# A
# Summer Internship Report
# On
# "Sentiment Analysis of Expert Lecture feedback"

(CE346 – Summer Internship - I)

**Prepared by**

Shubham Dhameliya(17CE026)
Jaydip Pansuriya(17CE064)

**Under the Supervision of**

Prof. Mayuri Popat

**Submitted to**

Charotar University of Science & Technology (CHARUSAT)
for the Partial Fulfillment of the Requirements for the
Degree of Bachelor of Technology (B.Tech.)
for Semester 5

**Submitted at**

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY
**Accredited with Grade A by NAAC**
**Accredited with Grade A by KCG**

cspit
Chandubhai S Patel Institute of Technology

**U & P U. PATEL DEPARTMENT OF COMPUTER ENGINEERING**
**(NBA Accredited)**
**Chandubhai S. Patel Institute of Technology (CSPIT)**
**Faculty of Technology & Engineering (FTE), CHARUSAT**
**At: Changa, Dist: Anand, Pin: 388421.**
**2020**

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Accredited with Grade A by NAAC
Accredited with Grade A by KCG

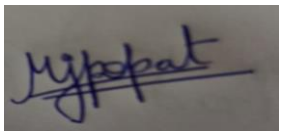
cspit
Chandubhai S Patel Institute of Technology

# CERTIFICATE

This is to certify that the report entitled "**Sentiment analysis of Student feedback**" is a bonafied work carried out by **Shubham Dhameliya(17CE026), Jaydip Pansuriya (17CE064)** under the guidance and supervision of **Prof. Mayuri Popat** for the subject **Summer Internship – I (CE346)** of 5th Semester of Bachelor of Technology in **Computer Engineering** at Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology & Engineering (FTE) – CHARUSAT, Gujarat.

We wish them good luck for their bright future.

Under the supervision of,



Prof. Mayuri Popat
Assistant Professor
U & P U. Patel Dept. of Computer Engineering
CSPIT, FTE, CHARUSAT, Changa, Gujarat

**Chandubhai S. Patel Institute of Technology (CSPIT)**
**Faculty of Technology & Engineering (FTE), CHARUSAT**
At: Changa, Ta. Petlad, Dist. Anand, Pin: 388421. Gujarat.

# ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We respect and thank Dr. Ritesh Patel, for providing us an opportunity to do the intern project work in Charusat University and giving us all support and guidance which made us complete the project duly. We are extremely thankful to him for providing such a nice support and guidance, although he had busy schedule.

We owe our deep gratitude to our project guide Prof. Mayuri Popat, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

I am thankful to and fortunate enough to get constant encouragement, support and guidance from all our friends and family members which helped us in successfully completing our project work.

# ABSTRACT

Sentiment analysis is a machine learning technique that detects polarity (e.g. a *positive* or *negative* opinion) within text, whether a whole document, paragraph, sentence, or clause. The study has been carried out to present the sentiment analysis model for improving the quality of teaching
in academic institutions, particularly at universities. The purpose of this study is to explore the different machine learning techniques to identify its importance as well as to raise interest in this research area. In this regard, the student feedback dataset will be collected from the students and faculties of Charusat university through the project itself. The dataset contains valuable information about the quality of teaching and learning. Methods/Statistical Analysis: The model used the Multinomial Naive Bayes, Stochastic Gradient Decent, Support Vector Machine, Random Forest and Multilayer Perceptron Classifier. The result was analyzed through the evaluation metrics i.e. Confusion Matrix, Precision, Recall and F-score. Findings: It is found that the performance of MNB and MLP remained effective as compared to other approaches. It is recommended that MNB and MLP should be used in the research context for the classification of the text. It has great significance for future researchers in sentences and text classification. The study helps for improving the quality of teaching in education system. And moreover, it will be upgrade by increasing the data samples of neutral comments in dataset.

# TABLE OF CONTENT

# LIST OF FIGURES

# CHAPTER 1:WEB APPLICATION

## 1.1 LOGIN/LOGOUT

Both students and faculties will have to first register themselves in order to login. With the help of credentials they can login into the application. The students and faculties can give the feedback, while subject coordinators can look at the analysis of the feedback.


