# Sentimental Analysis

***A project report submitted to***

**MALLA REDDY UNIVERSITY**

***in partial fulfilment of the requirements for the award of degree of***

**BACHELOR OF TECHNOLOGY**

**in**

## COMPUTER SCIENCE AND ENGINEERING (AI & ML)

### Submitted by

**V.Durga Mahesh : 2111CS020127**

### P.Durga Shashinadh : 2111CS020128

**B.Esha Tanmai : 2111CS020129**

### K.Eshwar Sai Pranay : 2111CS020130

### Fardeena Thoufiya : 2111CS020131

### Sd.Fouziya Thabassum : 2111CS020132

***Under the Guidance of***

**Prof.Balaiah**

**Assistant Professor**



**COLLEGE CERTIFICATE**

This is to certify that this is the bonafide record of the application development entitled, “**Sentimental Analysis”** submitted by V. Durga Mahesh(2111CS020127), P. Durga Shashinadh (2111CS020128), B. Esha Tanmai(2111CS020129),E.Eswar Pranay(2111CS020130),FardeenaThoufiya(2111CS020131),Sd.Fouziya Thabassum(2111CS020132). B. Tech III year I semester, Department of CSE (AI &ML) during the year 2022-2023. The results embodied in the report have not been submitted to any other university or institute for the award of any degree or diploma.

### INTERNAL GUIDE HEAD OF THE DEPARTMENT

**Prof Balaiah Dr.Thayyaba Khatoon CSE(AI&ML)**

Acknowledgement

My Major Project would not have been successful without the help of several people .I would like to thank the personalities who were part of my major project in numerous ways, those who gave us outstanding support from the birth of the project.

I am extremely thankful to our honorable Pro-Vice Chancellor, ***VSK Reddy*** sir for providing necessary infrastructure and resources for the accomplishment of my project.

I highly indebted to **Prof. Kailasa Rao** sir Dean, School Of Engineering ,for his support during the tenure of the project.

I am very much obliged to our beloved Prof. **Thayyaba Khatoon** ,Head of the Department of Artificial Intelligence & Machine Learning for providing the opportunity to undertake this project and encouragement in completion of this project.

I hereby wish to express our deep sense of gratitude to Prof. Hariharan for the esteemed guidance ,moral support and invaluable advice provided by them for the success of the project.

I am also thankful to all the staff members of Computer Science and Engineering department who have cooperated in making our project a success. We would like to thank all our parents and friends who extended their help, encouragement and moral support either directly or indirectly in our project work.

ABSTRACT

Sentiment analysis is a technique used to classify the polarity of sentiments and opinions expressed in text data. This classification may be binary, distinguishing between positive and negative sentiments, or multiple, identifying emotions such as happiness, anger, sadness, and disgust. In this project, we aim to develop a sentiment analysis system using Python, a popular language for data science and machine learning. The sentiment analysis system will use machine learning techniques, including natural language processing and classification algorithms, to analyze text data and determine the sentiment scores. The system will be trained using a labeled dataset, which includes text data and their corresponding sentiment labels. The ultimate goal of this project is to develop a robust and accurate sentiment analysis system that can be applied to various applications, such as social media monitoring, customer feedback analysis, and market research. Additionally, this project will provide an opportunity to learn about machine learning techniques, including text preprocessing, feature extraction, and model selection, and apply them to real-world problems. In summary, this project aims to develop a sentiment analysis system using Python, which can classify the polarity of sentiments and opinions expressed in text data accurately. It will leverage machine learning techniques and can be applied to various applications, including social media monitoring, customer feedback analysis, and market research

