



MACHINE LEARNING BASED SENTIMENT ANALYSIS FOR FEEDBACK DATA

Dr.Nancy Jasmine Goldena¹ and J.Fathima Irfana²

Department of Computer Applications, Sarah Tucker College, Tirunelveli-7.

ABSTRACT

Feedback taken from students will help teaching-learning more effectively. This also helps to understand the learning behavior of students. Feedback is usually taken at the end of the unit. The feedback of the student must be taken in real-time which helps to know the learning behavior of the student. The feedback from the students is collected using Google Forms. The data used by us for study purposes is taken from STC college students. Sentiment analysis is the process of identifying and detecting subjective information using natural language processing, text analysis, and computational linguistics. In short, sentiment analysis aims to extract information on the attitude of the writer or speaker towards a specific topic or the total polarity of a document. It is the process of determining the opinion or tone of a group of words. It helps us gain attitude and motion behind a series of the word. Knowing about the user's feedback can come to a greater aid in knowing the user as well as improving the organization. Here an example of student's data is taken for study purposes. Analyzing the student feedback will help to help to address student-related problems and help to make teaching more student-oriented. In this work, the Naïve Bayes algorithm is used to predict the feedback data. This project is developed using MatLab.

1. INTRODUCTION

Sentiment analysis (also called: sentiment mining, sentiment classification, opinion mining, subjectivity analysis, review mining or appraisal extraction, and in some cases polarity classification) can deal with the computational handling of subjective, sentiment, and opinion in the text. It plans to realize the attitude or opinion of a writer for a certain topic or goal. Opinion mining is very much essential in e-commerce websites, furthermore advantageous with an individual. An ever-increasing amount of results are stored on the web as well as the number of people would acquiring items from the web is increasing, as a result, the users' reviews or posts are increasing day by day. The reviews toward shipper sites express their feeling. Any organization, for example, web forums, discourse groups, blogs, etc., there will be an extensive add-up for information. Records identified with items on the Web, which are functional to both makers and

clients. The process of finding user opinion about the topic or product or problem is called opinion mining. It can also be defined as the process of automatic extraction of knowledge through opinions expressed by the user who is currently using the product about some product is called opinion mining. Analyzing the emotions from the extracted opinions is defined as Sentiment Analysis. The goal of opinion mining and Sentiment Analysis is to make the computer able to recognize and express emotion. This work concentrates on mining reviews from the websites like Amazon, which allows users to freely write the view. It automatically extracts the reviews from the website. It also uses a machine-learning algorithm to classify the review as a positive and negative review. In the end we have used quality metric parameters to measure the performance of each algorithm

2.LITERATURE REVIEW

In [1], the authors have proposed Feedback Evaluation is a necessary part of any institute to maintain and monitor the academic quality of the system. Traditionally, a questionnaire-based system is used to evaluate the performance of teachers of an institute. Here, they propose an automatic evaluation system based on sentiment analysis, which shall be more versatile and meaningful than the existing system. In their proposed system, feedback is collected in the form of running text, and sentiment analysis is performed to identify important aspects along with the orientations using supervised and semi-supervised machine learning techniques. However, the information we get may be very subjective or person-based. To get correct information about products and services, we need a large number of opinions.

In [2], they have presented sentiment analysis and machine learning principles to find the correlation between "public sentiment" and "market sentiment". They used Twitter data to predict public mood and use the predicted mood and previous days' DJIA values to predict the stock market movements. In that paper, they tested a hypothesis based on the premise of behavioral economics, that the emotions and moods of individuals affect their decision-making process, thus, leading to a direct correlation between "public sentiment" and "market sentiment". They performed sentiment analysis on publicly available Twitter data to find the public mood and the degree of membership into 4 classes - Calm, Happy, Alert, and Kind (somewhat like fuzzy membership).

In [3], they have proposed to help higher education institutions make the best use of their student feedback. This guide is based on a HEFCE-funded project undertaken by the Centre for Higher Education Research and Information (CHERI). The purpose of this Guide is to help higher education institutions make the best use of their student feedback. All institutions collect feedback from their students and in many different forms. They use it to improve the quality of the education they provide. In recent years, there has been a shift in the balance between informal and formal types of student feedback with a greater emphasis on the latter. Now, new devolved forms of national quality assurance promise to give an important role to students and there is also an expectation that information from student feedback will be used to inform the choices of students when applying to higher education. Thus, as the importance attached to student feedback increases, ensuring that feedback is collected effectively and used wisely becomes an increasing priority for

higher education institutions. This Guide draws on the experiences of the sector to highlight some of the good practices that exist as well as some of the problems that institutions are experiencing in using student feedback. Its focus is upon the use of student feedback to enhance the quality of teaching and learning. Other purposes are acknowledged but are not the main emphasis of this publication. Data mining-related survey present in [4].

In [5], the authors have proposed that substantial efforts have been made recently to compare the effectiveness of traditional course formats to alternative formats (most often, online delivery compared to traditional on-site delivery). This study examines, not the delivery format but rather the evaluation format. It compares traditional paper and pencil methods for course evaluation with electronic methods. Eleven instructors took part in the study. Each instructor taught two sections of the same course; at the end, one course received an online course evaluation, the other a traditional pencil and paper evaluation.

In [6], the authors have proposed to develop a teacher's performance evaluation tool using opinion mining with sentiment analysis. The study may help to identify the strengths and weaknesses of the faculty members based on the positive and negative feedback of the students either in English or in Filipino language. The proposed system provides the sentiment score from the qualitative data and numerical response rating from the quantitative data of teachers evaluation. It will also graphically represent the evaluation result including the percentage of positive and negative feedback of the students. Thus, the school administrators and educators will be more aware about the sentiments and concerns of the students.