Fig 1.1 Login page

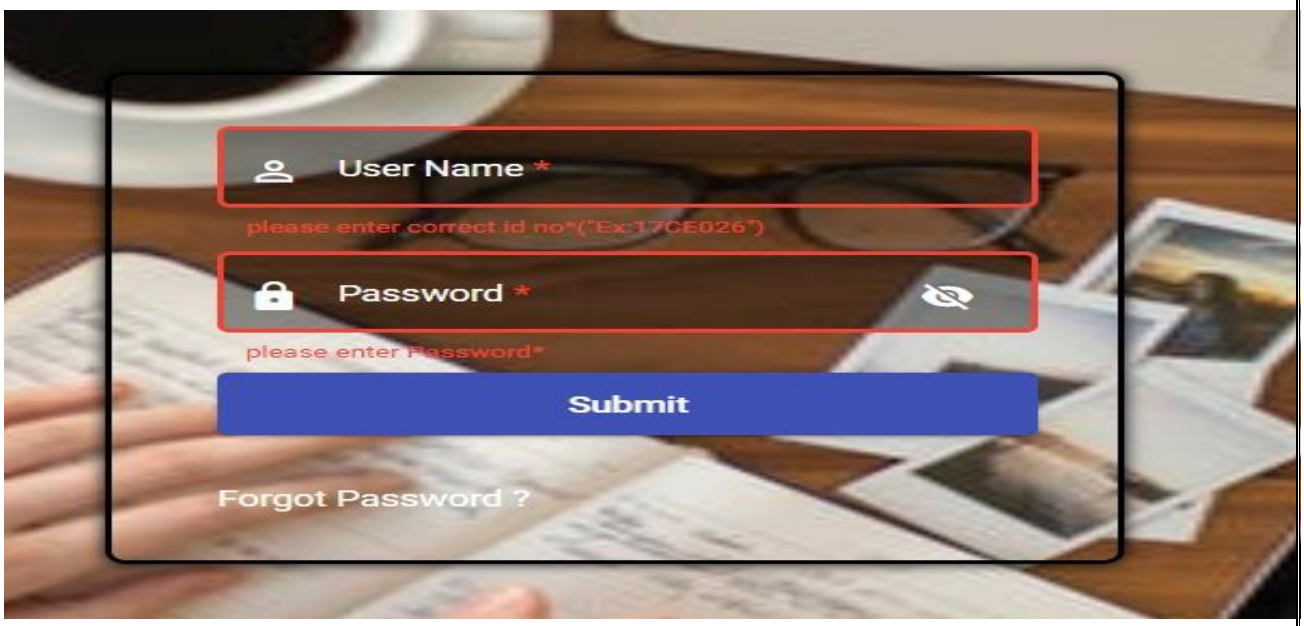



Fig 1.2 Login page

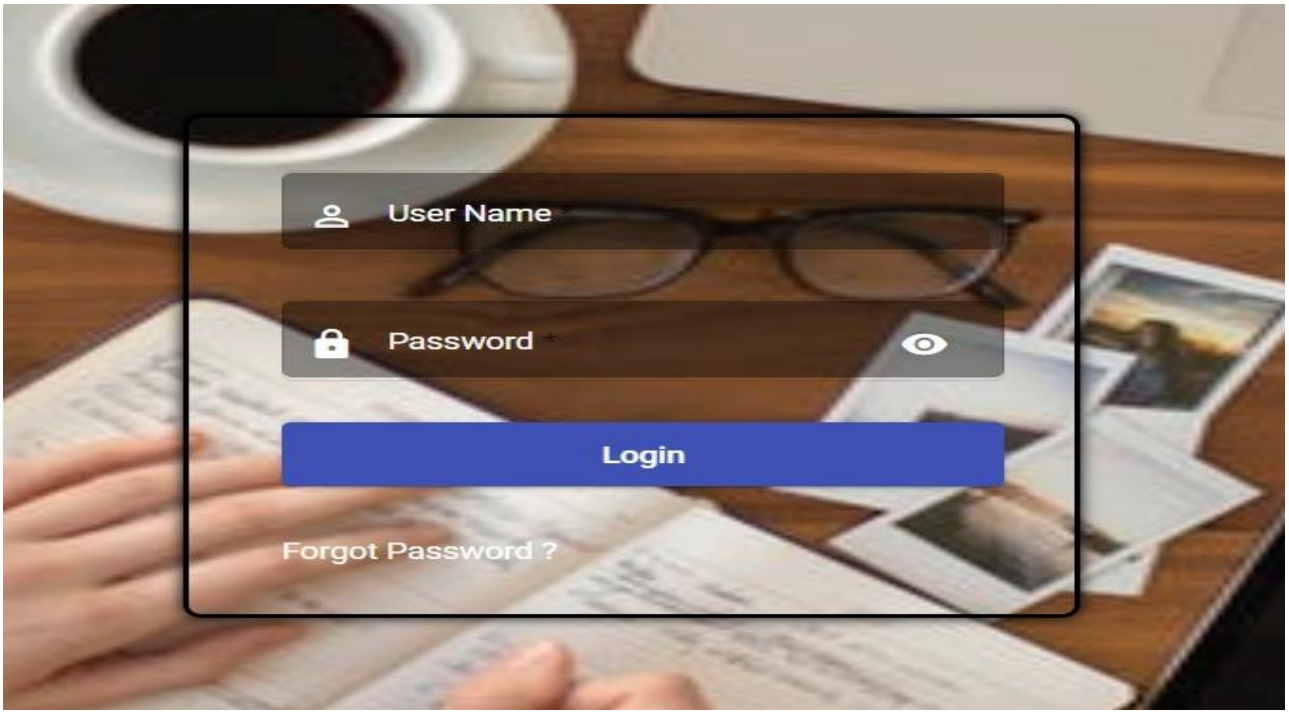



Fig 1.3 Login page

## 1.2 FORMS

There is a separate feedback form for the faculties and students. While logging in, the information required to direct you to your respective feedback form will be fetched from your credentials. Your feedback will be stored in the database.

