

# Emotion Identification between POMS and Multinomial Naive Bayes Algorithm Using Twitter API

Asharani S Dandoti<sup>1\*</sup>, Sunil M Sangve<sup>2</sup>

<sup>1,2</sup>Department of Computer Engineering, Zeal College of Engineering and Research, Pune, India

Corresponding Author: Asharani S Dandoti\_ashadandoti@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7i7.1419> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 14/Jul/2019, Published: 31/Jul/2019

**Abstract**— The analysis of social networks is a very challenging research area while a fundamental aspect concerns the detection of user communities. The existing work of emotion recognition on Twitter specifically depends on the use of lexicons and simple classifiers on bag-of words models. The vital question of our observation is whether or not we will enhance their overall performance using machine learning algorithms. The novel algorithm a Profile of Mood States (POMS) represents twelve-dimensional mood state representation using 65 adjectives with combination of Ekman's and Plutchik's emotions categories like, anger, depression, fatigue, vigour, tension, confusion, joy, disgust, fear, trust, surprise and anticipation. These emotions classify with the help of text based bag-of-words and LSI algorithms. The contribution work is to apply machine learning algorithm for emotion classification, it takes less time for classification without interfere human labeling. The Multinomial Naïve Bayes classifier works on testing dataset with help of huge amount of training dataset. Measure the performance of POMS & Multinomial Naïve Bayes algorithms on Twitter API. The result shows with the help of Emojis for emotion recognition using tweet contents.

**Keywords**— Emotion Recognition, Text Mining, Twitter, Recurrent Neural Networks, Convolutional Neural Networks, Multinomial Naïve Bayes Classifier

## I. INTRODUCTION

Emotions can be defined as conscious affect attitudes, which constitute the display of a feeling. In recent years, a large number of studies have focused on emotion detection using opinion mining on social media. Due to some intrinsic characteristics of the texts produced on social media sites, such as the limited length and casual expression, emotion recognition on them is a challenging task. Previous studies mainly focus on lexicon-based and machine learning based methods. The performance of lexicon-based methods relies heavily on the quality of emotion lexicon and the Performance of machine learning methods relies heavily on the features. Therefore, we work with three classifications That is the most popular, and has also been used before by the researchers from computational linguistics and natural language processing (NLP). Paul Ekman defined six basic emotions by studying facial expressions. Robert Plutchik extended Ekman's categorization with two additional emotions and presented his categorization in a wheel of emotions. Finally, Profile of Mood States (POMS) is a psychological tool that defines a six-dimensional mood state representation using text mining. The novel algorithm a

Profile of Mood States (POMS) generating twelve-dimensional mood state representation using 65 adjectives with combination of Ekman's and Plutchik's emotions categories like, anger, depression, fatigue, vigour, tension, confusion, joy, disgust, fear, trust, surprise and anticipation. Past work commonly considered just a single Emotion arrangement. Working with various characterizations at the same time not just empowers execution correlations between distinctive Emotion classifications on a similar sort of information yet additionally enables us to build up a solitary demonstrate for anticipating numerous characterizations in the meantime.

### 1.1 Motivation

The framework created dependent on our proposed methodology would have the capacity to consequently recognize what individuals feel about their lives from twitter messages. For instance, the framework can recognize:

- Percentage of individuals communicating better degrees of life fulfillment in one gathering rather than another gathering,
- Percentage of those individuals who feel happy and chipper,

- Percentage of those individuals who encounter quiet and peaceful, and
- Percentage of those individuals who communicating large amount of frenzy or trouble.

## II. RELATED WORK

In [2] paper, explore whether open mind set as estimated from expansive scale accumulation of tweets posted on twitter.com is corresponded or even predictive of DJIA values. The result show that adjustment in the general population inclination state can to be sure be followed from the substance of extensive scale twitter channels by methods for or may be straight for word content preparing methods and that such changes react to an assortment of socio-social drivers in an exceedingly separated way. Advantages are: Build the execution open mind-set investigation from twitter channels offers a programmed, quick, free expansive scale expansion to this toolbox that might be stream lined to quantity an assortment of measurements of the general population disposition state. Disadvantages are: It avoids geographical and cultural sampling mistakes.

