**SENTIMENTAL ANALYSIS**

**MAJOR PROJECT**

***Submitted by***

**ABDULLAH**

**2019-310-004**

***In partial fulfillment for the award of the degree of***

## B. Tech (Computer Science & Engineering)

***Under the supervision of***

## Dr. SHAHAB SAQUIB SOHAIL



**Department of Computer Science & Engineering School of Engineering Sciences & Technology**

**JAMIA HAMDARD**

**(Deemed to be University)**

**New Delhi-110062**

**(2023)**

**DECLARATION**

I, **Abdullah** student of **BTech CSE (Enrollment No.: 2019-310-004)** hereby declare that the Project entitled **“Sentimental Analysis”** which is being submitted by me to the Department of Computer Science, Jamia Hamdard, New Delhi in partial fulfillment of the requirement for the award of the degree of **BT-CSE,** is my original work and has not been submitted anywhere else for the award of any Degree, Diploma, Associate ship, Fellowship or other similar title or recognition.

**Name: Abdullah Signature: Date: 16/05/2023**

**Place: New Delhi**

**ACKNOWLEDGMENT**

I would like to express my gratitude and appreciation to all those who gave me the possibility to complete this report. Special thanks are due to my supervisor **Dr. Shahab Saquib Sohail** whose help, stimulating suggestions and encouragement helped me in all time of fabrication process and in writing this report. I also sincerely thanks for the time spent proofreading and correcting my many mistakes.

I would also like to acknowledge with much appreciation the crucial role of the staff in the Computer Laboratory, who gave me the permission to use the system for programming and coding for the project work.

Many thanks go to all lecturers and supervisors who have given their full effort in guiding me in achieving the goal as well as their encouragement to maintain my progress in track. My profound thanks go to all my classmates, especially to my friends for spending their time in helping and giving support whenever I need it in fabricating my project.

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TOPICS** | **Page No.** |
|  | **List of figures** | **5** |
|  | **Objectives** | **6-7** |
|  | **System requirements specification** | **8** |
|  | **Problem statement** | **9** |
|  | **Methodology** | **10** |
|  | **Libraries used** | **11** |
|  | **Machine learning model performance** | **12-13** |
|  | **Entity-relation diagram** | **14** |
|  | **Use case diagram** | **15** |
|  | **Screenshots** | **16-18** |
|  | **Conclusion** | **19** |
|  | **Limitations** | **20** |
|  | **Bibliography** | **21** |