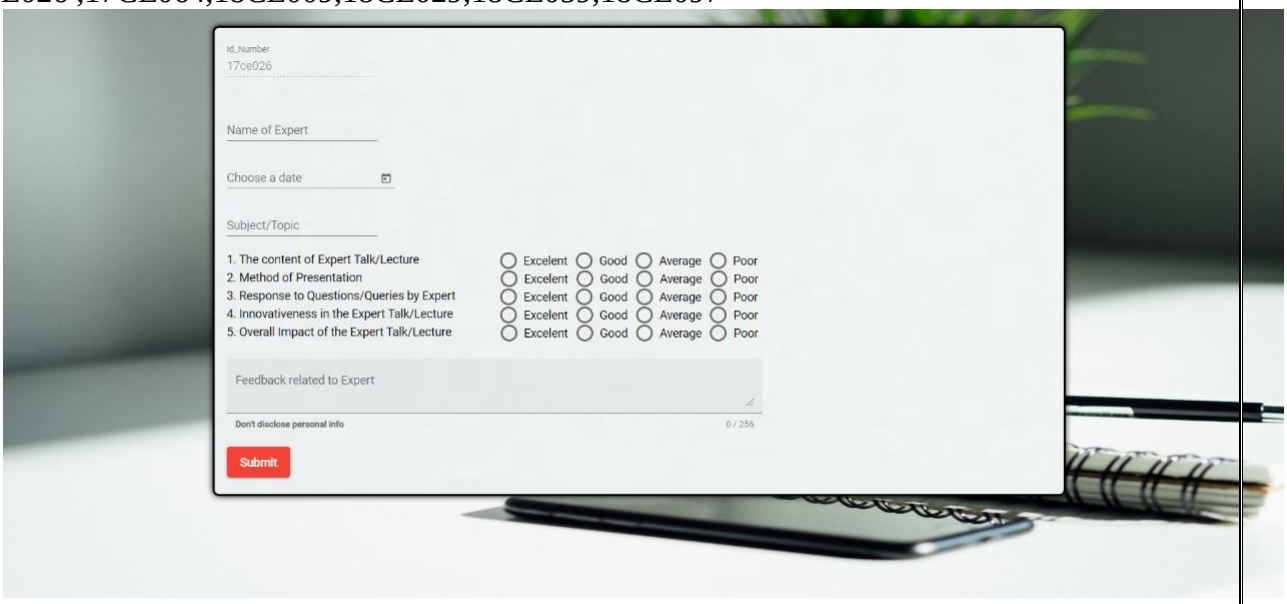



Fig 1.2 Student feedback form

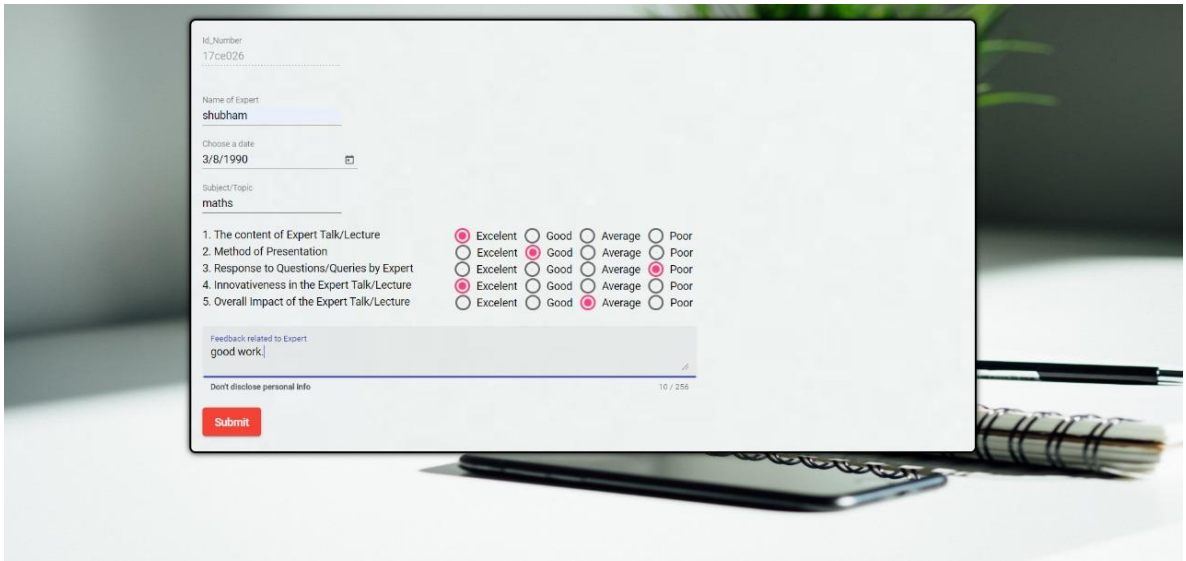



Fig 1.3 Student feedback form

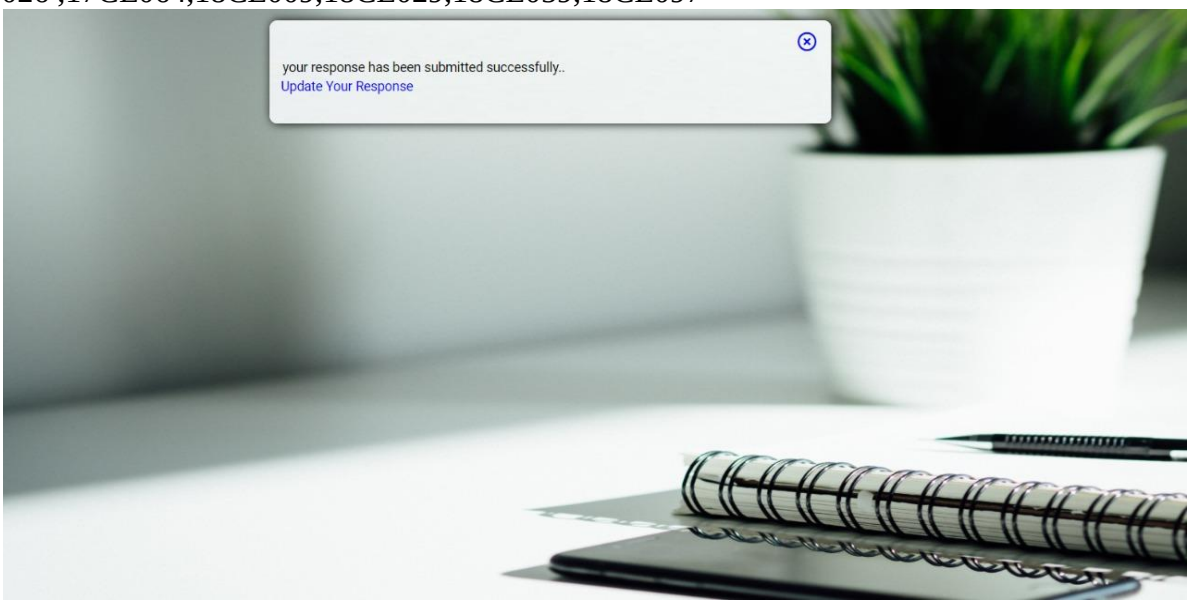



Fig 1.4 Update Feedback

## 1.3 DATABASE

We have used firebase to store all the data of the application.