**CONTENTS**

**CHAPTER NO. TITLE PAGE NO.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1. | INTRODUCTION |  |
|  | 1.1 Problem definition | 1 |
|  | 1.2 Objective of project | 1 |
| 2. | 1.3 Limitations of project  ANALYSIS   * 1. Introduction   2. Software Requirement specification      1. Software Requirement | 1  2  2 |
|  | 2.2.2 Hardware Requirement | 2 |
|  | 2.3 Existing System | 3 |
|  | 2.4 Proposed System | 4 |
|  | 2.5 Architecture | 4 |
| 3. | DESIGN |  |
|  | 3.1 Introduction | 5 |
|  | 3.2 UML Diagram | 5 |
|  | 3.3 Data Set Descriptions | 6 |
|  | 3.4 Data Preprocessing Techniques | 6-7 |
|  | 3.5 Methods & Algorithms | 7-9 |
|  | 3.6 Building a Model | 10 |
|  | 3.7 Evaluation | 11 |
| 4. |  | DEPLOYMENT AND RESULTS |  |
|  |  |  |  |
|  |  | 4.1 Source Code | 11-22 |
|  |  | 4.2 Final Results | 23-28 |
| 5. |  | CONCLUSION |  |
|  |  | 5.1 Project Conclusion | 29 |
|  |  | 5.2 Future Scope | 29 |
| 6. |  | REFERENCES | 30 |

# INTRODUCTION

## PROBLEM DEFINITION:

The problem of sentiment analysis on a product review website project can be broken down into the following tasks:

Collecting and preprocessing a dataset of product reviews

Defining the criteria for positive, negative, and neutral sentiment

Developing a model that can accurately classify the sentiment of reviews

Evaluating the performance of the model on a holdout dataset

Deploying the model to the product review website and using it to analyze new reviews in real-time

The success of the sentiment analysis project depends on the accuracy of the model in classifying the sentiment of reviews. By accurately identifying the sentiment of reviews, companies can improve their products and services and provide a better experience for their customers.

## OBJECTIVE OF PROJECT:

The main objectives of the sentiment analysis project are:

**Customer feedback analysis:** To understand customer opinions and preferences regarding a particular product or service by analyzing their reviews. By analyzing customer feedback, companies can identify areas for improvement and make data-driven decisions to improve their products and services.

**Reputation management:** To monitor and manage the online reputation of a product or service by analyzing customer reviews. Companies can use sentiment analysis to identify negative reviews and address them promptly to improve their online reputation.

**Competitive analysis:** To compare the sentiment of reviews of a particular product or service against its competitors. By analyzing the sentiment of reviews, companies can identify their strengths and weaknesses compared to their competitors.

**Customer experience improvement:** To improve the overall customer experience by analyzing feedback from customer reviews. By analyzing customer feedback, companies can identify pain points and make changes to improve the customer experience.

**Sales and marketing optimization:** To use sentiment analysis to optimize sales and marketing efforts. By analyzing customer feedback, companies can identify the features or aspects of their products that are most valued by customers and highlight them in their sales and marketing efforts.

Overall, the objective of sentiment analysis on a product review website project is to help companies gain valuable insights into customer feedback and make data-driven decisions to improve their products and services and provide a better experience for their customers.

## 1.3 LIMITATIONS OF THE PROJECT:

**Accuracy of sentiment analysis:** Sentiment analysis is not always accurate, and the results may be influenced by factors such as sarcasm, irony, and context. Therefore, the accuracy of the sentiment analysis model should be evaluated carefully to ensure that the website's results are reliable.

**Data bias**: The accuracy of the sentiment analysis model can be influenced by the training data used to develop it. If the training data is biased or incomplete, the model's results may be inaccurate. Therefore, it is important to ensure that the training data is diverse and representative.

**Limited languages:** Sentiment analysis may only be performed in specific languages, so the website may not be able to handle text inputs in languages that are not supported by the sentiment analysis model.

**Technical complexity:** The development and deployment of machine learning algorithms can be complex and require significant technical expertise. Therefore, the project may require a skilled team of developers and data scientists.

**Cost:** The development and maintenance of a sentimental analysis website can be expensive, including the cost of hardware, software, and personnel. Therefore, the cost of the project should be carefully evaluated before proceeding.

**Privacy and security concerns:** The website will store user data, so privacy and security concerns should be addressed to protect user information from unauthorized access or disclosure.

Overall, the limitations of a sentimental analysis website project should be carefully considered before proceeding to ensure that the project is feasible and effective

* 1. **INTRODUCTION:**

# 2. ANALYSIS

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative our neutral.it is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. However, analysis of social media streams is usually restricted to just basic sentiment analysis and count based metrics.