In [7], they have proposed Universities collect qualitative and quantitative feedback from students upon course completion in order to improve course quality and students' learning experience. Combining program-wide and module-specific questions, universities collect feedback from students on three main aspects of a course namely, teaching style, content, and learning experience. The feedback is collected through both qualitative comments and quantitative scores. Current methods for analyzing the student course evaluations are manual and majorly focus on quantitative feedback and fall short of an in-depth exploration of qualitative feedback.

In [8], the authors have proposed Twitter is a micro blogging site in which users can post updates (tweets) to friends (followers). It has become an immense dataset of the so-called sentiments. In this paper, we introduce an approach to selection of a new feature set based on Information Gain, Bigram, Object-oriented extraction methods in sentiment analysis on social networking side. In addition, They also proposed a sentiment analysis model based on Naive Bayes and Support Vector Machine. Its purpose is to analyze sentiment more effectively. This model proved to be highly effective and accurate on the analysis of feelings.

In [9], the authors have proposed Students' real-time feedback has numerous advantages in education, however, analysing feedback while teaching is both stressful and time consuming. To address this problem, they proposed to analyse feedback automatically using sentiment analysis. Sentiment analysis is domain dependent and although it has been applied to the educational domain before, it has not been

previously used for real-time feedback. To find the best model for automatic analysis we look at four aspects: pre-processing, features, machine learning techniques and the use of the neutral class. We found that the highest result for the four aspects is Support Vector Machines (SVM) with the highest level of preprocessing, unigrams and no neutral class.

In [10], the authors have described a rule-based sentiment analysis algorithm for polarity classification of financial news articles. The system utilizes a prior polarity lexicon to classify the financial news articles into positive or negative. Sentiment composition rules are used to determine the polarity of each sentence in the news article, while the Positivity/Negativity ratio (P/N ratio) is used to calculate the sentiment values of the overall content of each news article. The performance of the Sentiment Analyser was evaluated using a dataset of manually annotated financial news articles collected from various online financial newspapers.

3. METHODOLOGY

MODULES

Data Collection

Preprocess

- Tokenizing
- Data cleansing
- Removal of stop words

Feature extraction

- TF, IDF, N grams
- Group words as Positive or Negative

Sentiment Analysis

Final call on positive/negative review

PRE PROCESS

We have collected data for study purpose form STC College, Tirunelveli. The data has feedbacks collected from students regarding teaching of an organization. The feedback data has collected from students there are 121 records. The tool used for analysis is “Microsoft Excel”. We uses sentiment analysis for analyzing the feedback ,Recently, several websites encourage researchers to express and exchange their views, suggestions and opinions related to scientific papers. Sentiment analysis aims at determining the attitude of a writer with respect to some topics or the overall sentiment polarity of a text, such as positive or negative. Sentiment analysis depends on two issues sentiment polarity and sentiment score. Sentiment polarity is a binary value either positive or negative.

As we have used excel as backend tool for analysis similarly we can use Python tool as front end. The methodology we used here is sentiment analysis, the sentiment analysis can also be applied on various twitter data to know about various trends, polarity and we can analyze public opinion, it also can be applied on various reviews data, to study customer surveys , public talks on a product. On other hand, we find another approach in the mining of sentiment is on the web. Web opinion mining aims at extracting summarize, and track various aspects of subjective information on the Web. This can prove helpful for advertising companies or trend watchers. By a synopsis of Sentiment analysis defection (also called as opinion mining) that refers to the use of natural language processing(NLP), text analysis (TA) and computational linguistics (CL) to identify and extract subjective information in source materials.

- With actual values obtained, the prediction results are compared.
- The various quality parameters are evaluated and the prediction results are compared.
- Pre-processing involves removal of unimportant features from the data. In this phase, several techniques like Stemming and Stop word removal are applied to data set for noise reduction and facilitating feature extraction.

DATA SOURCE

For any consumer to make an effective decision on buying a product, online product reviews serve as excellent sources. The consumer can get product related information through online reviews. Star ratings of a product serve as excellent cues for decision making as they provide a quick indication of a review.

Data preparation

The sentiment data set used in this work contains a set of product review sentences which were categorized as positive, negative or neutral class. The neutral class is important because not all sentences have sentiments. The neutral class is not only considered as a state between positive and negative classes. Also in binary classes, we force the words to be classified as either positive class or negative class leaving no room for neutrality. This leads to over fitting and becomes vulnerable to situations where due to randomness, a particular neutral word occurs more times in positive or negative class examples. For constructing the dataset, we collected staff, students, parents and alumni feedbacks of our college. Staff feedbacks have been chosen to perform our proposed research. In a total of 220 reviews, sentences are crawled using staff feedbacks and parser.

Then the feedbacks are extracted and sentiment class is assigned based on the review score in the data format (sentiment class is positive if review score is >3 , if review score is < 3 the class label is assigned negative or else the class label is assigned neutral). Amongst 121 feedback sentences, 9 are positive reviews, 13 are negative reviews and 99 are neutral review sentences.

Feature Extraction

The product attributes in the review sentences are to be collected by part of speech (POS) tagging by considering each of positive, negative and neutral class reviews. The main aim is to extract the features about a particular product and to analyze the features which people like or dislike. It is also important to identify the product features that people are interested in discussing. These product features contribute to the crucial step. For example, consider a sentence from the reviews of a digital camera:

“The staff’s Teaching is very clear.”

