

# A Novel Approach for Analysis of Feedback in Banking System using Machine Learning Algorithm

Maheswaran M

Student Scholar, E&CE department,  
Coorg Institute of Technology,  
Ponnampet, India

Benaka Santhosh S

Assistant Professor, E&CE Dept.,  
Coorg Institute of Technology,  
Ponnampet, India

**Abstract** - A novel approach for analysis of feedback in banking system using machine learning algorithm is introduced, opinion mining and sentiment analysis techniques are used to search out the emotion of the customers input on staffs about their working speed, time sense and customer support etc... This examination includes the apparatus of a blend of AI and common language preparing systems on underway input data accumulated from module investigation overview consequences of banking system. Additionally, to offer a grade by grade clarification of the technique of accomplishment of opinion mining on or after customer's remarks using the open source tool like Python, take a look of comparative overall performance study, take a look of remarks and some extracted alternatives etc. The consequences are compared and analyzed to find out the higher overall performance with relevance. Several evaluation standards are designed for various techniques.

**Keywords** - *AI (Artificial Intelligence), ML (Machine Learning), Opinion mining, Sentiment analysis, Supervised learning, Natural Language Processing*

## I. INTRODUCTION

In present lifestyle banks are facing lots of problems with their customers like bad loans, customer dissatisfaction etc. and also huge competition in day to day life, due to these factors so many small sector banks are closed permanently and merged to a large sector banks. In order to make it a profitable it is very much important to analyze the mindset of customers about their bank frequently. The bank must consistently analyze their customer's feelings by gaining feedback from their customer's asking to fill some forms like feedback, review, compliment forms etc. In order to achieve the positive result to lead the bank in profitable side, it is very important to analyze huge number of data (feedback) given by their customer and it is very huge problem by the humans because human can make some human errors while analyzing large data to overcome this error it is better to feed the data to a machine learning algorithm, and once the data is trained with high accuracy it is better to analyze the customer's feedback better than human. In order to achieve this with the help of machine learning we have introduced a positive approach with satisfying accuracy.

To study and analyze these large data we are using some Machine learning algorithms like k-NN, SVM, Naïve Bayes. From among the numerous systems for aggregating the customer's comments, surveys contain a vital position and most

of the establishments commence surveys in several paperwork. The focal point is on utilizing opinion mining, data mining, and sentiment analysis approach for categorizing the customer's feedback received all through component estimate survey. The collected data is mined using some techniques like data mining, opinion mining and sentiment analysis. The mined and preprocessed datasets be subjected to the models SVM, Naïve Bayes (NB), k- Nearest Neighbors (k-NN) applied with the usage of Python. By using these (Machine Learning) ML models we are classifying the data into positive, negative and also neutral data's, by training the mined data and fed to the ML model, after feeding the data to the ML model, check the accuracy of the model which we used to model the algorithm.

The proportional effectiveness of these algorithms in the selected utility circumstance are evaluated and accuracy is taken into account. Accuracy is that the quantitative relation of true positives towards the whole type of positives dataset. The work aims to dig deeper into the feedback information of an establishment. Presently the feedback information is employed to report solely the performance of the staff's. The paper proposes ways to research the feedback information victimization data processing techniques for a higher understanding of the bank or financial supporters, customers, and bank staffs. The format of feedback varies from establishment to establishment, thus there cannot be a general technique which will appropriate all. The feedback information from the customer's is analyzed by victimization completely different data processing techniques. The feedback given by the customers is used for analyzing all the parameters regarding bank environment, staffs support etc. This Paper surveys maximum data processing techniques that are applied for analyzing feedback information.

## II. LITERATURE SURVEY

The paper [1] focused on victimization Opinion Mining method for classifying the customer's comments acquired during the survey. From survey of this paper the authors gave some knowledge about what are the different ways for data preprocessing techniques like extraction, stop words elimination, stop word removal method, stemming, and tokenization, etc. The data which is gained from the customer's via feedback form are preprocessed before feeding to a machine learning algorithm to obtain expected result. Data must be preprocessed which means the data should be cleaned by

removing punctuations and unnecessary words, symbols etc. The way to remove these junks from the data, is data preprocessing. To achieve a good accuracy the data must be mined with high care. The preprocessing techniques are analyzed with the help of trial & error method with that knowledge the authors are proceeding towards further steps. In our project we are using only limited data preprocessing techniques which are suitable for analyzing the data given by customer's and also data preprocessing techniques which are used taken into much deeper, this will helps to clean the data with high accuracy level.

The author in paper [2] conveys that how to extract keywords from the data. These keywords are the subset of words that contains the most important information of the data or document. Stop words elimination is also explained in paper [2]. The most common words are in text documents are prepositions, articles, and pro-nouns etc. that does not provide the meaning of the documents. These words are treated as stop words. In our project we are removing few stop words which are common words that doesn't change the meaning of the data. By removing stop words the data become much cleaner and precise which helps our model to analyze the data easily.