Sentimental analysis is required as it stores data in an efficient, cost-friendly. Sentimental analysis solves real-time issues and can help you solve all the real-time scenario. According to the survey,80% of the world’s data is unstructured. the data needs to be analyzed and be in a structured manner whether it is in the form of emails, texts, documents or feedbacks.

Required Sentimental analysis has wide range of applications. For suppose we take social media if for instance the comments o social media side as Instagram, over here all the reviews are analyzed and categorized as positive, negative, and neutral. In Customer services in the play store, all the comments in the form of feedback ratings up to five stars with the help of the sentimental analysis approach.in this way by using sentimental analysis we can know the opinion or behaviour of the reviewer.

**2.2 SOFTWARE REQUIREMENT SPECIFICATION:**

The software requirements specification (SRS) for a sentimental analysis website outlines the functional and non-functional requirements of the website, including the features, user interface, and performance parameters. The interaction between different customers ad contractor is done because its necessary to fully understand needs of customers. Depending upon information gathered after interaction, SRS is developed which describes requirements of software that may include changes and modifications that is needed to be done to increase quality of product and to satisfy customer’s demand.

* + 1. **Software Requirements:**
       1. Jupyter Notebook.
       2. Anaconda navigator.
       3. Html
       4. Css
       5. JavaScript

## Hardware Requirements:

* + - 1. Minimum 1GHz i.e., basic CPU and GPU
      2. Basic Ram i.e., 4-8 GB is optimal
      3. Storage Space for Dataset and saving real-time analysis data

## EXISTING SYSTEM

A sentiment analysis project involves using natural language processing and machine learning techniques to analyze text data and identify the sentiment expressed within it. The existing system of a sentiment analysis project typically consists of the following components:

**Data Collection:** The first step in building a sentiment analysis system is to collect data from various sources. This data can be in the form of tweets, comments, reviews, or any other form of text data.

**Data Preprocessing**: Once the data is collected, it needs to be preprocessed to remove irrelevant information, stop words, and noise. Preprocessing techniques may include tokenization, stemming, and lemmatization.

**Sentiment Analysis Algorithm:** The core component of a sentiment analysis system is the sentiment analysis algorithm. There are various techniques for sentiment analysis, such as rule-based, lexicon-based, and machine learning-based approaches. The algorithm should be trained on a labeled dataset that consists of positive, negative, and neutral sentiment.

**User Interface:** A user interface is required to interact with the sentiment analysis system. The user interface can be designed using HTML, CSS, and JavaScript. The interface should allow users to input text data and display the sentiment analysis results.

**Backend:** The backend of a sentiment analysis system can be built using a programming language like Python or Java. The backend is responsible for processing the user input and running the sentiment analysis algorithm.

**Deployment:** Once the sentiment analysis system is built, it needs to be deployed to a server or a cloud platform for users to access it.

**Continuous Improvement:** The sentiment analysis system should be continuously improved based on user feedback and the changing language trends. This may involve retraining the sentiment analysis algorithm on new data or updating the preprocessing techniques.

## PROPOSED SYSTEM



Sentiment analysis on product review websites is a valuable task that can provide insights into how customers feel about a particular product or service. Here are some proposed systems for sentimental analysis on product review websites:

**Rule-based system:** This approach involves creating a set of rules that can classify the sentiment of the text. The rules can be based on specific keywords or patterns that are commonly associated with positive or negative sentiment.

**Machine learning-based system:** This approach involves training a machine learning model on a dataset of product reviews that have been labeled with their corresponding sentiment. The model can then be used to classify new reviews based on their sentiment.

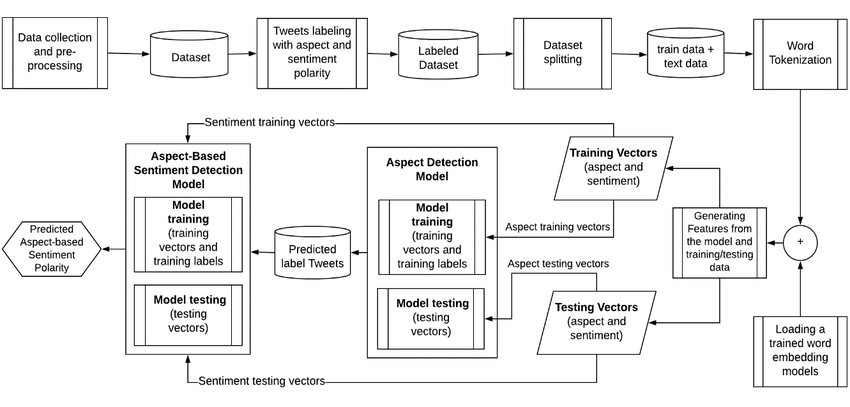
**Hybrid system:** This approach combines both rule-based and machine learning-based approaches to improve the accuracy of the sentiment analysis. The rule-based system can be used as a pre-processing step to identify the sentiment of the text, which can then be used to train a machine learning model.