**TABLE OF CONTENTS**

# LIST OF FIGURES

* 1. **Sentimental Analysis**
  2. **Sample Output of the program**

**4.1 Confusion Matrix**

**5.1 Snapshots of all libraries used**

**6.1 Accuracy of model**

**6.3 Accuracy graph**

**7.1 Entity-Relational diagram**

**8.1 User case diagram**

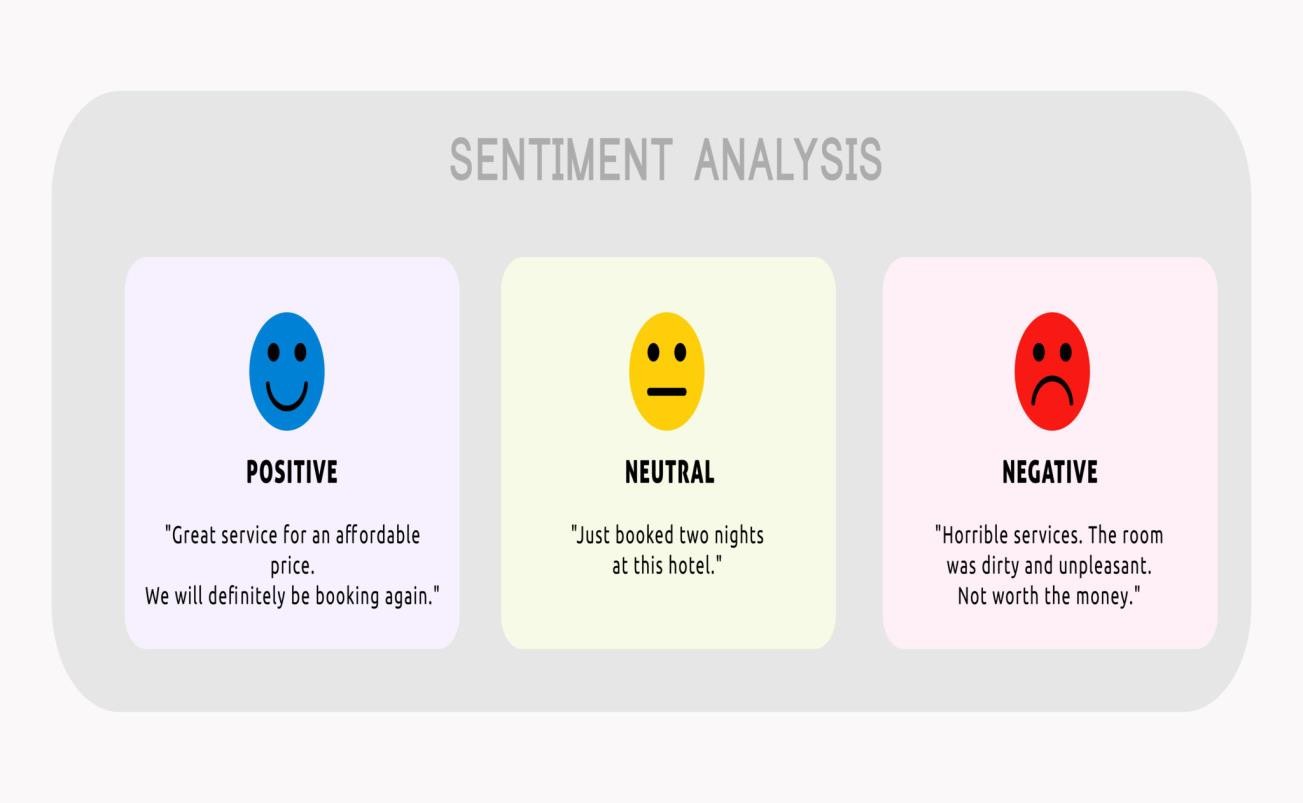
* 1. **Output-1**
  2. **Output-2**
  3. **Output-3**
  4. **Output-4**
  5. **Output with an error**
  6. **First limitation**

# Sentimental Analysis using Random Forest classification

**OBJECTIVE:**

The objective of sentiment analysis is to accurately extract people's opinions from an oversized variety of unstructured review texts and classifying them into sentiment categories, i.e., positive, negative, or neutral. typically “positive” and “negative” also are thought-about.

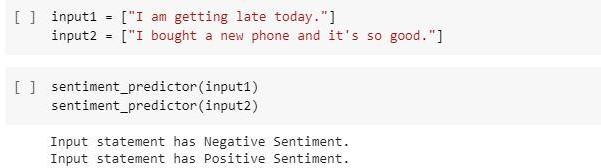
Sentiment analysis is additionally known as opinion mining, that refers to the utilization of language process and text mining to spot the emotional info from text materials.



**Fig 1.1: Sentimental Analysis**

# HERE IS AN EXAMPLE WHAT Sentimental Analysis DOES:

**Predicted Output** -



**Fig 1.2: Output of the Program**

# Specific System Requirements

* + 1. **Google Colab:**

It is a product from Google analysis. Colab permits anybody to put in writing and execute impulsive python code through the browser, and is particularly like minded to machine learning, knowledge analysis and education. a lot of technically, Colab may be a hosted Jupyter notebook service that needs no setup to use, whereas providing access freed from charge to computing resources as well as GPUs.

# Python

Python is a computer programming language often used to **build websites and software, automate tasks, and conduct data analysis**. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

# Random Forest Classifier

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

# PROBLEM STATEMENT

Sentiment analysis is a powerful marketing tool that enables product managers to understand customer emotions in their marketing campaigns. It is an important factor when it comes to product and brand recognition, customer loyalty, customer satisfaction, advertising and promotion's success, and product acceptance.