The paper [3] broke down budgetary web journals and online news articles to build up an open state of mind dynamic forecast display. For securities, exchanges, referencing the points of view of conduct back and the qualities of online money related networks. An open mind-set time arrangement forecast show is likewise exhibited, incorporating highlights from informal organizations and conduct fund, what's more, utilizes huge information examination to evaluate passionate substance critique on current stock or money related issues to estimate changes for Taiwan stock record. Advantages are: It is helpful for highlight word development and handling speed more widely used. Disadvantages are: Only uses for stock prices.

In [4] paper the product of profound repetitive neural system to the test of sentence-level assessment articulation extraction. DSEs (direct subjective expressions) comprise of explicit notices of individual states or discourse occasions communicating non-open states; and ESEs (expressive subjective expressions) include articulations that demonstrate assumption, emotion, and so forth etc.; without expressly passing on them Advantages are: Profound RNNs outflanked past(semi)CRF baselines; accomplishing new cutting edge results for fine-grained on sentiment articulation extraction.. Disadvantages are: RNNs do not have access to any highlights other than word vectors.

In [5] paper break down constituent tweets for all the more quietly communicated data, for example assumption (enormous or terrible), the emotion (pleasure, sadness, anger, and so forth and etc.), the reason or then again purpose for the tweet (to call attention to a slip-up, to help, to disparage,

etc.), and the style of the tweet (straightforward, articulation, mockery, overstatement, and numerous others). There are two area; on clarifying content for supposition ,style, and classifications ,for example reason what's more on programmed classifiers for recognizing these classifications. Advantages are: Using a large no of uniquely designed highlights like those concerning emojis, accentuation, extended words and invalidation alongside unigrams, bigrams, and emotion vocabularies include the SVM classifier accomplished a higher exactness.. Disadvantages are: Does not outline tweets. Does not consequently recognizing other semantic jobs of emotions. For example: degree, reason, and sympathy target.

In [6] paper, i) Show how a ton of web based life data can be used for gigantic scale open-vocabulary character acknowledgment; ii) Dissect which features are judicious of which personality estimation and iii) Present a novel corpus of 1.2m English tweets(1,500 makers) remarked on for sex introduction and MBTI. Advantages are: The identity qualifications, specifically Introvert extrovert (IE) and Thinking feeling (TF), can be anticipated from web based life information with unwavering quality. The expansive-scale, open-vocabulary investigation of client attributes can help enhance characterization exactness.

The paper [7] build up a perform multi-task DNN for learning portrayals different assignments, not just utilizing expansive measure of cross-undertaking information, yet additionally profiting by a regularization impact that prompts progressively broad portrayals to help under-takings in new space. A perform multiple tasks profound neural system for portrayal learning, specifically concentrating on semantic order (question arrangement) and semantic data recovery (positioning for web search)under takings. Exhibit solid results on question arrangement and web seek. Advantages are: The MT- DNN powerfully performs using solid baselines overall web search and question grouping undertaking. Disadvantages are: The query classification fused either as characterization or positioning assignment not thorough investigation work.

In [8] article, demonstrate emotion-word hash tags are great manual marks of emotions in tweets. in tweets. Proposes a strategy to produe an extensive dictionary of word emotion relationship from this emotion-marked tweet corpus. This is the primary dictionary with genuine esteemed word emotion affiliation scores. Advantages are: Using hash tagged tweets can gather a lot of marked information for any emotion that is utilized as a hash tag by tweets. Disadvantages are: This paper works just on given content not equivalent word of that content.

The paper [9] centres on concentrate two major NLP assignments, Discourse Parsing and Sentiment Analysis.

The enhancement of 3 autonomous recursive neural nets: for the key sub commitments of talk parsing, explicitly structure expectation and connection forecast; the 1/3 web for assessment forecast. Advantages are: The latent Discourse highlights can help support the execution of a neural emotion analyser. Preparing and the individual models are a request of extent quicker than the multi-entrusting model. Disadvantages are: Difficult expectations to multi-sentential content.