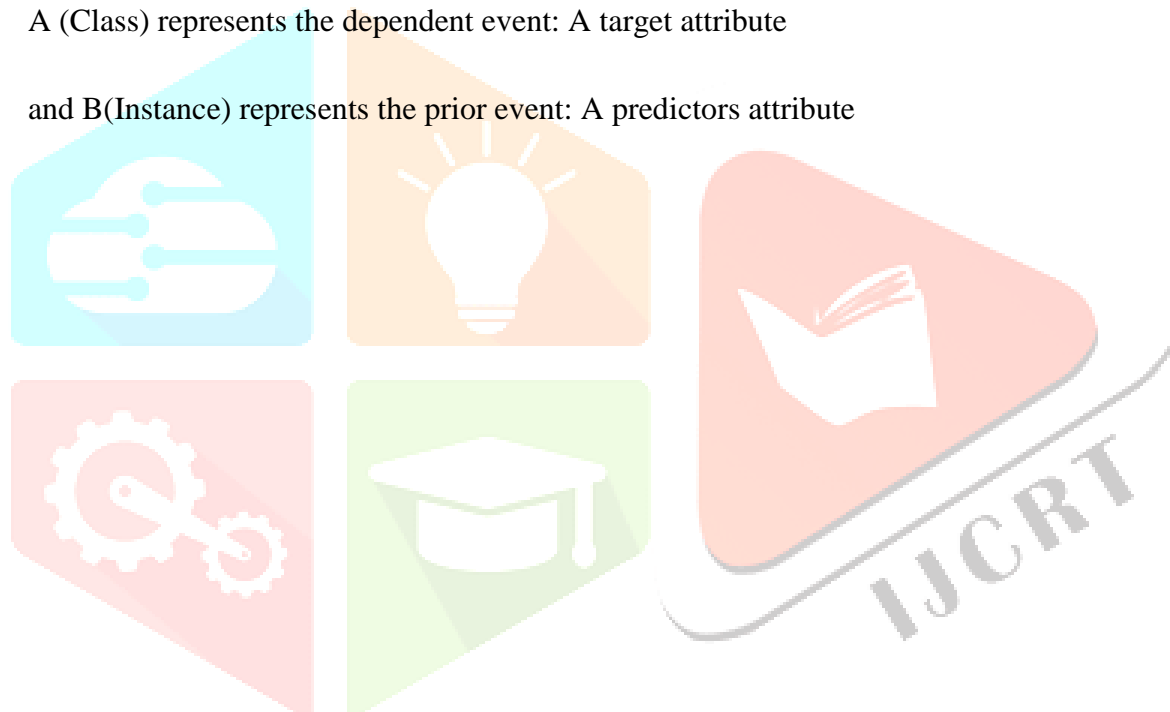
In the above sentence, it is clear that the user is convinced with the quality of the picture of the camera. Here the user talks about picture which is considered as a feature. In certain product review sentences, features are explicitly stated whereas in some reviews, it is very hard to find the feature as they are implicit. In our research, we have focused only on the explicitly stated features and are identified by noun or noun phrases. The identification of implicit features are left for future work. Usually nouns or noun phrases in review sentences are considered as product features. For this research, we extracted only noun phrases from a document based on the consideration that the feature terms are nouns. The NLP Stanford parser has been used to parse each review and to split text into sentences and for each word POS tag is produced.

Each sentence considered for review processing along with the POS tag information of each word in the sentence is stored in review database. The next step is the generation of frequent features. A transaction file is created for this purpose. The transaction file contains only preprocessed nouns/noun phrases of the sentence. All the frequent pattern item sets are identified by association mining. An item set is nothing but a set of words that occurs together in a paragraph. The need for using association mining in this process is to identify the frequent item sets that are likely to be product features. The infrequent noun or noun phrases are expected to be non-product features. Apriori algorithm is applied on the transaction set of nouns/noun phrases. Each resulting frequent item set is a possible feature, if minimum support of the product review sentences are more than 1%. All the candidate frequent features generated association mining are not genuine features. Compactness pruning and redundancy pruning are used to discard the unlikely features. For redundancy pruning a support lower than the minimum p-support of three is used.