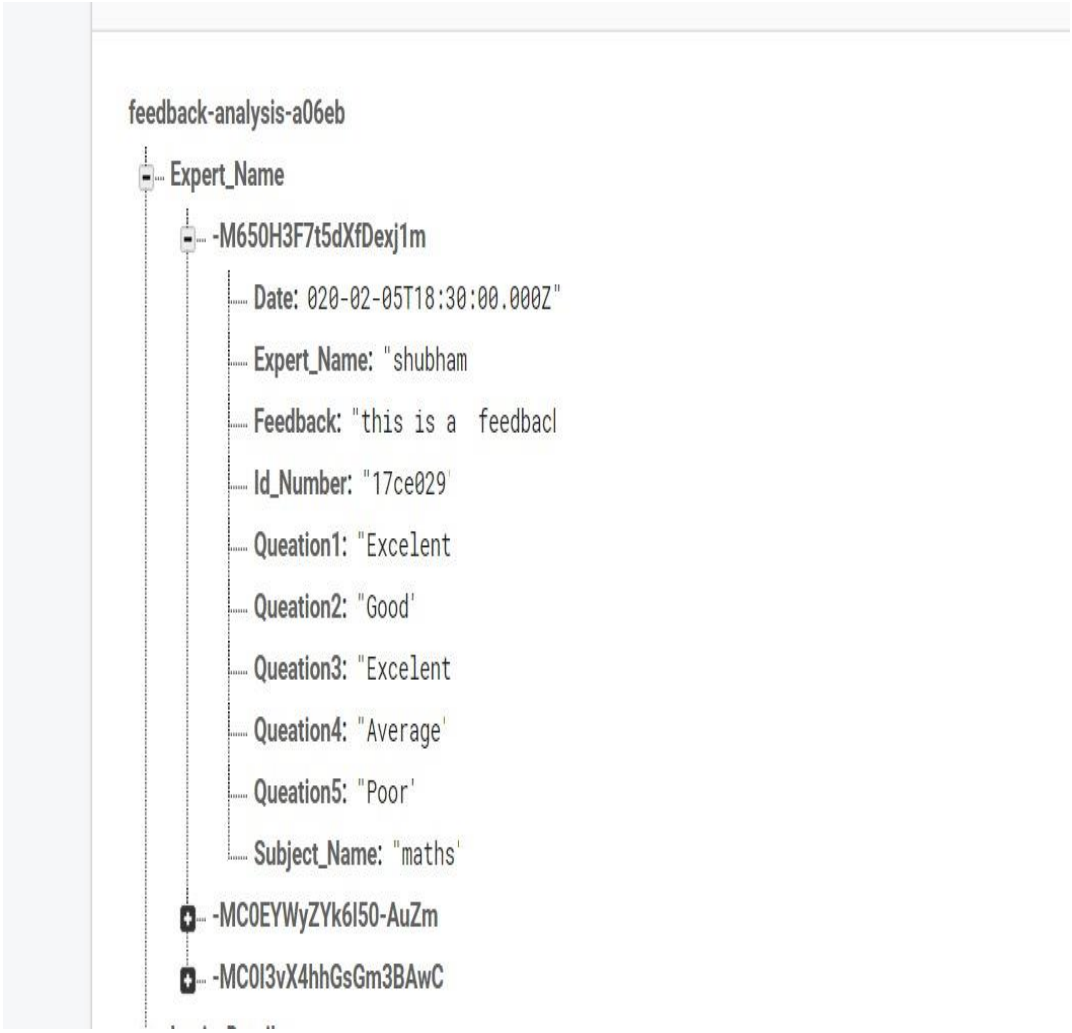



Fig 1.5 Firebase data

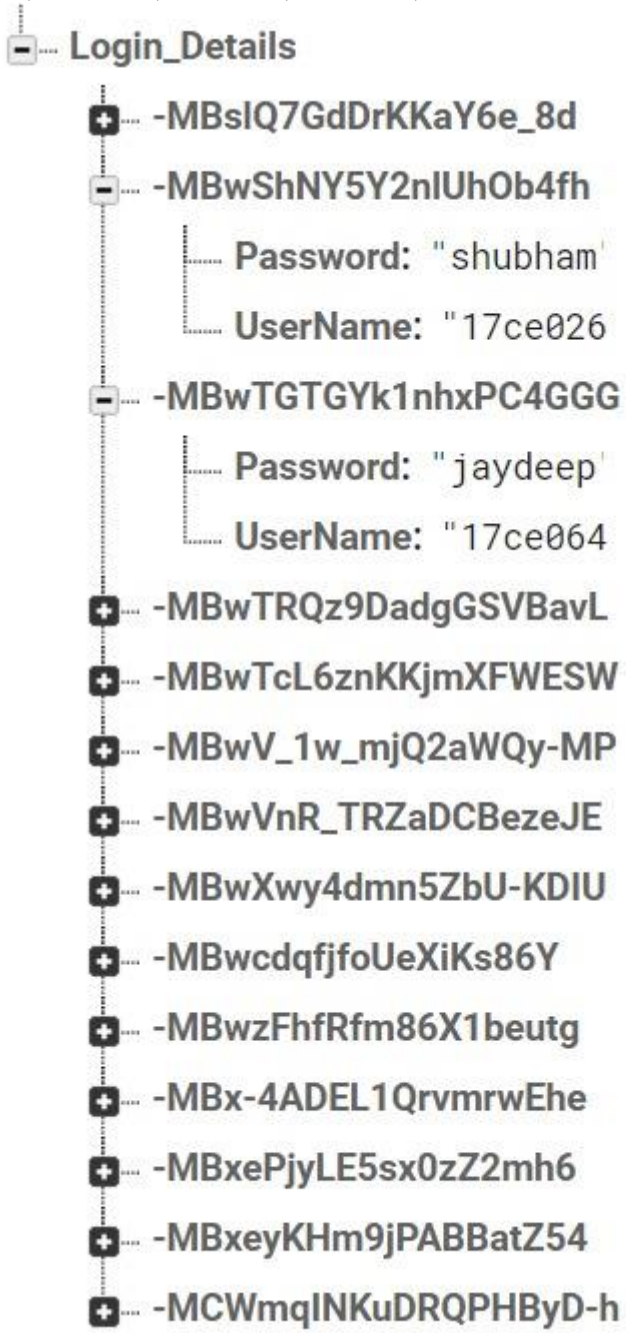



Fig 1.6 Login Data

# CHAPTER 2: TEXT PROCESSING TECHNIQUES

## 2.1 REMOVING NUMBERS AND PUNCTUATIONS

In this step we remove numbers, punctuations marks and other unnecessary things which is not used full for sentiment analysis.