In [10] paper, Twitter Spam has turned into a fundamental downside nowadays. Current works have practical experience in applying machine learning procedures for Twitter spam location that construct utilization of the connected math choices of tweets. and in this manner the execution of existing machine learning based classifiers diminishes. This issue is alluded to as “Twitter Spam Drift” and the anticipated plan can find “changed”. Spam tweets from unlabelled tweets and consolidate them into classifier’s preparation procedure. Various examinations are performed to assess the proposed plan. The outcomes demonstrate that our proposed Lfun plan can essentially enhance the spam identification precision in certifiable situations.

The paper [11] Anomaly recognition is utilized in different applications like discovery of misrepresentation, arrange examination, observing traffic over systems, fabricating and ecological programming. The information streams which are produced are constant and changing after some time. Thesis the motivation behind why it turns out to be about hard to identify the exceptions in the current information which is colossal and ceaseless in nature. This method increases the speed of outlier detection by 20 times and the speed goes on increasing with the increase with the number of data attributes and input data rate.

### III.OPEN ISSUES

The capability of the human face to speak emotional states through facial expressions is widely recognized, and beyond research has hooked up the importance and universality of emotional facial expressions. However, latest evidence has found out that facial expressions of emotion are maximum as it should be diagnosed while the perceiver and expresser are from the equal cultural in institution. Paul Ekman discloses outward appearances to characterize a lot of six all around unmistakable fundamental feelings: anger, disgust, fear, joy, sadness and surprise. Robert Plutchik characterized a wheel-like graph with a lot of eight fundamental, pairwise differentiating emotions. Joy sadness, trust disgust, fear anger and surprise anticipation. Think about every one of these feelings as a different classification, and neglect distinctive dimensions of forces that Plutchik characterizes in his wheel of emotions.

### Disadvantages:

#### A. Ekman’s Facial expressions limitations:

**1. Image quality:** Image top notch impact how properly facial-notoriety calculation artistic creations. The photo appealing of scanning video is very low in relationship with that of a virtual advanced camera.

**2. Image size:** when a face-location calculation finds a face in a picture or on the other hand in a video catch process, the general size of that confront contrasted and the selected picture estimate influences how well the face will be perceived.

**3. Face angle:** The general point of view of the objectives confront impacts the prominence rating significantly when a face is selected inside the acknowledgment programming, by and large two or three edges are utilized (profile, frontal and 45-degree are common).

**4. Processing and storage:** Despite the fact that exorbitant definition video is very low in goals when as contrasted and advanced camera pictures, it despite everything it possesses recognizable measures of plate space. Handling each edge of video is the huge test, so typically just a portion (10 percent to 25 percent) is really go through an acknowledgment framework.

#### B. Plutchik’s algorithm limitations:

1. The FPGA Kit uses hardware that is expensive. Thus, making this approach a cost ineffective technological solution.
2. Also, there is an additional dimension which involves a lot of tedious calculations.

## IV.PROPOSED METHODOLOGY

POMS are a psychological tool [1] for evaluating the people mind set. It characterizes sixty-five adjectives which are evaluated using the concern based on the five-point scale. Every adjective adds one of the six classes. For instance, if person feels irritated them he belongs to the category “anger”. If the score is at the higher level for the adjective, it adds more general score of the respective classes, with the exception of for loose and proficient whose commitments to their respective classes are negative. POMS merge all related ratings into a six-dimensional mood state consisting of classes: anger, depression, fatigue, vigour, tension and confusion. Contrast to the first framework, we disposed the adjective blue, because it just once campers to an emotion and not just colour, and word- sense disambiguation instrument where failed at recognizing the two implications. We additionally evacuated adjectives relaxed and loose, which non positive contributions, as tweets containing there adjective would have counter models.

Contribution of this paper is to implement the novel algorithm a Profile of Mood States (POMS) generating twelve-dimensional mood state representation using 65 adjectives with combination of Ekman's and Plutchik's emotions categories like, anger, depression, fatigue, vigour, tension, confusion, joy, disgust, fear, trust, surprise and anticipation. The machine learning algorithm gives less time consumption without interfere human labelling. The Multinomial Naïve Bayes classifier works on testing dataset with help of huge amount of training dataset. It gives same result as POMS tagging methods. The contribution work is prediction of Emojis for emotion recognition using tweet contents.

### A. Architecture

The Fig.1 shows the proposed system architecture of emotion recognition system.