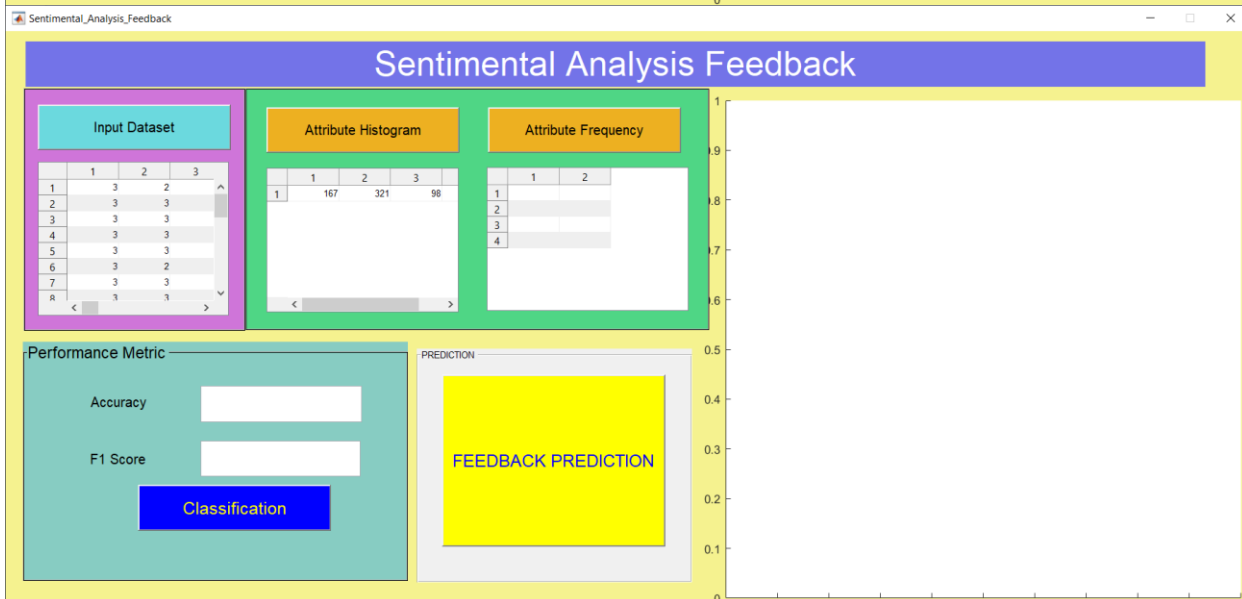
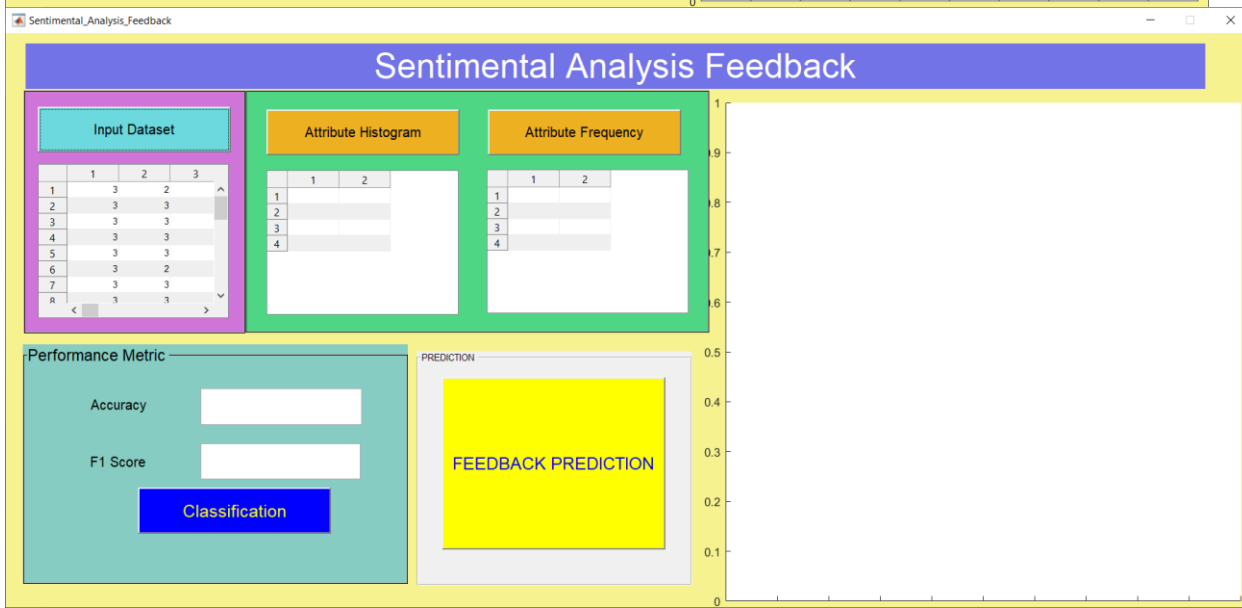
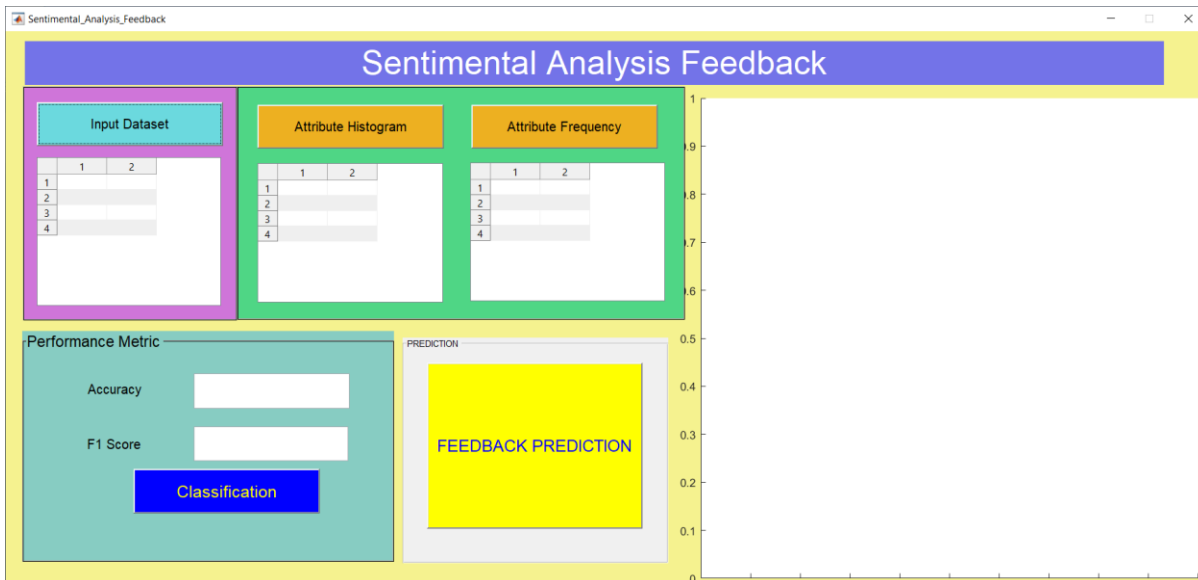
CLASSIFICATION