**Aspect-based sentiment analysis:** This approach focuses on analyzing the sentiment of specific aspects of the product, such as the quality of the product, the customer service, or the price. This approach can provide more detailed insights into how customers feel about different aspects of the product.

**Deep learning-based system**: This approach involves using deep neural networks to perform sentiment analysis on product reviews. This approach has been shown to achieve state-of-the-art performance on various sentiment analysis tasks.

Ultimately, the choice of system will depend on factors such as the size of the dataset, the level of accuracy required, and the resources available for implementing the system.

* 1. **ARCHITECTURE**



# 3.DESIGN

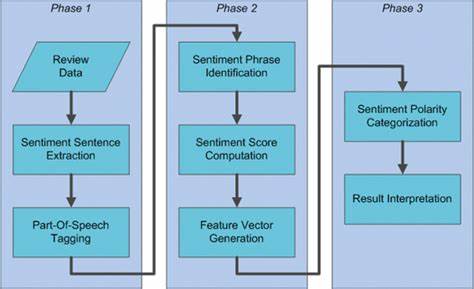
## INTRODUCTION

Sentiment analysis is a powerful tool that can help businesses understand how their customers feel about their products or services. In this project, we will be using sentiment analysis to analyze product reviews on a website and extract valuable insights from them. Our goal is to build a system that can accurately identify the sentiment expressed in a review, whether it is positive, negative, or neutral.

To accomplish this, we will be using natural language processing (NLP) techniques to analyze the text of each review and classify it based on the sentiment expressed. We will be using machine learning algorithms to train our system to recognize patterns in the text and make accurate predictions about the sentiment of each review.

By implementing sentiment analysis on product reviews, we can provide valuable insights to businesses, allowing them to identify areas where they can improve their products or services and better understand their customers' needs and preferences. This can ultimately lead to increased customer satisfaction, improved brand reputation, and higher sales.

## UML/ USECASE DIAGRAM

****

* 1. **DATA SET DESCRIPTIONS**

we need a dataset of product reviews with sentiment labels. The dataset should be representative of the products and the customers we are interested in analyzing. Here is a brief description of the dataset we will use:

**Dataset size:** We will need a large enough dataset to train a robust sentiment analysis model. A dataset with at least 10,000 reviews is recommended.

**Data source:** The dataset can be collected from various sources such as e-commerce websites, social media platforms, or review websites. It is essential to make sure the dataset is relevant to the products we are analyzing.

**Data format:** The data should be in a structured format, such as a CSV or JSON file, with each row containing the text of a review and its corresponding sentiment label (positive, negative, or neutral).

**Data pre-processing:** Before using the dataset, we may need to perform some pre-processing steps, such as removing irrelevant information (such as product names or dates) and cleaning up the text (removing stop words, punctuation, and special characters).

**Data labeling:** The dataset should be labeled with sentiment labels. We can use various techniques to label the dataset, such as manual labeling or using pre-trained sentiment analysis models to automatically label the reviews.

Overall, having a diverse and representative dataset that accurately captures the sentiment of the product reviews is crucial for building an effective sentiment analysis model.

## DATA PREPROCESSING TECHNIQUES:

There are several data processing techniques that we can use for sentimental analysis on a product review website project. Here are some of the most common techniques:

**Text normalization:** Text normalization refers to the process of converting the text to a standard format, such as converting all text to lowercase, removing punctuation, and expanding contractions.

**Tokenization:** Tokenization is the process of breaking down the text into individual words or phrases, called tokens, which can then be analyzed separately.