## 2.2 REMOVING STOP WORDS

In English language there are many grammar words like this, that, is and so on which is not represent any sentiment so we need to remove it so that we get accurate sentiment.

## 2.3 TOKENISER

Tokenization is the act of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded. The tokens become the input for another process like parsing and text mining

After performing above step we have data that we can use for Analysis.

# CHAPTER 3: SENTIMENT ANALYSIS BASED ON VARIOUS ALGORITHMS

## 3.1 KNN ALGORITHM:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
it is used for the Classification problems.
K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
**Example:** Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

**We get Accuracy : 74.2144%**

## 3.2 SVM MODEL:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

However, primarily, it is used for Classification problems in Machine Learning.
The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

**We get Accuracy: 78.38114%**

## 3.3 NAÏVE BAYES ALGORITHM :

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
It is mainly used in *text classification* that includes a high-dimensional training dataset.
Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
**It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

**Accuracy : 52.0833%**

## 3.4 DECISION TREE:

Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**
In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
The decisions or the test are performed on the basis of features of the givendataset.
***It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.***
It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
In order to build a tree, we use the **CART algorithm,** which stands for **Classification and Regression Tree algorithm.**
A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

**We get Accuracy: 67.110%**

## 3.5 RANDOM FOREST:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, ***"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."*** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**We get Accuracy: 74.043%**

After analyzing various models, we found that SVM algorithms were more accurate than the rest and have been used for the project.

# CHAPTER 4: RESULT

With the help of python library matplotlib we create all charts base on data given by students . The students were asked 5 questions and based on the feedback the following graphs have been derived.