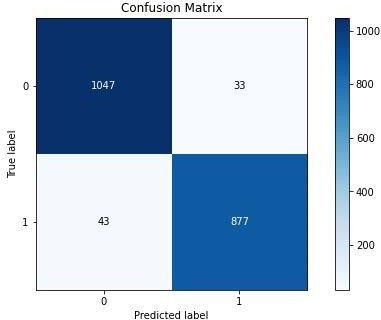
Understanding the psychology of consumers can help product managers and customer success managers to alter their product roadmap with greater precision. The term emotion-based marketing is a broad umbrella phrase that encompasses emotional customer responses, such as "positive," "negative," "neutral," "negative," "uptight," "disgust," "frustration," and others. Understanding the psychology of customer responses can also increase product and brand recall.

# METHODOLOGY:

It uses unsupervised learning like Random Forest Classifier to produce the sentimental Analysis using the Sentimental Analysis dataset.

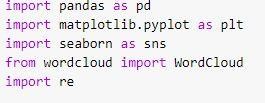
First, we will do the pre-processing of the dataset and then extract the features which are the sentiments from people’s online post after that we will train the model using the extracted sentiments using Random Forest Classifier, we classify them into two sub-classes ‘Positive’ and ‘Negative’ then we test the trained model using random inputs to check its validity and accuracy.

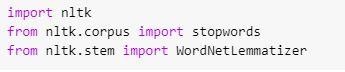
The accuracy generated by this random forest classifier model is 96.2% and the confusion Matrix generated is



**Fig 4.1 Confusion Matrix**

# LIBRARIES USED IN PYTHON





**Fig 5.1 Snapshot of all the Libraries used**

**Pandas:** Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named [Numpy](https://www.activestate.com/products/python/python-packages/), which provides support for multi-dimensional arrays. As one of the most popular data wrangling packages, Pandas works well with many other [data science](https://www.activestate.com/products/python/python-data-science/) modules inside the Python ecosystem, and is typically included in every Python distribution, from those that come with your operating system to commercial vendor distributions like ActiveState’s [ActivePython](https://platform.activestate.com/featured-projects).

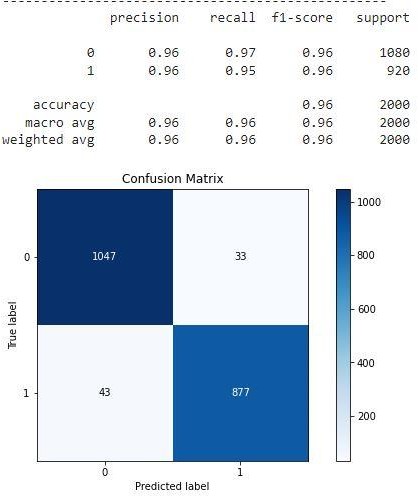
**Matplotlib**: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

**Nltk:** NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

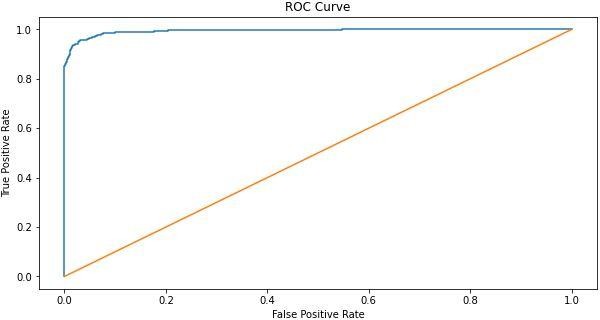
# Machine Learning Model Performance

Accuracy is **one metric for evaluating classification models**. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition: Accuracy = Number of correct predictions Total number of predictions.

