

# SENTIMENT ANALYSIS TEXT TYPE FEEDBACK SYSTEM USING ARTIFICIAL INTELLIGENCE

Cibi Chandra, Anil Kumar

PG Scholar, Dept. of Computer Science and Engineering, SVS college of Engineering Coimbatore, Tamilnadu, India  
Assistant Professor, Dept. of Computer Science and Engineering, SVS college of Engineering Coimbatore, Tamilnadu, India

\*\*\*

**Abstract** - Feedback is used to analyze the performance in various organizations like academic, agricultural and many other organizations. At present most of the systems use grading technique to evaluate the performance. But this method always does not reveal the sentiments of the reviewer. For example, in students' feedback in an academic system, if grading is used it does not reveal the student sentiments towards a particular individual.

The same scenario exists in agricultural field also. The farmers can't give their feedback about the agricultural Rice yield present in the country through grading. Thus, textual feedback is always found to be effective. In this work we propose a textual feedback to analyze the sentiments of the reviewer in agriculture as well as academic system.

Sentimental analysis is a method for identifying the sentiment expressed in texts. The need of Sentiment Analysis of text has gained more importance in today's situations faced by the people of the world. In machine learning technique, it uses unsupervised learning or supervised learning. Classification problem can be carried out using several algorithms like support vector machine, naïve Bayes, random forest. In lexicon-based method sentiment polarity of the textual content is detected using sentiment lexicon. A lexicon is a list of words with associated sentiment polarity. Hybrid approach is a combination of lexicon-based and machine learning methods. The training data set is labelled using sentiment lexicon and this is used to for the machine learning model. Then testing data is evaluated using this model. MATLAB software is used for this setup

**Key Words:** Naïve Bayer, Support Vector Machine,

## 1. INTRODUCTION

Social Media and Micro blogging platforms like Face book, Twitter, and Tumbler dominate in spreading encapsulated news and trending topics across the globe at a rapid Space. Large organizations take feedback about their products and services to increase their marketing

Sites Twitter, Face book, Instagram, google+ offer a platform to people to voice their opinions. people quickly post there. This type of vast information on these sites can use for marketing and social studies. Therefore, sentiment analysis has wide applications and includes emotion mining, polarity, and classification and influence analysis.

Feedback is the statement sent to an entity about its past behavior from which the entity can analyze the future and current behavior to achieve the expected result. Feedback plays an important role in education and learning by helping to adopt new knowledge and prevent repetitive mistakes. Feedback is a process which helps the organization to monitor, evaluate, and regulate the overall working environment. Good feedback practice provides useful information to the organization in improving the teaching and learning experience. Similarly, the agricultural field can be improved by taking regular feedback from farmers.

In a more practical sense, our objective here is to take a text and produce a label (or labels) that summarize the sentiment of this text, e.g. positive, neutral, and negative. For example, if we were dealing with agricultural reviews, we would want the sentence 'the incentives to help farmers were helpful' to be labeled as Positive, and the sentence 'The amount for relief to farmers in budget is not enough' is' labeled as Negative.

Depending on the feedback given by students or from famers in agricultural field it can be classified as textual or grading (Liker-scale based score) form. In Liker-scale based score questions are provided to the students and are asked to answer those questions using a rating-based scale. the sentiment of the student is known using textual feedback technique. In this textual form student are given with set of questions and they need to answer it in sentences. It is helpful to both the academic administration and instructor to overcome the issues related to their organization. In this paper, the student feedback with varied opinion is collected. The aim is to extract expressions of opinion and classify it as negative, positive or neutral using machine learning techniques. Similarly, in agricultural field the feedbacks are collected from farmers by government officials. Various necessities of farmers are identified by analyzing these feedbacks. In this work the data is collected from farmers using questioners.

Methods like Artificial neural network, machine learning algorithms such as Support Vector Machine, Multinomial Naïve Bayes Classifier, and Random Forest are used for classifications. The results of classifiers are compared to find the best method for classification. The result shows that Artificial Neural Network has better performance over the other methods.