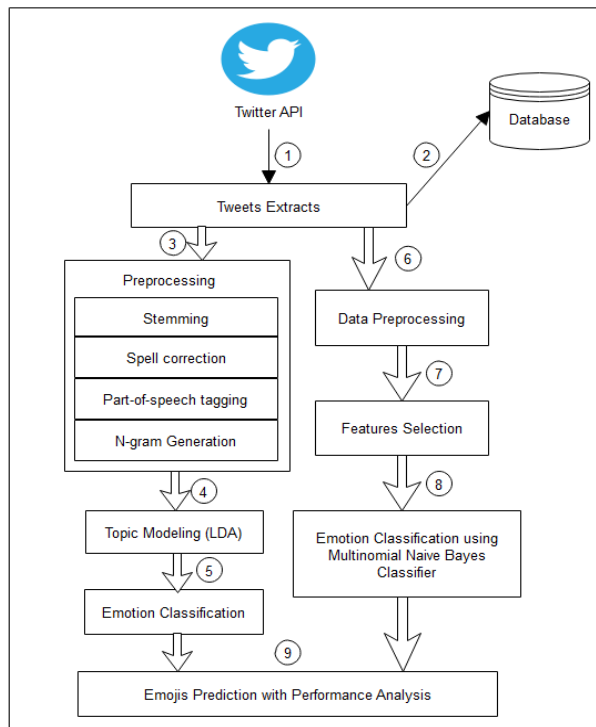


Figure 1. System Architecture

This system is working on Twitter API tweets dataset collecting at first step. There are two parts of the emotion recognition system. First, using NLP language algorithms and second one is working on machine learning classifier algorithms.

The third step is the pre-processing which includes stemming, spell correction using Porter algorithm. The NLP using Part-of-Speech tagging represents the tags of every word which is very helpful for identifying adjectives. The N-gram generation method is to design the similarity score of

tweets using text clustering algorithm. The fourth step is the topic modelling using Latent Dirichlet Allocation (LDA) algorithm for extraction of topics using clustered tweets. Finally identifying the emotion of the tweets with the help of adjectives includes in emotion categories.

The sixth step is the data pre-processing which shows at machine learning algorithm pre-processing step including data transformation and normalization methods. The seventh step is to select the features which required for detection of emotion. The Multinomial Naïve Bayes classifier is an algorithm for solving the problem that arises during the training of Binarized Naive Bayes algorithms. This is used for emotion classification system for multi-class classifier at eighth step. Finally, the ninth step is used to predict the emojis with the help of emojis dataset and analyse the performance of emotion recognition system.

### B. Algorithms

#### 1. Sentiment Analysis using Senti word net Dictionary

```

Polarized Tokens List ← new List ()
While tokenized Ticket Has Next () do
Token ← tokenized Ticket. Next ()
Lemma ← token. Lemma
Polarity Score ← null
If Domain Dictionary. Contains (lemma, pos) then
If Senti Word Net. Contains (lemma, pos) and
Senti Word Net. Get Polarity (lemma, pos) ≠ 0 then
Polarity Score ← Senti Word Net. Get Polarity (lemma, pos)
Else
Domain Dic Token ← Domain Dictionary. Get Token
(lemma, pos)
If domain Dic Token. Polarity Orientation == "POSITIVE"
Then polarity Score ← Default Polarity. Positive
Else
Polarity Score ← Default Polarity. Negative
End if
End if
Polarized Tokens List. Add (token, polarity Score)
End if
End while
Return polarized Tokens List

```

#### 2. Latent Dirichlet Allocation (LDA) Algorithm:

The most vital, LDA offers a generative rendition that depicts how the archive in a dataset was made. In this condition, a dataset is a gathering of  $D$  archives. Record is a social affair of words. The most vital, LDA offers a generative rendition that depicts how the archive in a dataset was made. In this condition, a dataset is a gathering of  $D$  archives. Record is a social affair of words. Let  $\beta_i$  speak to the multinomial for the  $i$ th theme where the span of  $\beta_i$  is  $V$ :  $|\beta_i|=V$ . Particular these appropriations, the LDA generative process is as per the following: Steps:

1. for each document:

- (a) Pick any dissemination among the subjects (a multinomial of length K)
- (b) For each word in the report:
- (c) One of the k should be drawn prospectly which focus from the dissemination are subjects in a, state point  $\beta_j$
- (d) One of the v words should be draw prospectly from  $\beta_j$