The accuracy generated by this random forest classifier model is 96.2% and the confusion Matrix generated is

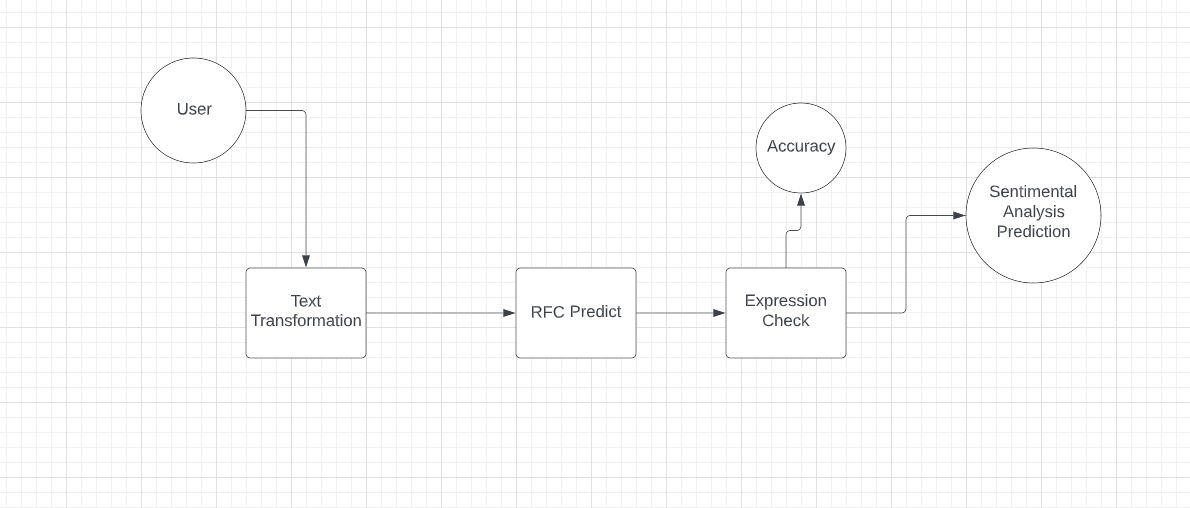


**Fig 6.1 Accuracy of the Model**



**Fig 6.2 Accuracy Graph**

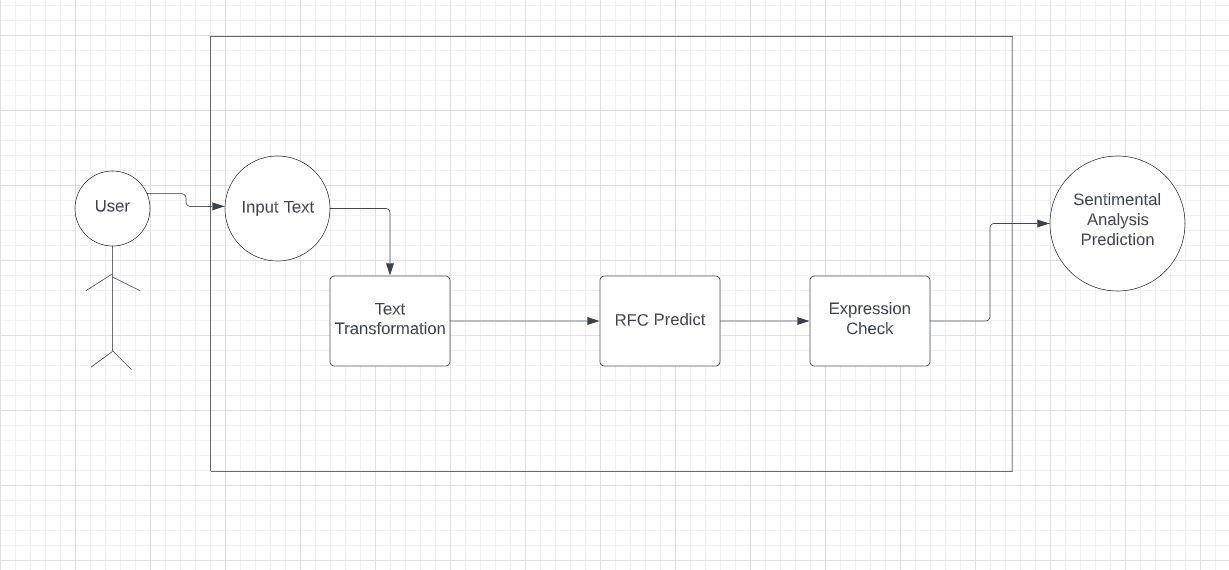
# ENTITY-RELATION DIAGRAM



**Fig 7.1 Entity Relation Diagram**

# USE CASE DIAGRAM

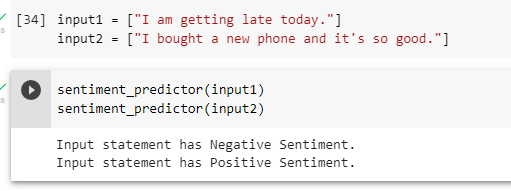
A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses.



