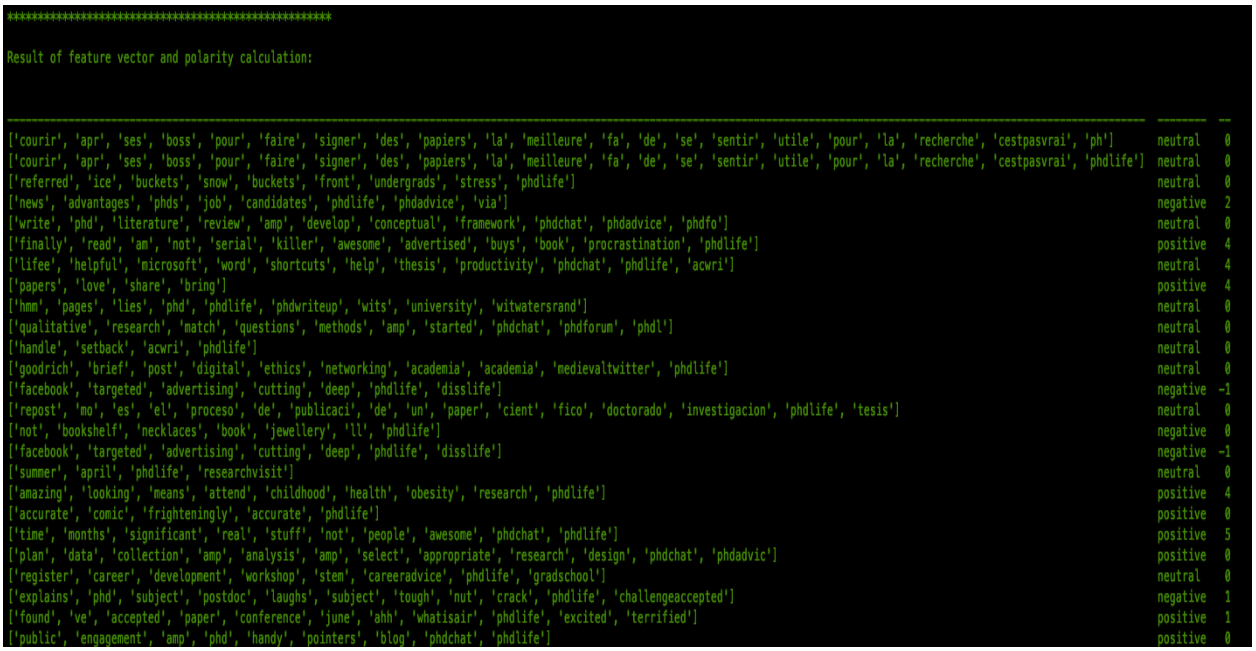
User is prompted to enter the keyword to get the tweets and we get READY as message showing that tweets has been successfully extracted.



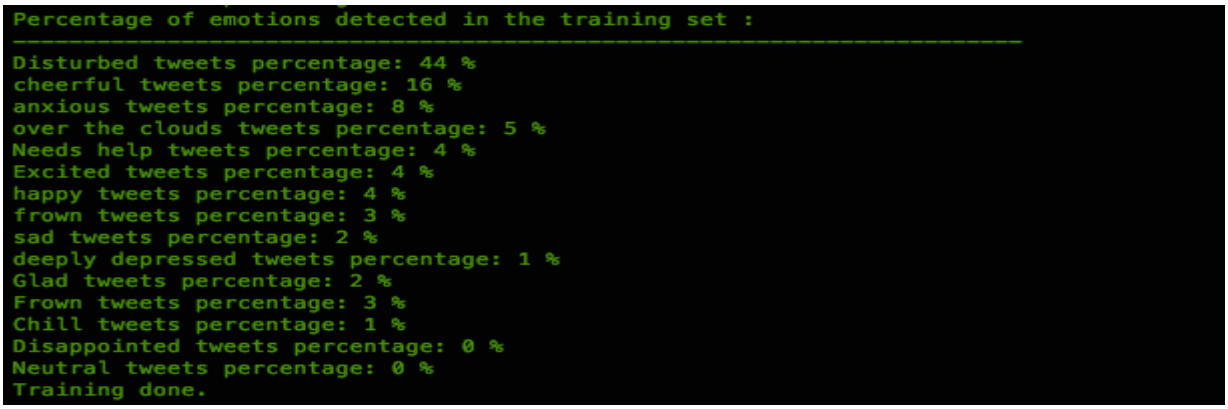
After the Ready message is displayed functions are called to do the pre-processing of the fetched data and we get result like the image below.