### Multi-Nominal Naïve Bayes Classifier Algorithm

Function Train Naïve Bayes (D, C)

Returns  $\log P(c)$  and  $\log P(w|c)$

Steps:

1. For each class  $c \in C$
2. Calculate  $P(c)$  terms
3.  $N_{doc}$  = number of documents in D
4.  $N_c$  = number of documents from D in class c
5.  $\logprior[c] \leftarrow \log \frac{N_c}{N_{doc}}$
6.  $v \leftarrow$  vocabulary of D
7.  $bigdoc[c] \leftarrow$  append( $d$ )
8. For  $d \in D$  with class C
9. For each word w in V
10. Calculate  $P(w|c)$  terms  
 $count(w|c) \leftarrow$   
 $\#$  of occurrences of w in  $bigdoc[c]$
11.  $\loglikelihood[w, c] \leftarrow \log \frac{count(w, c) + 1}{\sum_{w' \in V} (count(w', c) + 1)}$
12. return  $\logprior, \loglikelihood, V$

## V. RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. Some functions used in the algorithm are provided by list of jars like Twitter-core and Twitter-stream jars etc.

Tweets are retrieved in a streaming way, and Twitter provides the Streaming API for developers and researchers to access public tweets in real time. The aim of this paper is to bridge the gap by carrying out a performance evaluation, which was from two different aspects of NLP and machine leaning algorithms. The Unison model is the combination of Ekman's, Plutchik's and POMS emotion categories and the Multinomial Naïve Bayes Classifier algorithm uses for emotion recognition performance. Precision is the ratio of

correctly predicted positive observations to the total predicted positive observations. Recall is the ratio of correctly predicted positive observations to the all observations in actual class. F-measure is the weighted average of Precision and Recall. Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Finally showing the accuracy is as compared to unison model and Multinomial Naïve Bayes Classifier algorithm. And it gives results better than unison model within short time period.

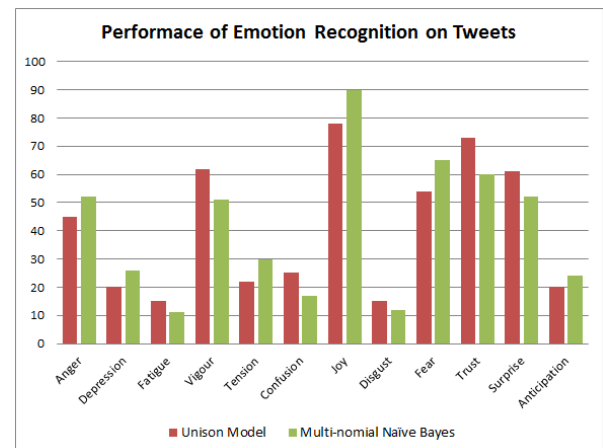


Figure 2. Comparison of emotion identification using Unison Model versus Multinomial Naïve Bayes Classifier Algorithm

TABLE I Performance Analysis between Unison Model versus Multinomial Naïve Bayes Classifier

	Unison Model	Multinomial Naive Bayes
Precision	68.45	78.70
Recall	79.44	72.64
F-Measure	72.11	84.31
Accuracy	80.29	93.26

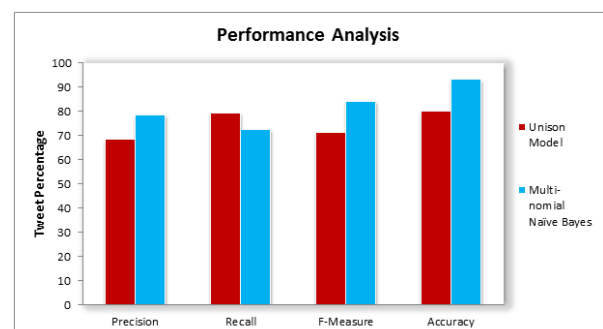


Figure 3: Performance Analysis between existing and proposed system

Fig. 3 shows the performance analysis between Unison Model and Multinomial Naïve Bayes classifier algorithm.

The graph shows the Unison Model increases accuracy as compare to previous algorithms. But, the Multinomial Naïve Bayes classifier algorithm gives better results than Unison Model. And Multinomial Naïve Bayes executes faster than Unison model.