**Stop word removal:** Stop words are common words such as "the" or "and" that do not carry much meaning in the text. Removing stop words can help reduce the amount of noise in the data and improve the accuracy of the sentiment analysis.

**Stemming and Lemmatization:** Stemming and lemmatization are techniques used to reduce words to their root form. Stemming involves removing the suffix of a word, while lemmatization involves converting the word to its base form. This can help reduce the number of unique words in the data and improve the accuracy of the sentiment analysis.

**Feature extraction:** Feature extraction involves selecting the most relevant features from the text that can help predict the sentiment. This can be done using techniques such as TF-IDF (term frequency-inverse document frequency) or word embeddings.

**Sentiment lexicons**: Sentiment lexicons are lists of words and phrases that are associated with a particular sentiment. We can use sentiment lexicons to assign a sentiment score to each review based on the frequency of positive and negative words in the text.

Overall, using a combination of these data processing techniques can help us preprocess the data effectively and improve the accuracy of our sentiment analysis model on the product review website project

## 3.5 METHODS AND ALGORITHMS

Majorly 4 types of supervised machine learning algorithms are being using in implementing the project. They are

* + - Rule Based Methods
    - Machine Learning Algorithm
    - Deep Learning Algorithm
    - Ensemble Methods
    - Lexicon Based Methods

There are various methods and algorithms that can be used for sentiment analysis on a product review website project. Here are some common methods and algorithms:

**Rule Based Methods**: Rule-based methods use pre-defined rules and patterns to analyze the sentiment of a text. For example, a rule-based method might assign a positive sentiment score if the text contains words such as "good" or "excellent" and a negative sentiment score if the text contains words such as "bad" or "terrible."

**Machine Learning Algorithm**: Machine learning algorithms can be trained to recognize patterns in the text and make predictions about the sentiment of a review. Common machine learning algorithms used for sentiment analysis include logistic regression, support vector machines (SVM), and Naive Bayes classifiers.

**Deep Learning Algorithm**: Deep learning algorithms, such as recurrent neural networks (RNN) and convolutional neural networks (CNN), can be used for sentiment analysis. These algorithms can learn to represent the meaning of a sentence or a review in a high-dimensional space, allowing for more accurate sentiment analysis.

**Ensemble methods**: Ensemble methods combine the predictions of multiple algorithms to improve the overall accuracy of the sentiment analysis. For example, we can combine the predictions of a rule-based method, a machine learning algorithm, and a deep learning algorithm to get a more accurate prediction of the sentiment.

**Lexicon Based Methods** : Lexicon-based methods use sentiment lexicons, which are lists of words and phrases that are associated with a particular sentiment. The sentiment of a text can be calculated based on the frequency of positive and negative words in the text.

Overall, the choice of method and algorithm for sentiment analysis on a product review website project depends on the specific requirements and goals of the project. A combination of methods and algorithms can also be used to improve the accuracy of the sentiment analysis.

* 1. **BUILDING A MODEL**

## To build models for sentiment analysis on a product review website project, we can follow these general steps:

## Data collection : Collect a dataset of product reviews with sentiment labels. The dataset should be representative of the products and customers we are interested in analyzing.

## Data pre-processing: Preprocess the dataset using techniques such as text normalization, tokenization, stop word removal, stemming and lemmatization, feature extraction, and sentiment lexicons.

## Feature engineering: Extract relevant features from the pre-processed data, such as bag-of-words representations or word embeddings.

## Model selection: Select a suitable machine learning or deep learning algorithm for the sentiment analysis task. Common algorithms include logistic regression, support vector machines (SVM), Naive Bayes classifiers, recurrent neural networks (RNN), and convolutional neural networks (CNN).

## Model training: Train the selected algorithm on the pre-processed and feature-engineered dataset. Use techniques such as cross-validation and hyperparameter tuning to optimize the performance of the model.

## Model evaluation: Evaluate the performance of the trained model on a holdout dataset using metrics such as accuracy, precision, recall, and F1-score.

## Model deployment: Deploy the trained model to the product review website and use it to automatically analyze the sentiment of new reviews.

It is important to note that building an effective sentiment analysis model requires a combination of domain knowledge, data pre-processing techniques, feature engineering, and algorithm selection. It is also crucial to continuously monitor the performance of the model and update it as necessary to maintain its accuracy over time.