At last we get extracted features of all the tweets with their sentiments and polarity.



Emotions based on the polarity score.



User is then asked to input tweet to for which user wants prediction of emotion.

7.4 Code Structure

Below is the table explaining the code structure in the order they are executed with name and their functionality:

Functions name	Description
main()	Main method which prompts user to enter the value for hash tags and performs calls to clean, pre-process, generate feature vector and to determine sentiment and polarity of each tweet in the tweets file. Also makes call to SVM function
get_tweets()	function to fetch the tweets and parse them
cleaning()	remove links and special characters from tweet using regex
get_tweet_sentiments()	set sentiment using text blobs sentiment method
getStopWordsList()	getStopWordList function uses the stopwords.txt file to remove the words which do not display any emotion and are rendered trivial for the analysis rt and url are appended to remove retweets and urls too.
processTweet()	process_tweet function converts tweets to lower case, replace links with URL and user with AT_USER, removes additional white spaces and removes #

getFeatureVector()	Feature vector generation - split the tweet into words .Remove punctuations if any. remove stop words from the tweet
replaceTwoOrMore()	Look for 2 or more repetitions of character if there are two or more occurrences of a word like yay yay or okayyyyy - replace it with yay and okay.
calculatescores()	Determine the polarity of each tweet from the feature vector generated for each tweet, compare if the elements of the feature vector exist in the corpus and if they do, sum the impact value for the element. this way sum for all the elements in each feature vector
trainsvm()	Using the sentiment and polarity of each tweet, determine the emotion and train the SVM. Also, test the classifier by prompting user to enter any tweet and see the prediction