In paper [3] authors are projected analysis of students feedback concerning to gain a formal approach towards the students feedback which uses some of the machine learning algorithm models such as SVM, Naïve Bayes (NN), and ANN (Artificial Neural Network) with an accuracy level up to 89.7% but in our project the accuracy is increased by using versatile algorithms of machine learning. The models are selected with the help of trial and error method, and chosen the best suited model for the mined data to get a good result. We have chosen three models namely k-NN, SVM, Naïve Bayes from these three models, in SVM using poly kernel we got a good accuracy level of 94.6% on classifying the mined data.

### III. METHODOLOGY

The data is collected from the customers by using webpage forms, or some other way which can easily reach out the customer's place. The collected data consists of textual statements, and with some other impurities like extra symbols, unwanted spaces. Later, the collected data is subjected to data mining i.e., data preprocessing which helps to get a good accuracy level. For our project we are using Comma separated values data structure. Datasets are of two types, they are linear dataset and nonlinear dataset. The linear dataset is the one which is having equal properties whereas the non-linear dataset is the one which is having non-equal properties. Machine learning works well for linear datasets. The linear dataset contains reviews given by the customer's.

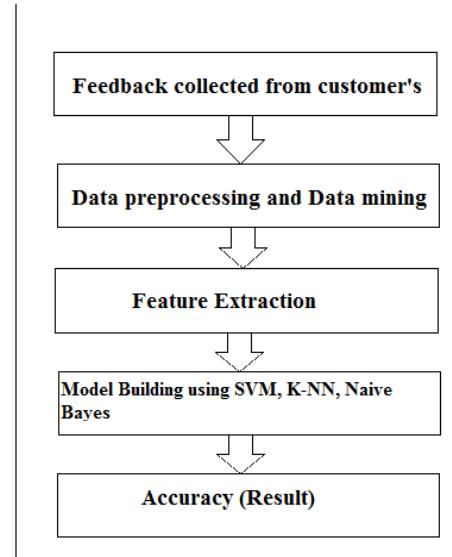


Fig .1 Methodology/Flow diagram

The dataset contains nearly 1000 reviews with their customer\_id, customer\_name, review. In that review contains the feedback about the staff's present in the bank in the form of textual statements.

Pre-processing is the progression of concentrated effort the data from redundant elements. It enlarges the accurateness of the results by dropping errors in the data. Not by means of pre-processing, such as enchantment corrections, may lead the system to disregard important words. Pre-processing and concentrated effort of data are one of the most important tasks that must be one before dataset be able to be used for machine learning. The real-world statistics is incomplete and incompatible. So, it is necessary to be cleaned. There are many general pre-processing techniques, of which the majority common is: tokenization, lemmatization, convert text to lower or upper case, eliminate punctuation, take away numbers, take away unwanted white space, take out repeated letters, and get rid of stop words, stemming and negation. The data that is obtained after the pre-processing is done is given to the Machine learning algorithm.

Later the cleaned data is subjected to a Machine learning model. The algorithm which is used to build the machine learning model presents with their confusion matrix and accuracy rate. The output of this model results in a separate new column which shows the predicted output by the model.

### IV. EXPERIMENTAL RESULTS

In this work, python (jupyter notebook) is used to discover how well the algorithm's mechanism on the feedback given by the customers based on the accuracy values.

The datasets which we taken contains customers feedback which needs to preprocessed by removing certain unwanted things such as stop words, extra symbols etc., and also stemming of words. The preprocessed data is subjected to the Naïve Bayes Algorithm

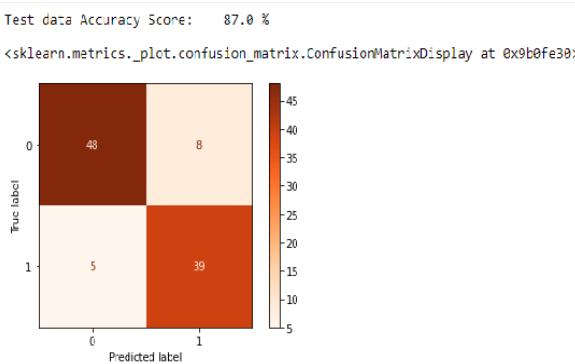


Fig 2(a) Result of Naïve Bayes algorithm

The Fig 2(a) represents that the model has 87% accuracy for the random values in the dataset. It shows that the model can calculate with satisfactory precision, recall for the classes of randomly chosen feedbacks from the given datasets.