## Evaluation



It is essential to ensure that the model accurately predicts the sentiment of the reviews. Here are some common evaluation metrics for sentiment analysis:

Accuracy: The proportion of correct predictions made by the model.

Precision: The proportion of positive predictions that are correct.

Recall: The proportion of positive reviews that are correctly predicted.

F1-score: The harmonic mean of precision and recall.

**AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** A measure of the model's ability to distinguish between positive and negative reviews.

**Confusion matrix:** A matrix that shows the number of true positive, true negative, false positive, and false negative predictions made by the model.

When evaluating the performance of a sentiment analysis model, it is important to use a holdout dataset that the model has not seen during training. This helps to ensure that the evaluation is unbiased and that the model can generalize well to new data.

In addition, it is important to consider the specific requirements and goals of the product review website project when evaluating the sentiment analysis model. For example, if the website's main goal is to identify negative reviews, then precision may be more important than recall. On the other hand, if the website's main goal is to identify all negative reviews, then recall may be more important then precision.

4.Deployment and Result:

4.1 Source code:

import tkinter as tk

from tkinter import messagebox

from tkinter import scrolledtext

from nltk.sentiment import SentimentIntensityAnalyzer

def analyze\_sentiment():

text = text\_entry.get("1.0", tk.END).strip()

if text:

sia = SentimentIntensityAnalyzer()

sentiment\_scores = sia.polarity\_scores(text)

compound\_score = sentiment\_scores['compound']

if compound\_score >= 0.05:

sentiment = "Positive"

elif compound\_score <= -0.05:

sentiment = "Negative"

else:

sentiment = "Neutral"

messagebox.showinfo("Sentiment Analysis Result", f"Sentiment: {sentiment}\nCompound Score: {compound\_score}")

else:

messagebox.showwarning("No Text", "Please enter some text.")

# Create the main window

window = tk.Tk()

window.title("Sentiment Analysis")

window.geometry("400x300")

# Create the label and text entry field

label = tk.Label(window, text="Enter text:")

label.pack()

text\_entry = scrolledtext.ScrolledText(window, height=5)

text\_entry.pack()

# Create the analyze button

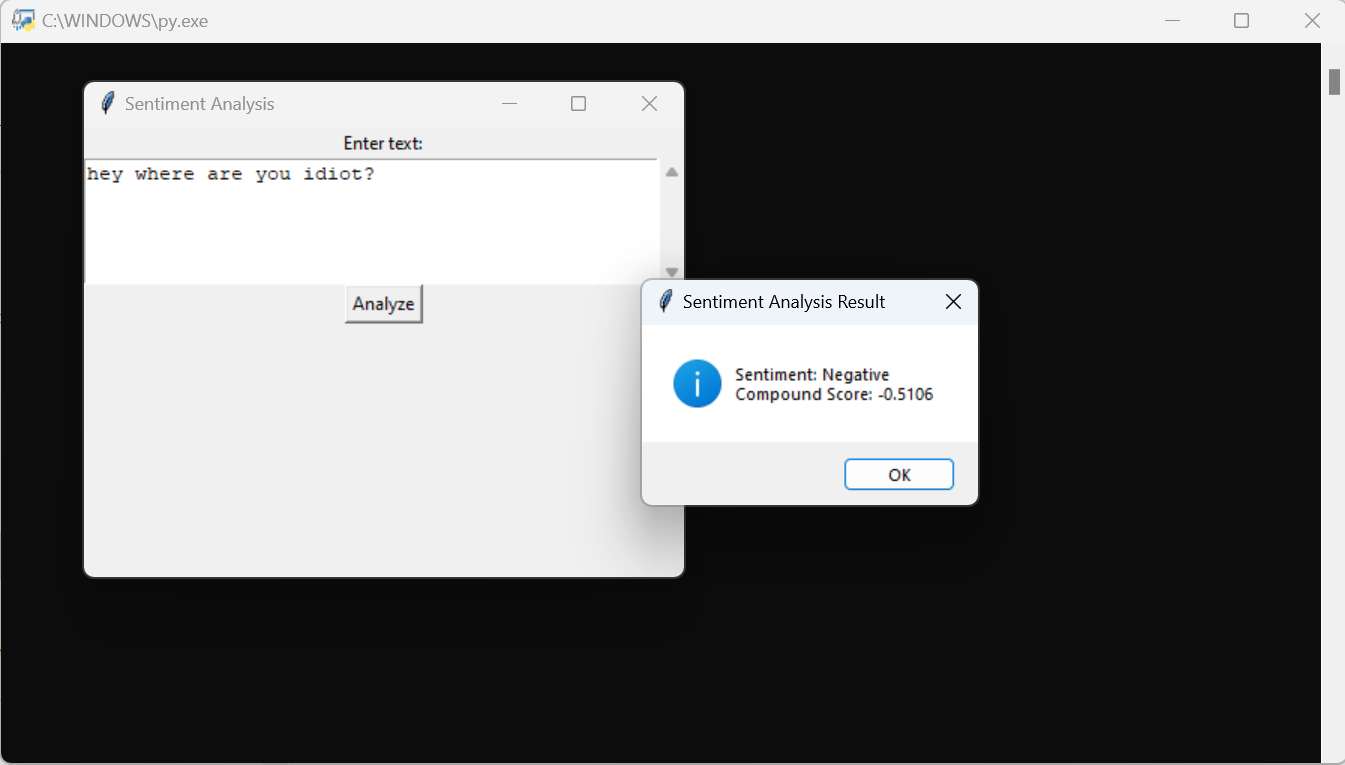
analyze\_button = tk.Button(window, text="Analyze", command=analyze\_sentiment)