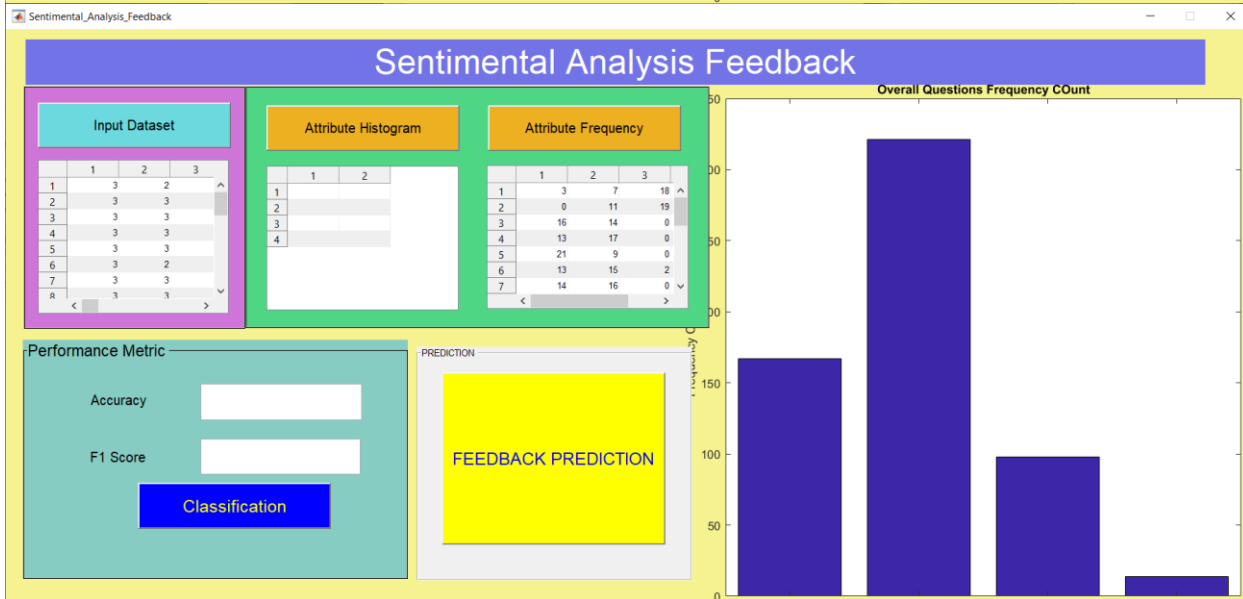
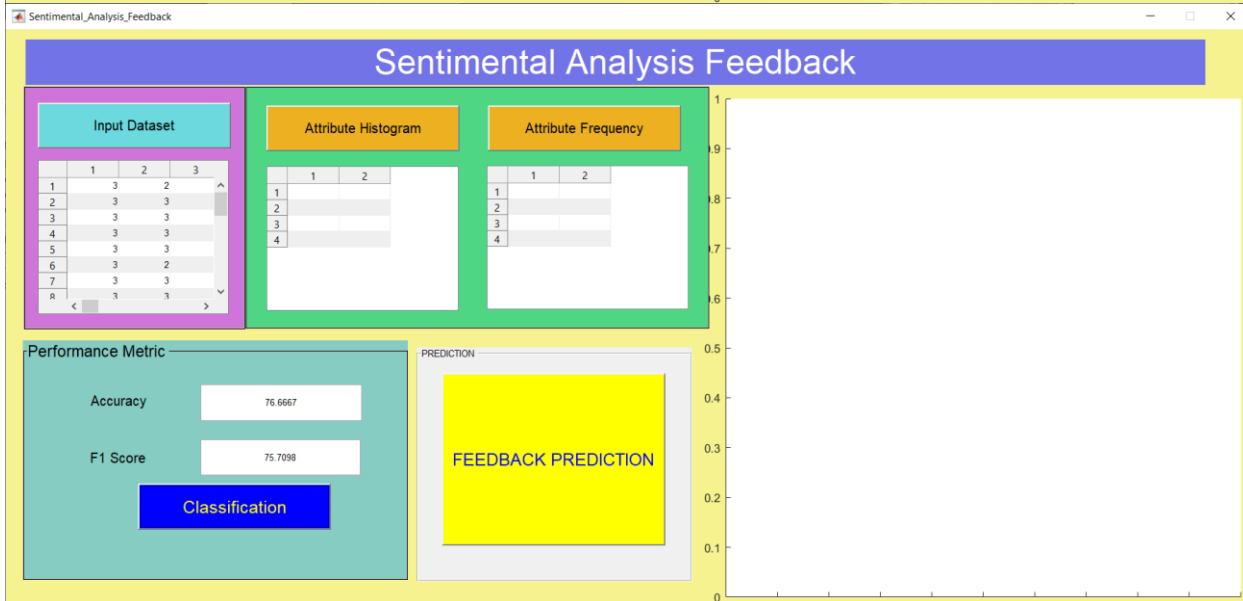
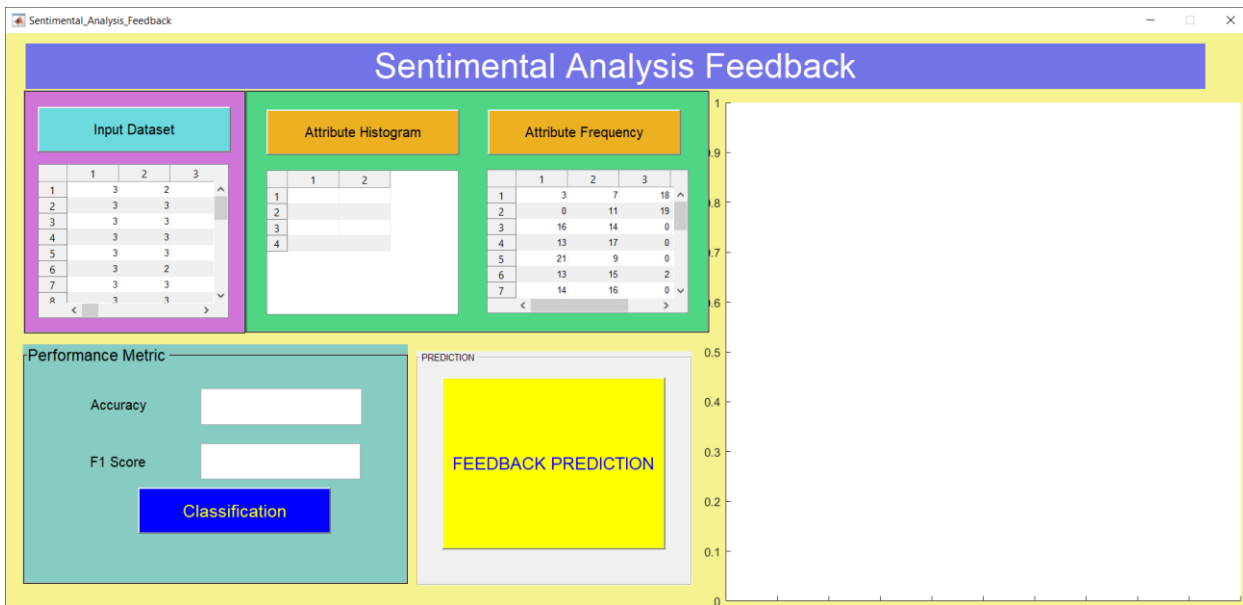
NAIVE BAYES