8. EXPERIMENTATION RESULTS

8.1 Interpretation of Results:

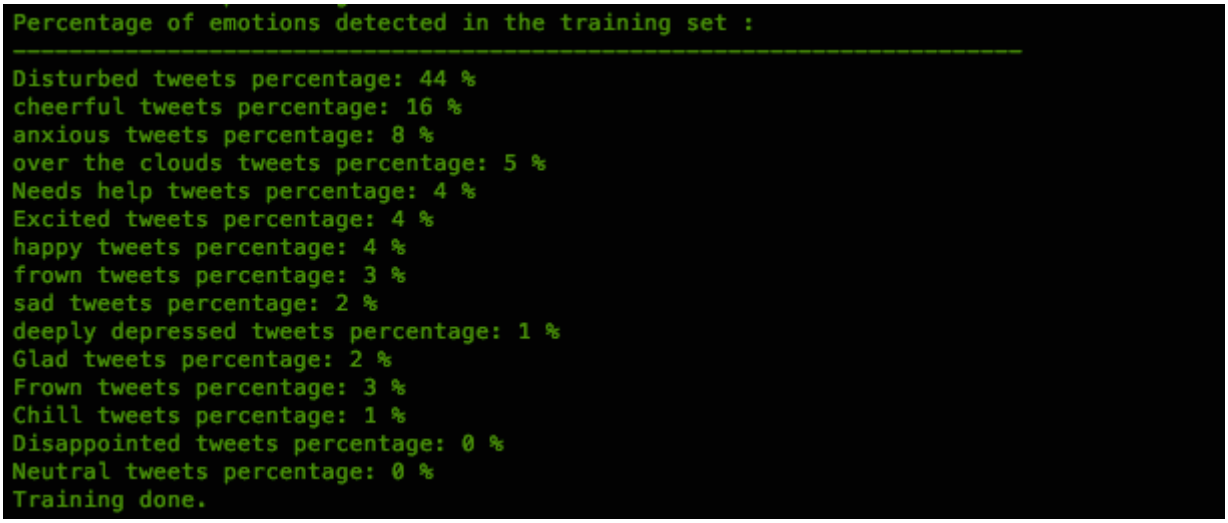
For our application result interpretation is straight forward.
When user enters a keyword, then processed tweets based on that keyword with the sentiment and polarity is displayed.

```
**************************************************

Result of feature vector and polarity calculation:

_____________________________________________________________________________________________________________________________________________________ ________ __
['courir', 'apr', 'ses', 'boss', 'pour', 'faire', 'signer', 'des', 'papiers', 'la', 'meilleure', 'fa', 'de', 'se', 'sentir', 'utile', 'pour', 'la', 'recherche', 'cestpasvrai', 'ph']          neutral   0
['courir', 'apr', 'ses', 'boss', 'pour', 'faire', 'signer', 'des', 'papiers', 'la', 'meilleure', 'fa', 'de', 'se', 'sentir', 'utile', 'pour', 'la', 'recherche', 'cestpasvrai', 'phdlife']  neutral   0
['referred', 'ice', 'buckets', 'snow', 'buckets', 'front', 'undergrads', 'stress', 'phdlife']                                                                                          neutral   0
['news', 'advantages', 'phds', 'job', 'candidates', 'phdlife', 'phdadvice', 'via']                                                                                                      negative  2
['write', 'phd', 'literature', 'review', 'amp', 'develop', 'conceptual', 'framework', 'phdchat', 'phdadvice', 'phdfo']                                                                  neutral   0
['finally', 'read', 'am', 'not', 'serial', 'killer', 'awesome', 'advertised', 'buys', 'book', 'procrastination', 'phdlife']                                                             positive  4
['lifee', 'helpful', 'microsoft', 'word', 'shortcuts', 'help', 'thesis', 'productivity', 'phdchat', 'phdlife', 'acwri']                                                                neutral   4
['papers', 'love', 'share', 'bring']                                                                                                                                                   positive  4
['hmm', 'pages', 'lies', 'phd', 'phdlife', 'phdwriteup', 'wits', 'university', 'witwatersrand']                                                                                        neutral   0
['qualitative', 'research', 'match', 'questions', 'methods', 'amp', 'started', 'phdchat', 'phdforum', 'phdl']                                                                          neutral   0
['handle', 'setback', 'acwri', 'phdlife']                                                                                                                                              neutral   0
['goodrich', 'brief', 'post', 'digital', 'ethics', 'networking', 'academia', 'academia', 'medievaltwitter', 'phdlife']                                                                 neutral   0
['facebook', 'targeted', 'advertising', 'cutting', 'deep', 'phdlife', 'disslife']                                                                                                       negative -1
['repost', 'mo', 'es', 'el', 'proceso', 'de', 'publicaci', 'de', 'un', 'paper', 'cient', 'fico', 'doctorado', 'investigacion', 'phdlife', 'tesis']                                     neutral   0
['not', 'bookshelf', 'necklaces', 'book', 'jewellery', 'll', 'phdlife']                                                                                                                negative  0
['facebook', 'targeted', 'advertising', 'cutting', 'deep', 'phdlife', 'disslife']                                                                                                       negative -1
['summer', 'april', 'phdlife', 'researchvisit']                                                                                                                                        neutral   0
['amazing', 'looking', 'means', 'attend', 'childhood', 'health', 'obesity', 'research', 'phdlife']                                                                                     positive  4
['accurate', 'comic', 'frighteningly', 'accurate', 'phdlife']                                                                                                                          positive  0
['time', 'months', 'significant', 'real', 'stuff', 'not', 'people', 'awesome', 'phdchat', 'phdlife']                                                                                   positive  5
```

User is then prompted to enter a tweet which then predicts the emotion.

```
Enter a tweet message to find its sentiment polarity: i hate my job
Calculating Polarity of your tweet....
Emotion Analysed:
['deeply depressed']
```

We get count of tweets based on their emotions



9. EVALUATION

We discuss about the advantages and limitation of the algorithm we implemented in this application