Fig 4.1 Overall feedback

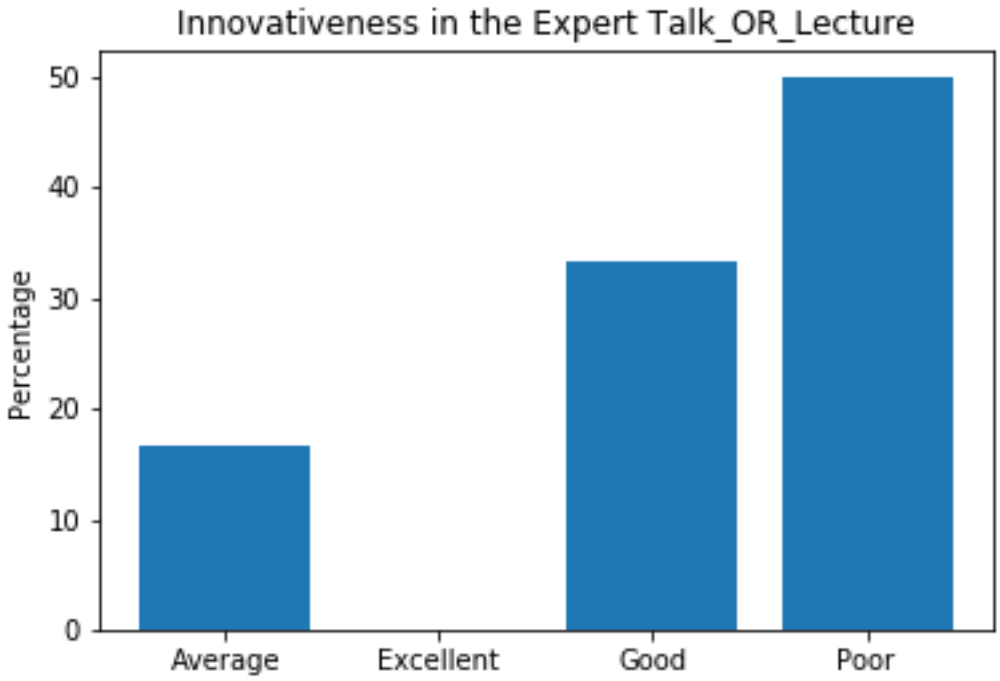



Fig 4.2 feedback for Q1

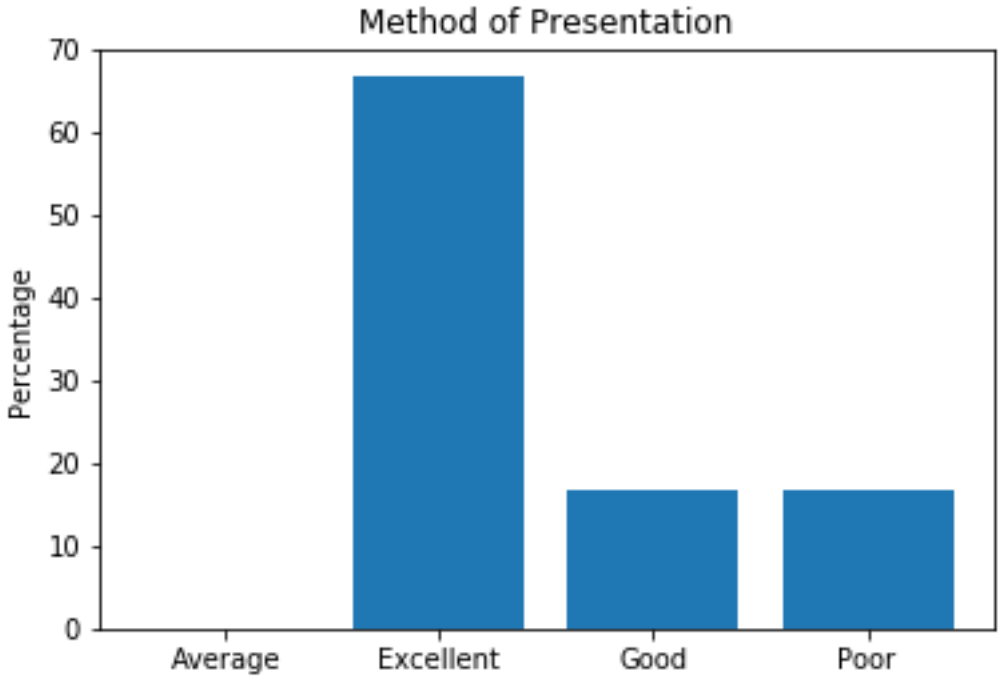



Fig 4.3 Feedback for Q2

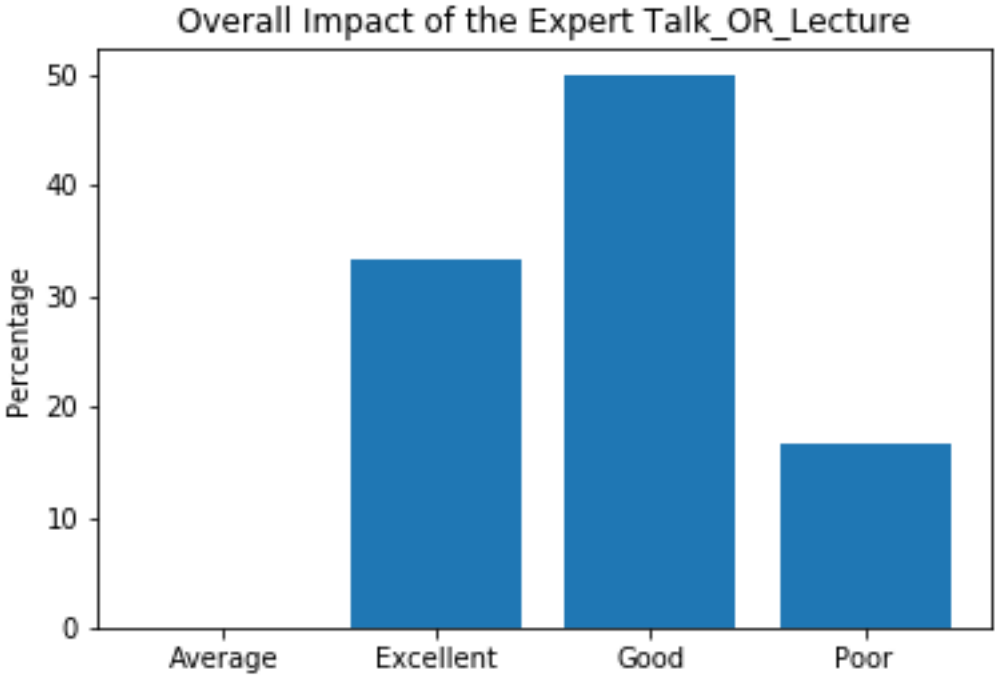



Fig 4.4 Feedback for Q3

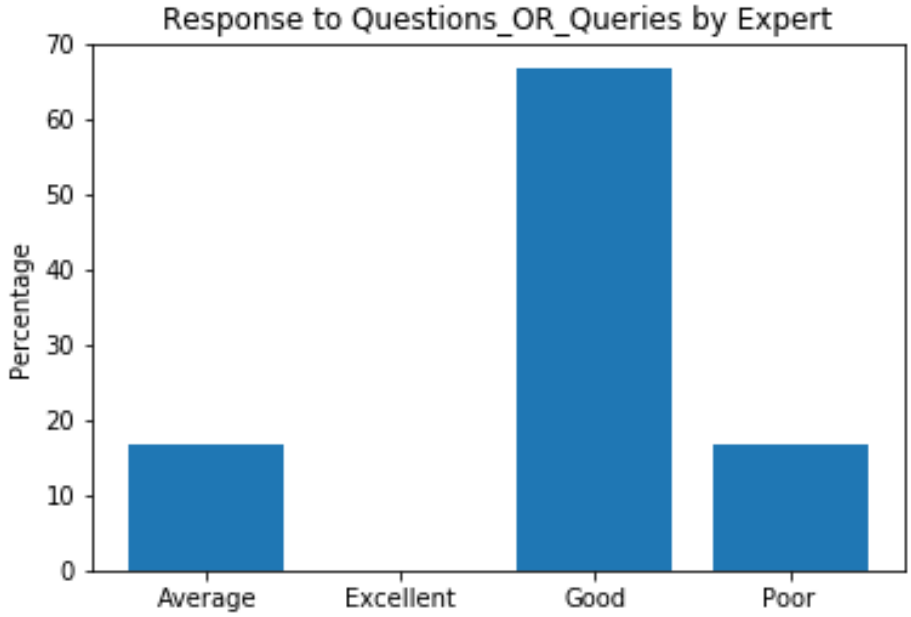



Fig 4.5 Feedback for Q4

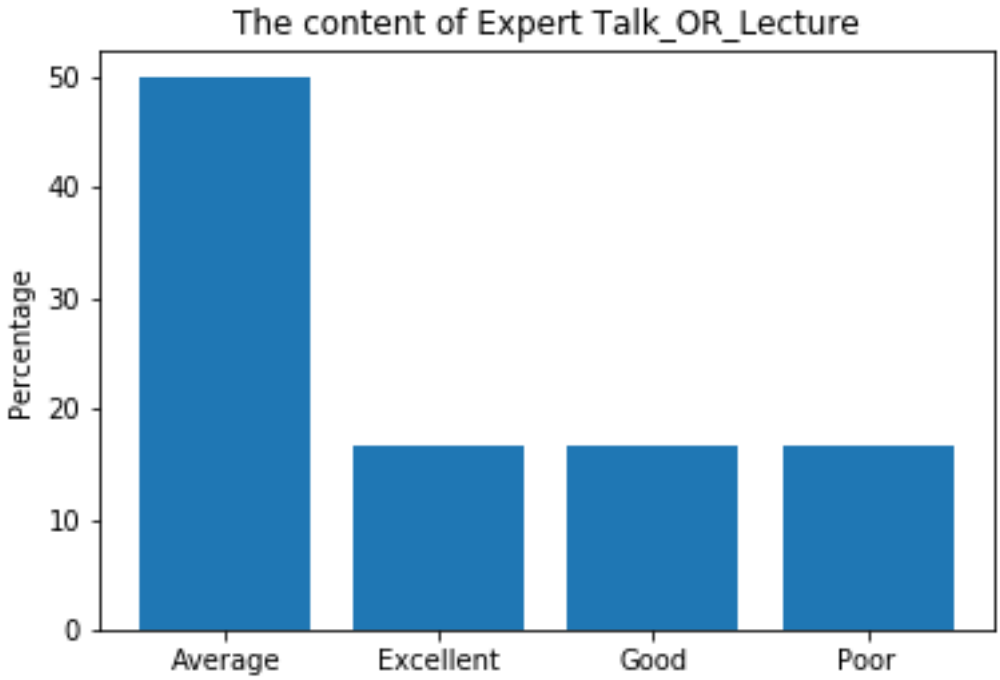



Fig 4.6 Feedback for Q5

# CONCLUSION

This study has been designed to analyses the discourse commonly used in an academic context. The study used the sentiment analysis survey by incorporating various machine learning techniques. It focuses particularly on students' feedback to strengthen the teaching and learning process at university. Student feedback dataset is collected from various feedback forms submitted b5y them to the university. Various machine learning techniques are used to analyze the text data. The student feedback dataset split into two parts training and testing; 30% for testing and 70% for training. Moreover, accuracy is evaluated through confusion matrix, Precision,Recall and F-score metrics. The various algorithms used to train the model include random forest(accuracy = 74.043%), decision tree (accuracy = 67.110%), Naïve Bayes model (accuracy = 52.0833%), SVM model (accuracy = 78.38114%) and KNN (accuracy = 74.2144%). From the accuracies of the given algorithms, it is concluded that SVM is more reliable than the others.

# REFERENCES

https://towardsdatascience.com/updated-text-preprocessing-techniques-for-sentiment-asnalysis-549af7fe412a

https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

https://www.javatpoint.com/machine-learning-naive-bayes-classifier

https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

https://www.geeksforgeeks.org/decision-tree-introduction-example/

https://www.javatpoint.com/machine-learning-random-forest-algorithm

https://towardsdatascience.com/random-forest-classification-and-its-implementation-d5d840dbead0

https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/

https://www.coursera.org/learn/javascript-jquery-json/home/welcome