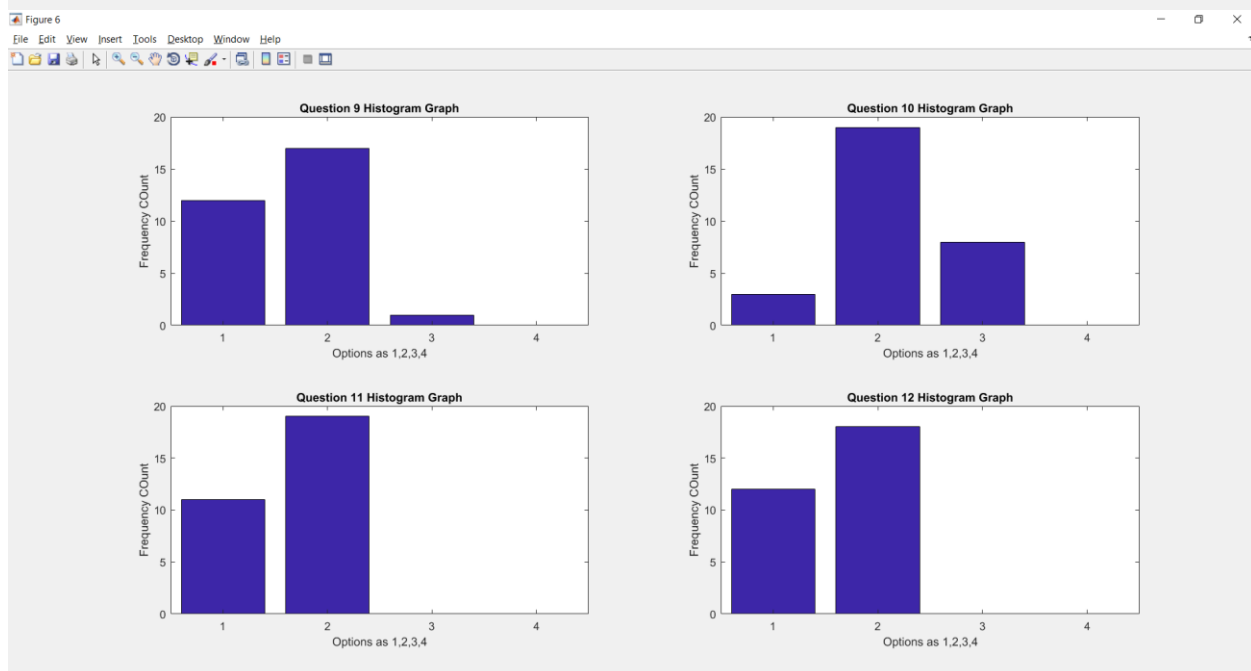
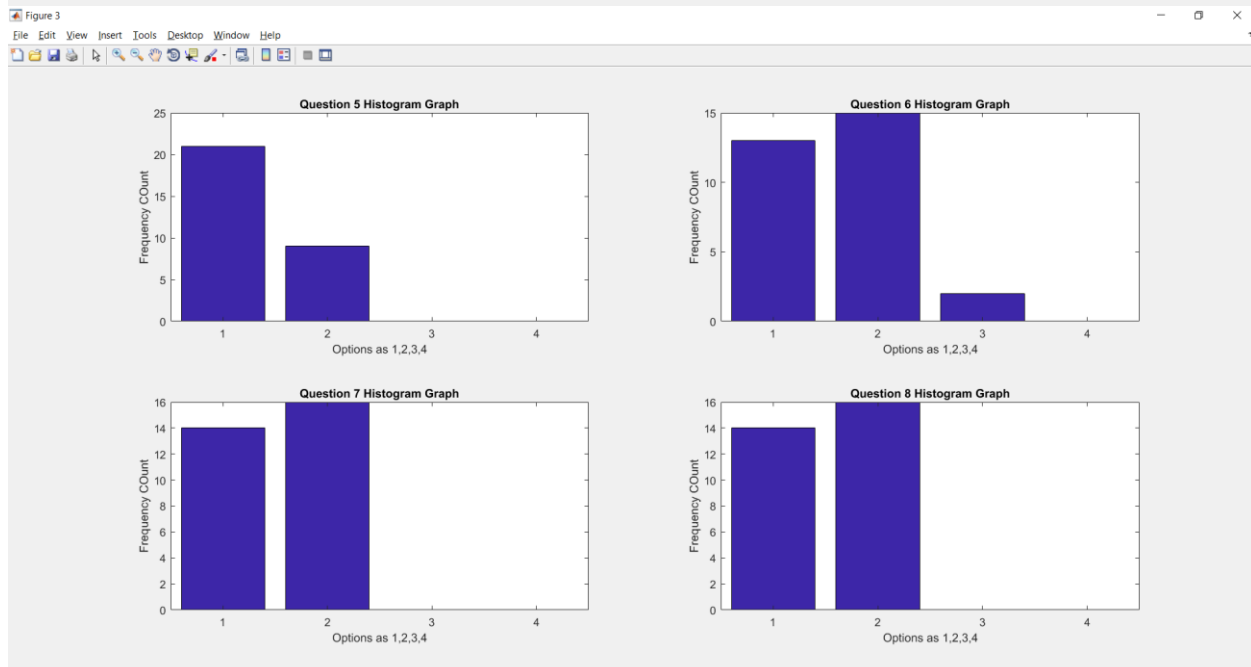
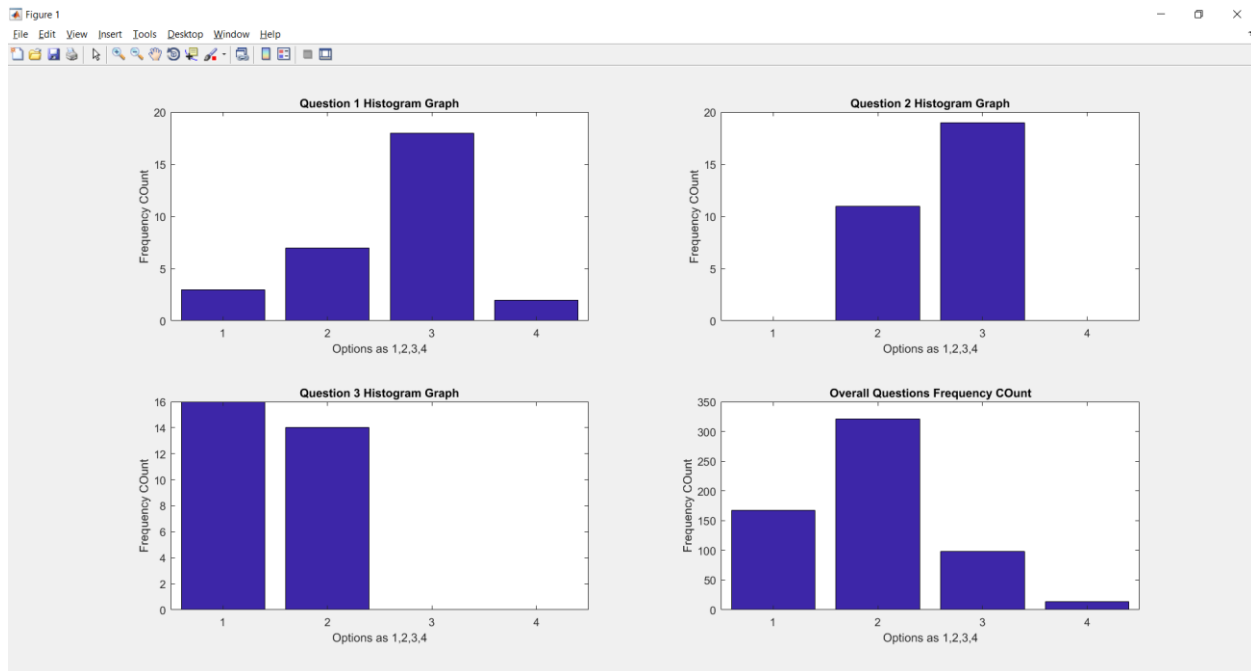
- The Naive Bayes algorithm is based on conditional probabilities. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data.
- Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem: Probability of event A given evidence B
- $\text{Prob}(A \text{ given } B) = \frac{\text{Prob}(A \text{ and } B)}{\text{Prob}(B)}$
- where:
- A (Class) represents the dependent event: A target attribute
- and B(Instance) represents the prior event: A predictors attribute