analyze\_button.pack()

# Start the main loop

window.mainloop()

4.2 Final Result:



In Our project we used python 3.11 and based on the code we have downloaded all the packages which are necessary in our project execution like nltk package, tkinter package etc.

By using the command senti.py we have executed our code in python 3.11

Cmd. And we got the output.

5.Conclusion:

5.1 project conclusion:

By analyzing sentiment, we can uncover patterns, trends, and sentiments hidden within textual data, allowing businesses and organizations to make informed decisions, tailor their products and services to customer preferences, and enhance customer satisfaction. Sentiment analysis also plays a vital role in social listening, enabling individuals and companies to understand public sentiment, detect emerging trends, and respond effectively to customer needs and concerns.

As sentiment analysis continues to evolve, it is essential to consider ethical implications, ensuring privacy, fairness, and transparency in the use of user-generated data. Additionally, ongoing research and development in the field will lead to more sophisticated models, enabling deeper understanding and more nuanced sentiment analysis.

In conclusion, sentiment analysis is a rapidly growing field that offers valuable insights into human sentiment and behavior, revolutionizing the way we understand and interact with textual data. By leveraging sentiment analysis effectively, individuals, businesses, and organizations can gain a competitive edge, make data-driven decisions, and foster stronger connections with their target audience.

5.2 Future Scope:

Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings and emotions regarding their brand. The brands through the use of social media data, survey responses and other hubs of user-generated content. By investigating and analyzing customer sentiments, these brands are able to get an inside look at consumer behaviors and, ultimately, better serve their audiences with the products, services and experiences they offer.

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens. This forecast also predicts broader applications for sentiment analysis – brandswill continue to leverage this tool, but so will individuals in the public eye, governments, nonprofits, education centers and many other organizations.

6.Reference:

Our project Sentimental Analysis code is suggested by chatGpt and also we have made some changes in the code .We have implemented using popular libraries such as NLTK and scikit-learn in Python.

This code uses the SentimentIntensityAnalyzer from the NLTK library to analyze the sentiment of a given sentence. It assigns sentiment scores to the sentence and interprets them to determine whether the sentiment is positive, negative, or neutral.

We also used Tkinter in our code which provides a wide range of widgets and options for building interactive GUI applications.

ScrolledText widget, which provides a text area that can handle longer texts. Users can enter or paste text into the widget. When the "Analyze" button is clicked, the analyze\_sentiment function is called. Inside this function, sentiment analysis is performed on the text extracted from the ScrolledText widget using the SentimentIntensityAnalyzer from the NLTK library. The sentiment result is then displayed in a message box using messagebox.showinfo().