9.1 Algorithm Advantage:

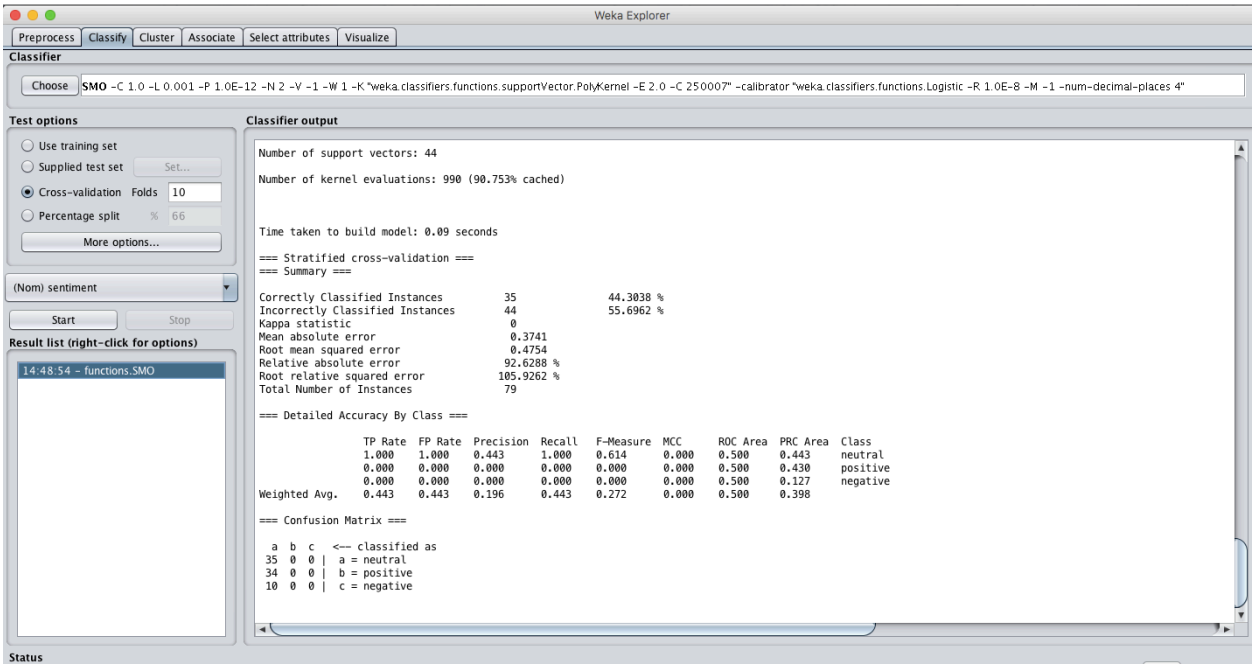
The main advantage of this algorithm is that it is different from conventional SVM in a way that it can predict more emotions then the former one. This quality of the algorithm gives it a edge in predicting sentiment in a more descriptive and deep manner specially if someone wants to analyze the student sentiment.

9.2 Algorithm Limitation:

1. The limitation of this algorithm is that it will only predict the emotions that are hardcoded in the code means if we have defined 10 emotions it will detect only those 10 emotions.
2. The other limitation is the training data availability. There are some keywords for which there are less tweets and if we try to predict emotions based on those keywords, accuracy will be compromised. It gives better result when it is trained with large data.

9.3 Comparison with WEKA

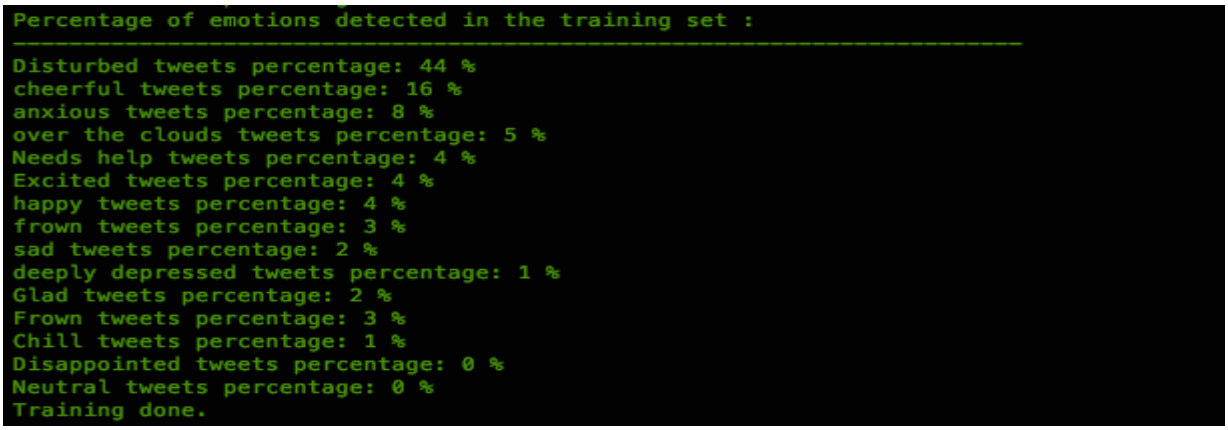
When we run SVM in WEKA, tweets are classified only in the categories of positive, negative and neutral. Same can be seen in the below result.