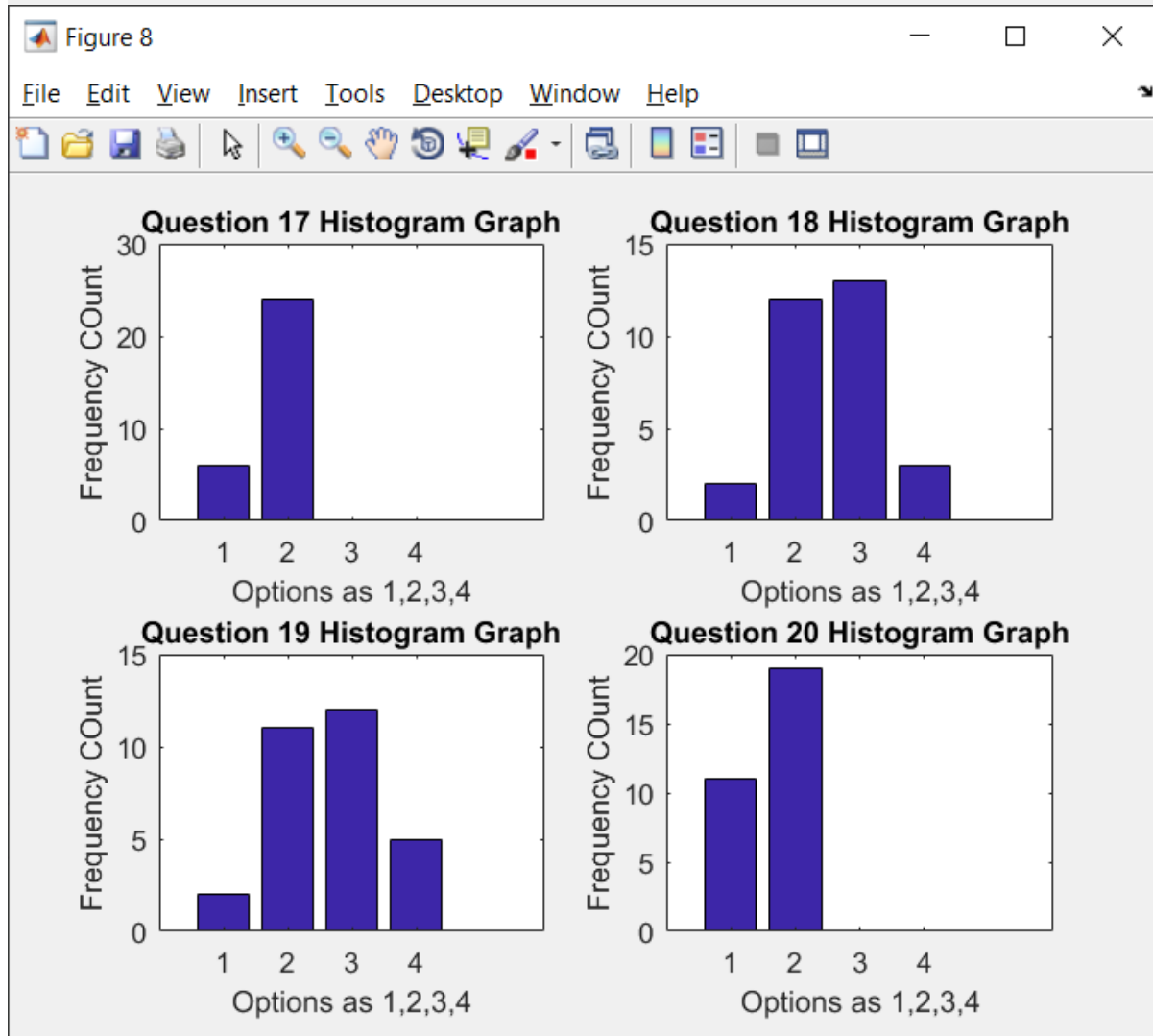
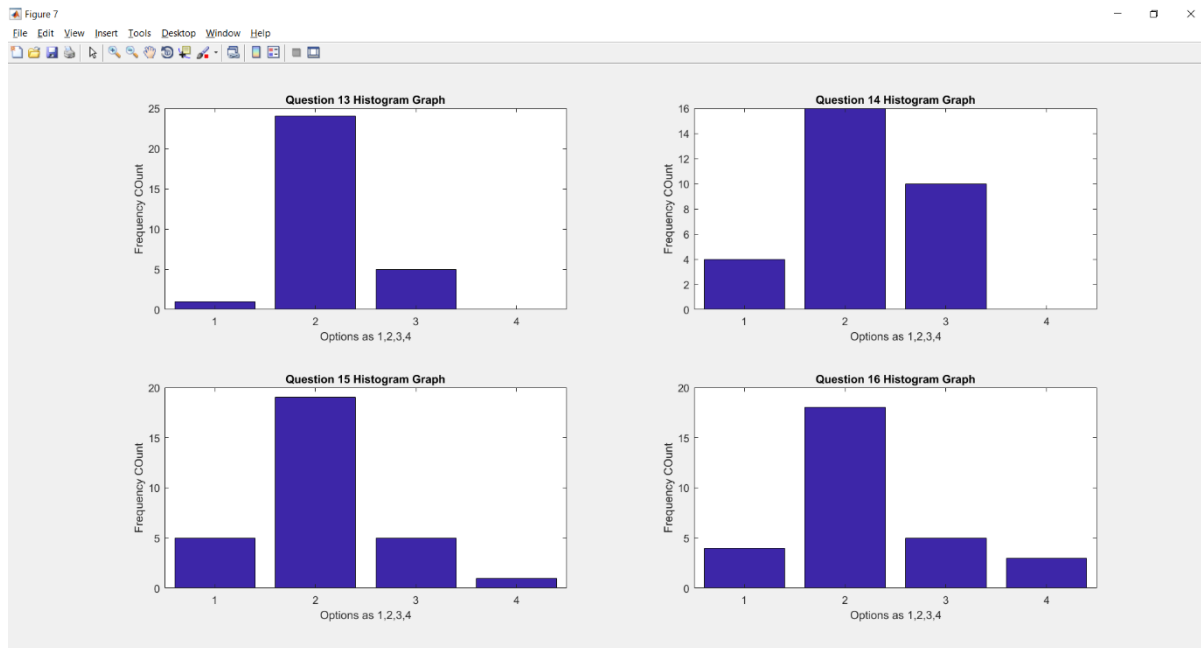