## VI. CONCLUSION

This project implements a novel algorithm Profile of Mood States (POMS) represents twelve-dimensional mood state representation using 65 adjectives with combination of Ekman's and Plutchik's emotions categories like, anger, depression, fatigue, vigour, tension, confusion, joy, disgust, fear, trust, surprise and anticipation. These POMS classifies the emotions with the help of bag-of-words and LSI algorithm. The machine learning Multinomial Naïve Bayes classifier is used to classify emotions, which gives results as accurate and less time consumption compares to POMS. Further work, after getting the emotion of the user, then recommending the tweet posts or motivational speech to the users when they are recognizing any negative emotion category like depression level.

## REFERENCES

- [1] NikoColneric and Janez Demsar, "Emotion Recognition on Twitter: Comparative Study and Training a Unison Model" IEEE Transactions on Affective Computing. PP. 1-1.10.1109/TAFFC. 2018. 2807817.
- [2] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," J. of Computational Science, vol. 2, no. 1, pp. 1-8, 2011.
- [3] J. Bollen, H. Mao, and A. Pepe, "Modelling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena", in Proc. of the 5<sup>th</sup> Int. AAAI Conf. on Weblogs and Social Media Modelling, 2011, pp. 450-453.
- [4] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," Information Processing and Management, vol. 51, no. 4, pp. 480-499, 2015.
- [5] B. Plank and D. Hovy, "Personality Traits on Twitter or How to Get 1,500 Personality Tests in a Week," in Proc. of the 6<sup>th</sup> Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92-98.
- [6] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval," Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912-921, 2015.
- [7] O. Irsoy and C. Cardie, "Opinion Mining with Deep Recurrent Neural Networks," in Proc. of the Conf. on Empirical Methods in Natural Language Processing. ACL, 2014, pp. 720-728.
- [8] S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," Computational Intelligence, vol. 31, no. 2, pp. 301-326, 2015.
- [9] B. Nejat, G. Carenini, and R. Ng, "Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis," Proc. of the SIGDIAL 2017 Conf., no. August, pp. 289-298, 2017.
- [10] S Kamble, SM Sangve, "Real time Detection of Drifted Twitter Spam Based On Features," International Journal of General Science and Engineering Research (IJGSER), ISSN 2455-510X, Vol 4(1), 2018, 21-23.
- [11] Harshad Dattatray Markad, SMS angve, "Parallel Outlier Detection for Streamed Data Using Non-Parameterized Approach," IJSE, Volume 8, Issue 2 July-December 2017.
- [12] M. Farhoodi and A. Yari, "Applying machine learning algorithms for automatic Persian text classification," 2010 6th International Conference on Advanced Information Management and Service (IMS), Seoul, 2010, pp. 318-323.
- [13] E. Tromp and M. Pechenizkiy, "Rule-based Emotion Detection on Social Media:" Putting Tweets on Plutchik's Wheel, arXiv preprint arXiv: 1412.4682, 2014.
- [14] S. Chaffar and D. Inkpen, "Using a Heterogeneous Dataset for Emotion Analysis in Text", in Canadian Conf. on Artificial Intelligence. Springer, 2011, pp. 6267.
- [15] S. Aman and S. Szpakowicz, "Identifying Expressions of Emotion in Text, in Int. Conf. on Text, Speech and Dialogue", vol. 4629. Springer, 2007, pp. 196205.
- [16] G. Mishne, "Experiments with Mood Classification in Blog Posts", in Proc. of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access, vol. 19, 2005, pp. 321327.

## Authors Profile

Ms.Asharani S Dandoti Complted Bachelor of Engineering from VTU. I am currently pursuing ME Computer Engineering at Zeal College of Engineering Pune. and I have 5years of experience in teaching. My main research work focus on Datamining,machine Learning.

Dr. Sunil M Sangve has total 20 years of experience in teaching and research. He has completed his P.hd in Computer science and Engineering from SRTM University Nanded.

He has more than 20 research publications in reputed international journal and conferences.

He is a life member of ISTE since 2004. His research work focuses on Network Security, Data Mining. Currently he is working as a professor at Dept of Computer Engineering , ZCOER Pune.