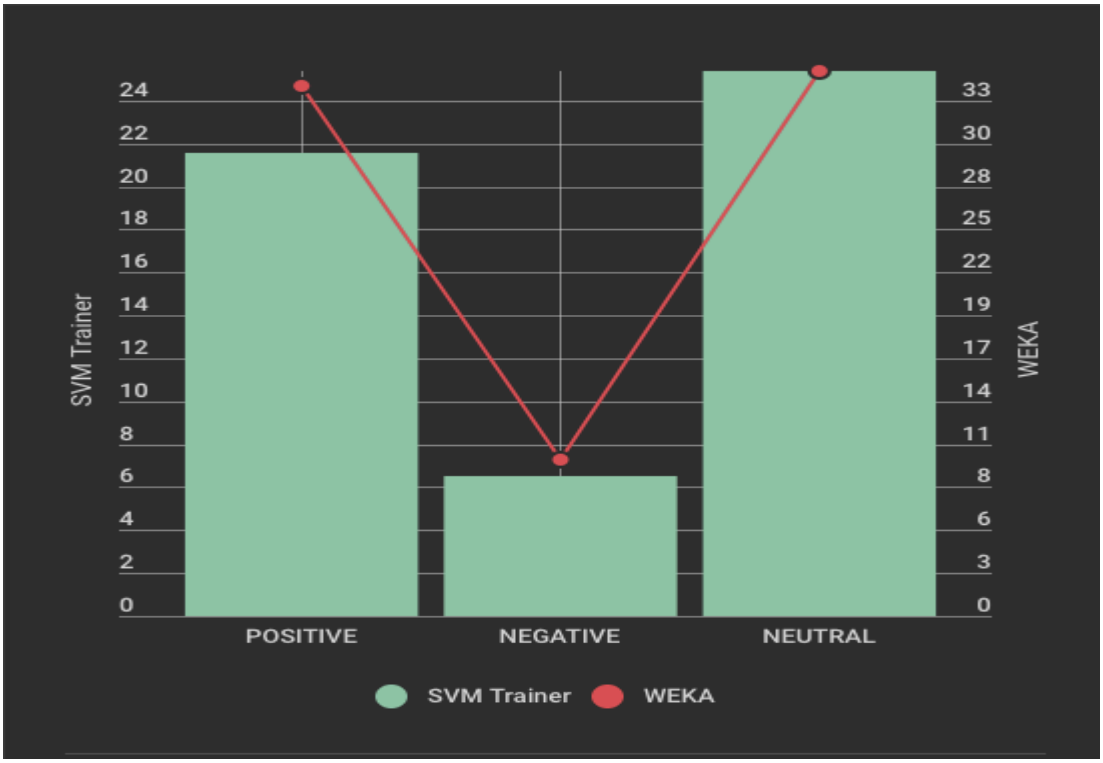
When we run our hybrid SVM we get more descriptive results and tweets are classified into more than three categories.

We get tweets predicted as sad, happy, excited, depressed etc

Same can be seen in the below result.



Below we can see the graph which shows the difference between accuracies in conventional SMV and hybrid SVM.



10. CONCLUSION AND FUTURE SCOPE

Future improvements to the hybrid algorithm can be made by introducing more labels in the multi class emotion classification. The algorithm can be further built by considering emoticons into account and adding their impact score to the decision weightage. Emoticons provide more emphasis to the tweets and can be helpful in classifying more accurately. The Classifier is a supervised algorithm and hence with more training data, can predict more accurately. The amount of data fed into the training algorithm has an impact on the accuracy of emotion prediction.