4. RESULTS









5. CONCLUSION

In the present paper, we have reported the analysis of feedback data of teaching from students, which is get from kaggle. We have done the sentiment analysis by using excel as tool. The investigation looked into variety of feedbacks and polarity about teaching from these analysis we get to know about which feedback student gives in more number about teaching according to that we can make some improvements in teaching, or most feedback are positive then we can maintain teaching level as earlier but if more feedback are negative then corrective actions and decisions should be taken. The study represents common students look over towards teaching and it is based on each review.

6. REFERENCE

- [1] [AnshulMittal,ArpitGoelinStock prediction using twitter Sentiment analysis.
- [2] Brennan, J. & Williams, R. (2004) Collecting and Using Student Feedback. A Guide to Good Practice (LTSN, York).
- [3] Dhanalakshmi V in Opinion mining from student feedback data using supervised learning algorithms.
- [4] Elaine Keane & Iain Mac Labhrainn, Obtaining Student Feedback on Teaching & Course Quality , CELT, April 2005.
- [5] GokarnIlaNitin,Asst.Prof.GottipatiSwapna,Prof.VenkyShankararaman in Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach.
- [6] GokarnIlaNitin ,Asst.Prof.GottipatiSwapna , Prof.VenkyShankararaman in Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach.
- [7] HarshaliP.Patil,MohammadAtiquein Sentiment analysis for Social media.
- [8] Tan Li Im, PhangWai San, Chin Kim On Center of Excellence in Semantic Agents Universities Malaysia in Rule-based Sentiment Analysis for Financial News
- [9] Mark McGuire,ConstanceKampf Aarhus University in Using Social Media Sentiment Analysis to Understand Audiences: A New Skill for Technical Communicators.
- [10] M.S.Neethu,R.Rajasree in Sentiment analysis in twitter using machine learning techniques.