## 2. Methodology

The student/farmers feedback collected is an input data. The data collected for agriculture ie rice production of india ministry of agriculture .For student feedback we use a website i.e. training data using which the system is trained. On receiving test samples, the trained system classifies the sentence as negative, neutral, and positive classes using machine learning algorithms. This outcome is graphically represented. The proposed methodology consists of six steps: preparing training data, collecting student feedback, feature extraction, model training, evaluation of test data, and graphical representation of the result as shown.

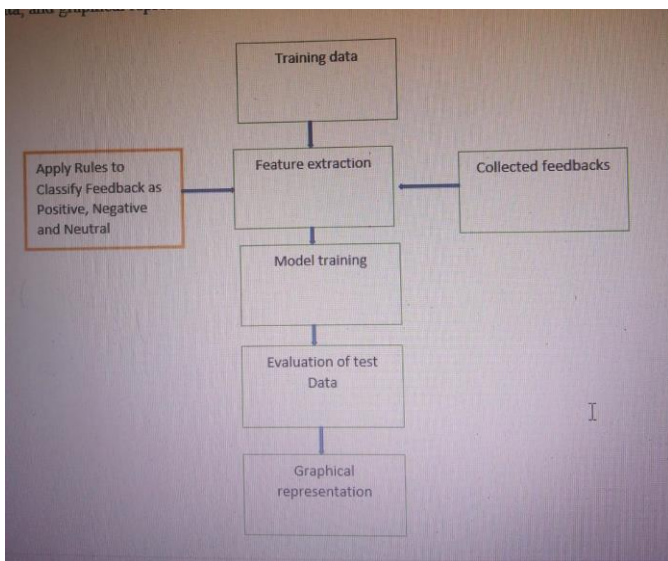


Fig4. 1 Methodology for sentiments analysis

### 2.1 Preparing Training Data:

Machine learning is classified into unsupervised and supervised learning. In supervised learning the sentences are given with label of classes, whereas labels are not provided in unsupervised learning. Training data is collected for each question which is used to train the system.

This training data is the responses in the form of sentence from the students. Each sentence is converted to lower case and is given as input to the Sentiment Intensity Analyser() method which is available in package VaderSentiment.VADER(Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis returns a sentiment score in the range -1 to 1 from negative to positive. Individual words have sentiment score between -4 and 4. The sentiment score (Compound Score) of a sentence is calculated by summing up the sentiment scores of each VADER-dictionary-listed word in the sentence and normalized to map it to the value between 1 and -1 [10]. The normalization used by Hutto is given below.

$$x/(x^2 + \alpha)$$

Where x is the sum of the sentiment scores of constituent words of a sentence and  $\alpha$  is the normalization parameter set to 15.If the compound score is  $\geq 0.05$  the sentence is classified as positive, if the compound score is greater than -0.05 and less than 0.05 the sentence is classified as neutral, if the compound score is  $\leq -0.05$  the sentence is classified as negative.

### 2.2 Collecting Student Feedback

The questions are sent to students through form and the student responses are collected and stored in Microsoft excel comma separated values file. forms containing questions.

### 2.3 Feature Extraction

This process involves feature extraction in a format required by machine learning algorithms from the datasets which consists of formats such as image and text. Both train and test data are subjected to feature extraction. Text documents can be tokenized and vocabulary of known words is built using Count Vectorizer.

### 2.4 Model Training

There are many different algorithms for text classification with machine learning. Multinomial Naïve Bayes Classifier (MNBC) the basic idea involved in naive bayes classification technique is to find the classes probabilities assigned to texts by using joint probabilities of classes and words.The features/predictors used by the classifier are the frequency of the words present in the document.

SVMs are supervised learning models that examine data used for regression analysis and classification. A support vector machine constructs a hyper plane or set of hyper planes in an n-dimension space, which can be used for classification, regression or another tasks are used to categorize the data points by mapping data to a high dimensional feature space, even when the data is not linearly separable. First a separator between the categories is found, and then the data is transformed so that hyper plane is drawn as a separator. At the end, new data characteristics are used to predict the class for a new record.