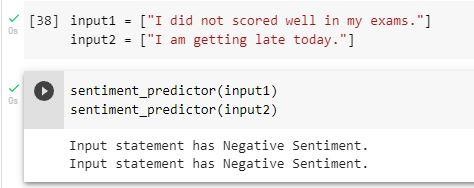
**Fig 8.1: Use Case Diagram**

First, you need to organize your four key elements — **system, actors, use cases and relationships**. Then, arrange them visually in a way that makes sense and will allow you to see immediately the connections between them.

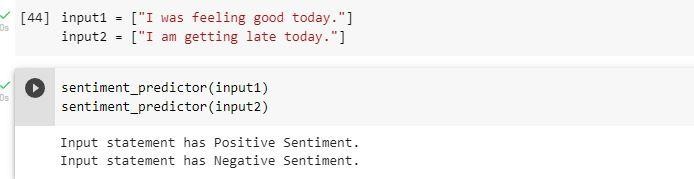
# SCREENSHOTS OF Predicted Answer on Google Colab



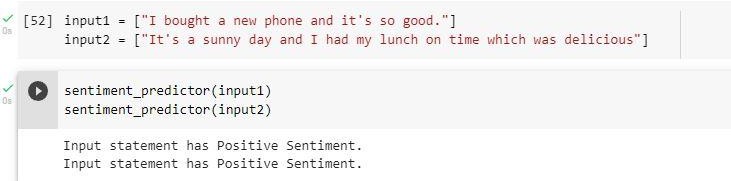
**Fig 10.1: Output-1**



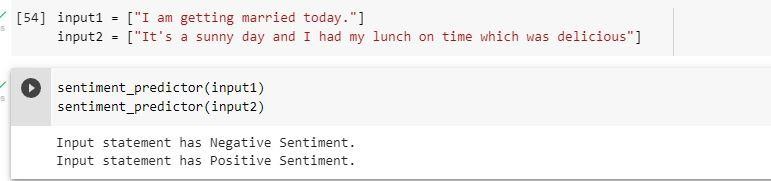
**Fig 10.2: Output-2**



**Fig 10.3: Output-3**



**Fig 10.4: Output-4**



**Fig 10.5: Output with a wrong Prediction**

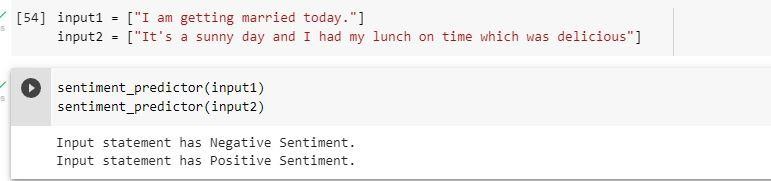
# CONCLUSION

Sentiment analysis deals with the classification of texts based on the sentiments they contain. This article focuses on a typical sentiment analysis model consisting of three core steps, namely data preparation, review analysis and sentiment classification, and describes representative techniques involved in those steps.

Sentiment analysis is an emerging research area in text mining and computational linguistics, and has attracted considerable research attention in the past few years. Future research shall explore sophisticated methods for opinion and product feature extraction, as well as new classification models that can address the ordered labels property in rating inference. Applications that utilize results from sentiment analysis is also expected to emerge in the near future.

# LIMITATIONS

The limitation of this **MODEL** is that it predicts a negative statement sometimes for a positive statement if not Neutral.



**Fig 11.1: First limitation**

# BIBLIOGRAPHY

Some references for help were: [https://towardsdatascience.com/sentiment-](https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17) [analysis-concept-analysis-and-applications-](https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17) [6c94d6f58c17](https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17)

<https://www.commsights.com/benefits-of-sentiment-analysis-for-businesses/>