11. REFERENCES

1.1 - https://fortunedotcom.files.wordpress.com/2015/10/cswccuwvaaaksly.png
1.3 - http://slides.com/mertkahyaoglu/twitter-sentiment-analysis-4#/2
1.4 - Mitali Desai and Mayuri A. Mehta, "A HYBRID CLASSIFICATION ALGORITHM TO CLASSIFY ENGINEERING STUDENTS' PROBLEMS AND PERKS", International Journal of Data Mining

& Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016
1.5 - http://political.analysis.twittertoolbox.in/images/screenshots/screens/7_Perctangge_of_originl_tweet.jpeg
[0] Mitali Desai and Mayuri A. Mehta, "A HYBRID CLASSIFICATION ALGORITHM TO CLASSIFY ENGINEERING STUDENTS' PROBLEMS AND PERKS", International Journal of Data Mining &

Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016
[1] I. King, J. Li and K. T. Chan, "A Brief Survey of Computational Approaches in Social Computing",
In Proc. of International Joint Conference on Neural Network, 2009, pp. 2699-2706.
[2] S. R. Barahate and V. M. Shelake, "A Survey and Future Vision of Data mining in Educational
Field", in Proc. 2nd Int. Conf. on Advanced Computing & Communication Technology, 2012, pp.
96-100.
[3] M. Dredze , "How Social Media Will Change Public Health", IEEE Intelligent Systems, 2012, pp.
1541-1672.
International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016
80
[4] P. A. Tess, "The role of social media in higher education classes (real and virtual) – A literature
review", Computers in Human Behavior, 2013, pp. 60-68.
[5] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education", Educause
Review, 2011, vol. 46, no. 5, pp. 30-32.
[6] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," in Systems,
Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , 2010, vol.40, no.6,
pp.601-618.
[7] X. Chen, M. Vorvoreanu and K. Madhavan, "Mining Social Media Data to Understand Students'
Learning Experiences", IEEE Transaction, vol. 7, no. 3, pp. 246-259, 2014.
[8] Weil, Kevin (VP of Product for Revenue and former Big Data engineer, Twitter Inc.), "Measuring
Tweets." Twitter Official Blog, February 22, 2010. [Online].

Available: http://www.internetlivestats.com/twitter-statistics. [Accessed: 19-Oct-2015].
[9] Krikorian, Raffi (VP, Platform Engineering, Twitter Inc.), "New Tweets per second record, and
how!" Twitter Official Blog. August 16, 2013. [Online].
Available: https://blog.twitter.com/2013/new-tweets-per-second-record-and-how. [Accessed: 19-Oct-
2015].
[10] Twitter Engineering, "200 million Tweets per day." Twitter Official Blog. June 30, 2011. [Online].
Available: https://blog.twitter.com/2011/200-million-tweets-per-day. [Accessed: 19-Oct-2015].
[11] "Twitter turns six." Twitter Official Blog. March 21, 2012. [Online].
Available: https://blog.twitter.com/2012/twitter-turns-six. [Accessed: 19-Oct-2015].
[12] N. Kasture and P. Bhilare, "An Approach for Sentiment analysis on social networking sites",
Computing Communication Control and Automation (ICCUBEA), 2015, pp. 390-395.
[13] S. Bhuta, A. Doshi, U. Doshi and M. Narvekar, "A review of techniques for sentiment analysis Of
Twitter data", Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014, pp. 583-
591.
[14] M. S. Neethu and R. Rajasree, "Sentiment Analysis in Twitter using Machine Learning Techniques",
in 4th Int. Conf. of Computing, Communications and Networking Technologies (ICCCNT), 2013,
pp. 1-5.
[15] S. Bahrainian and A. Dangel, "Sentiment Analysis uses Sentiment Features", in Int. joint Conf. of
Web Intelligence and Intelligent Agent Technologies, 2013, pp. 26-29.
[16] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and
semantic analysis", in 7th Int. Conf. on Contemporary Computing, 2014, pp. 437-442.
[17] B. Gokulakrishnan, P. Plavnathan, R. Thiruchittampalam, A. Perera and N. Prasath, "Opinion Mining
and Sentiment Analysis on aTwitter Data Stream", in Int. Conf. on Advances in ICT for Engineering
Regions, 2012, pp. 182-188.
[18] V. Singh and S. K. Dubey, "Opinion mining and analysis: A literature review", in 5th Int. Conf. on
Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 232-239.