Random Forest is a supervised learning method. It can be used for both classification and regression task. Random Forest algorithm creates a forest with number of trees.

### 2.5 Evaluation of test data

evaluation is the process of collecting and analyzing data, where the result can be used to determine whether an organization is effectively carrying out the planned activity. This corresponds to the final evaluation that the model goes through after the training phase has been completed. This step is critical to test the generalizability of the model by using test set we can get the working accuracy of the mode

### 3. EXPERIMENTAL SETUP

The first step in any of the Machine Learning tasks is to analyze the data. second step is to extract title and body There is no specific way to do this, this totally depends on the problem statement at hand and on the analysis, we do on the dataset. Pre-processing is one of the major steps when we are dealing with any kind of text models. Few mandatory pre-processing is converting to lowercase, removing punctuation, removing stop lemmatization/stemming. In our problem statement it seems like the basic pre-processing steps will be sufficient. During the text processing each sentence is split to words and each word is considered as a token after pre-processing Converting to lowercase is a very mandatory pre-processing step. Stop words we do not give any additional value to the document vector. There might be few problems such as U.S — us “United stated” being converted to “us” after the pre-processing we use numpy here because our data is stored in list of lists. So, what we are doing is we are first removing the stop words, and then symbols and then finally stop words because few words might still have apostrophe which are not stop words. This is the final and most important part of the pre-processing. For example, reading and reads are the same type of words which basically indicate an action read. we use a library called porter-stemmer. Porter-Stemmer removes the suffix or affix of a word. Lemmatization is a way to reduce the word to root synonym of a word. Unlike Stemming, Lemmatization makes sure that the reduced word is again a dictionary word. Word Net Lemmatizer is used to lemmatize word. When user gives a query 500 dollars or five hundred dollars. both these search terms are same. IR model treats them separately, as we are storing 500, dollar, five hundred as different tokens. to make IR model we need to convert 500 to five hundred. we use a library called num2word. The calculation of TF-IDF work in this case Giving different weighs to title and body.we need to calculate the TF-IDF for body and for title.We are going to calculate TF-IDF for body, multiply the whole-body TF-IDF values with alpha, iterate the tokens in title, and replace the title TF-IDF value in the body TF-IDF value of the (document, token) pair exists.

$$TF-IDF = body\_tf-idf * body\_weight + title\_tf-idf * title\_weight$$

$$body\_weight + title\_weight = 1$$

Matching score is to calculate the similarity, in this method; we add tf\_idf values of the tokens that are in query for every document. For example, if the query “hello world”, we need to check in every document if these words exist and if the word exists, then the tf\_idf value is added to the matching

score of that particular doc\_id. at last we will sort and take the top k documents. We use total\_vocab variable which has all the list of unique tokens to generate a index for each token, and And numpy of shape (docs, total\_vocab) to store the document vectors. For vector, we calculate the TF-IDF values, TF we can calculate from the query itself, and we can make use of DF for the document frequency, and finally we will store in a (1,vocab\_size) numpy array to store the tf-idf values, index of the token will be decided from the total vocab list