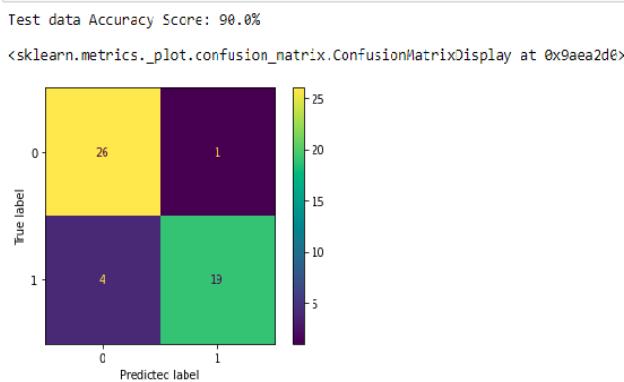


Fig 2(b) Result of K-NN algorithm

The Fig 2(b) describes that the accuracy of 90.0% by using the K-Nearest Neighbor algorithm. The accuracy of the K-NN representation is much better than that of the Naïve Bayes model.

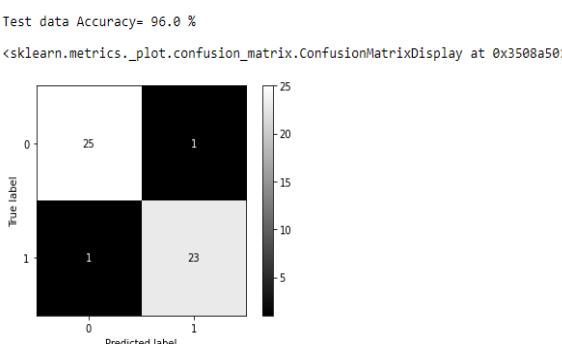


Fig 2(c) Result of SVM algorithm

The Fig 2(c) represents that the accuracy of 96.0% is obtained by using the support vector machine algorithm. It is clearly shown that the SVM algorithm have outperformed all other machine learning algorithms. So this model is used for further operations.

| Algorithm Models Used  | Accuracy Obtained |
|------------------------|-------------------|
| Naïve Bayes            | 87%               |
| K- Nearest Neighbor    | 90.0%             |
| Support Vector Machine | 96.0%             |

Table 1: Accuracy score from the algorithm

## V. CONCLUSION AND FUTURE WORK

Sentence level sentiment analysis become accustomed removes the remarked normal alternatives and evaluation words from the contribution dataset. A customer Feedback Mining System is work to inquire about staff's performance. This strategy will be helpful to improve bank manager to gain knowledge about their staff. Automating the customer's feedback may give several advantages together with saving time and creating economical report generation, etc. the utilization of opining mining will facilitate in summarizing the feedback report effectively and evaluating bank performance in the type of a summarized read might be helpful for the establishments. Several machine learning algorithms (models) are audited and referred to. Our results proves that the Support Vector Machine algorithm dominates all other algorithms which we used with its accuracy level.

In future work, we are trying to resolve the problems that occurred during data mining which means misplace of information during data cleaning and also loss of information while removing some of the words. The opinion mining are to be implemented more precise and accurate. The models are also trying to get a further more high accuracy level which needs a cleaned data and also it must be flexible. Moreover it is a web-based system the data is collected more precisely in further improvements we are trying to implement which can update customer name, age, and gender which will helps the manager or administration to improve their quality of service.

## VI. REFERENCES

- [1]. Vijayarani, Ilamathi, and Nithya "Preprocessing Techniques for Text Mining-An Overview", International Journal of Computer Science & Communication Networks, Vol 5
- [2]. Menaka S, Radha N "Text classificatioion Using Keyword Extraction Technique" International Journal of Advanced Research in Computer Science and Software Engineering
- [3]. Sharnitha Katragadda "Performance Analysis on Student Feedback using Machine Learning Algorithms" 6th International Conference on Advanced Computing & Communication Systems 2020(ICACCS)
- [4]. Sarpreet Kaurl, Rasleen Deol, "Students Feedback for Mining Their Opinions Using Supervised Learning Algorithm" International Journal of Engineering Science and Computing, June 2017
- [5]. Anushree Chitriv, Prof. A. Thomas "Analysis of Student Feedback by Using Data Mining Techniques", Helix 2018
- [6]. Khan, Khairullah, "Mining opinion components from unstructured reviews: A review," Journal of King Saud University – Computer and Information Sciences, Vol. 26, 2016.
- [7]. Trisha Patel, "Sentiment Analysis of Parents Feedback for Educational Institutes", International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 3, 2017.
- [8]. SungwhanMacKim and Rafael, "Sentiment Analysis in Student Experiences of Learning", Available at ResearchGate.com
- [9]. V. Dhanalakshmi; Dhivya Bino "Opinion mining from student feedback data using supervised learning algorithms", 2016 3rd MEC International Conference on Big Data and Smart City [ICBDSC]