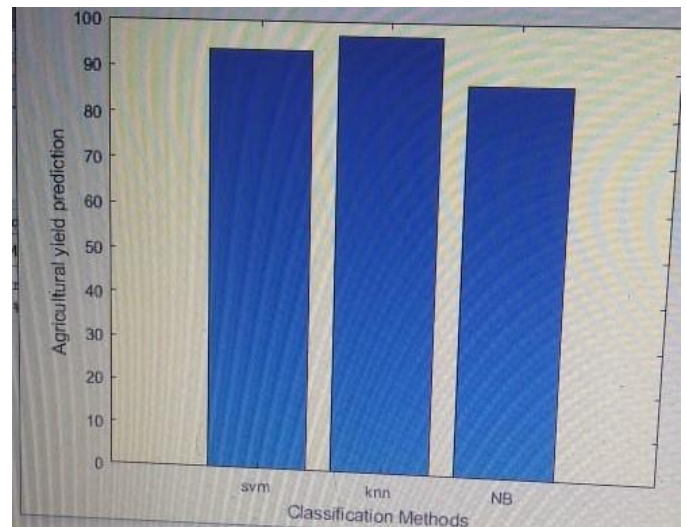


Chart -1: Agricultural Yield Production

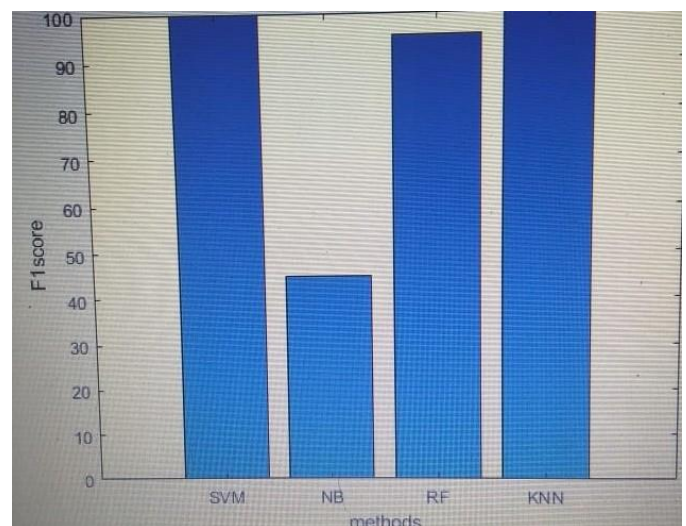


Chart-2: Academic feedback

### 3. CONCLUSIONS

Here a method is proposed to analyze the sentiments of members from various institutions. We specifically analyze the feedback from students of an institute also from farmers in a village. The data collected is employed for analysis.

Different machine learning techniques like SVM, NB, DT etc are employed and the accuracy of feedback analysis is analyzed. Then a recurrent Neural Network is employed for analysis. We propose the neural network has better performance over the existing machine learning techniques mentioned above.

## REFERENCES

- [1] A. P Jain, P. Dandannavar, "Application of Machine Learning Techniques to Sentiment Analysis", Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, pp. 628-632, 2018.
- [2] B. K. Bhavitha, A. P. Rodrigues, N. N Chiplunkar, "Comparative Study of Machine Learning Techniques in Sentimental Analysis". International Conference on Inventive Communication and Computational Technologies, pp.216221, 2017.
- [3] G. Qiang, "Research and Improvement for Feature Selection on Naive Bayes Text Classifier", 2nd International Conference on Future Computer and Communication, pp. V2-156-V215 9, 2010
- [4] H. P. Patil and M. Atique, "Sentiment Analysis for Social Media: A Survey," 2nd International Conference on Information Science and Security (ICISS), Seoul, pp. 14, 2015
- [5] J. Singh, G. Singh and R. Singh, "Optimization of sentiment analysis using machine learning classifiers", Human-centric Computing and Information Sciences, Volume 8, Number 1, 2018.
- [6] K. S. Krishnaveni, Rohit R Pai, V. Iyer, "Faculty Rating System Based on Student Feedbacks Using Sentimental Analysis", International Conference on Advances on Computing, Communication and Informatics, pp.1648-1653, 2017.0
- [7] M. A. Ullah, "Sentiment analysis of students feedback: a study towards optimal tools", International Workshop on Computational Intelligence (IWCI), pp. 17-180, 2016.
- [8] ] M. D. Devika, Sunitha, A. Ganesh "Sentiment Analysis: A Comparative Study on Different Approaches", Fourth International Conference on Recent Trends in Computer Science & Engineering, Procedia Computer Science 87, pp. 44 – 49, 2016.
- [9] T. M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.
- [10] W. Medhat, A. Hassan, H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, pp. 1093–1113, 2014.