SENTIMENT EXTRACTION

Presented By

Rohan Dodiya
190470107015
Department of Computer Engineering

Presented to

Prof. Dipesh Joshi

Table of Content

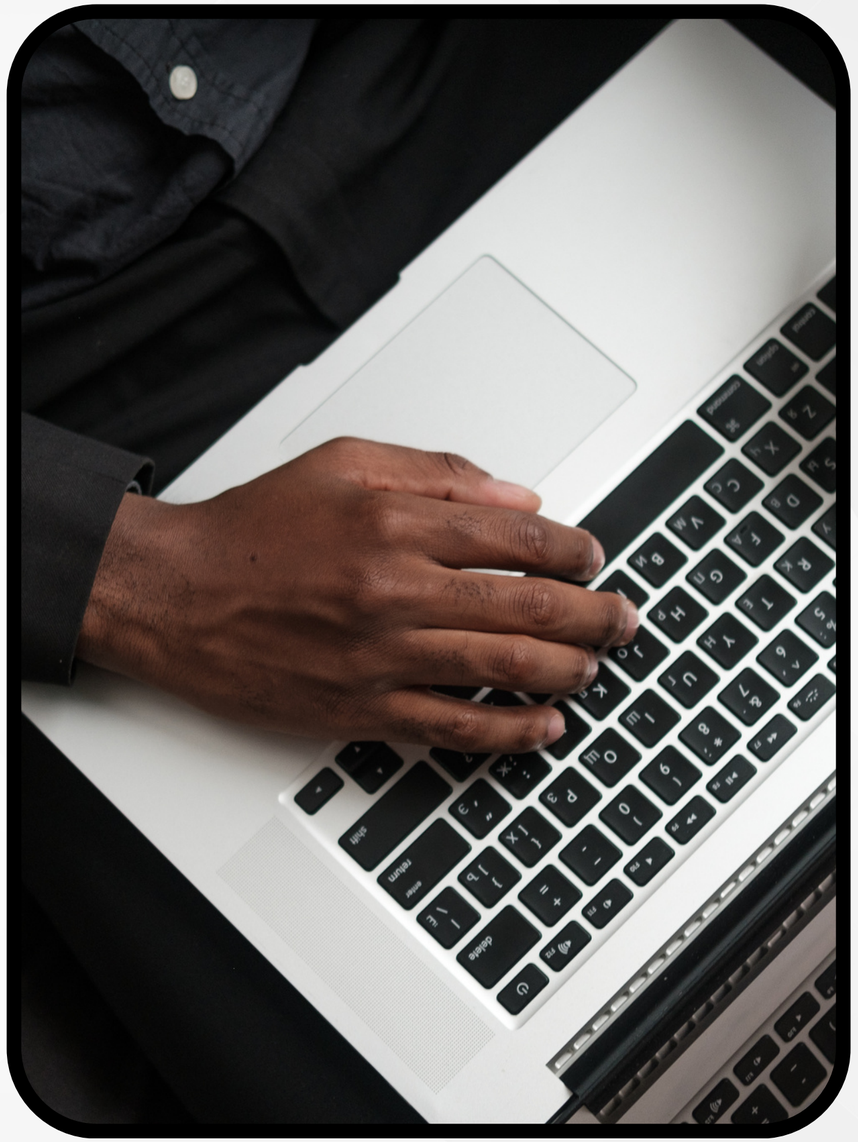
COMAPNY PROFILE

- In July 2021, Dhrima Solutions LLP was founded in the data science & analytics field, marking the start of an exciting chapter in the world of business intelligence.
- They offer a wide range of services including data analysis, predictive modeling, machine learning and data engineering
- The team is made up of experienced data scientist, analyst & engineer who have deep understanding of statistical analysis, programming language & machine learning algorithms.

dhrima
solutions LLP

INTRODUCTION TO INTERNSHIP

- I did my internship at Dhrima Solutions as data scientist intern.
- During my internship as a data science intern, I had the opportunity to collect and clean data from various sources. I applied statistical and machine learning techniques to build predictive models that could help businesses make informed decisions.
- Throughout my internship, I had the privilege of working alongside experienced data scientists who provided valuable mentorship and guidance.
- By the end of my internship, I had gained hands-on experience with various tools and techniques, as well as skills in data wrangling, statistical analysis, and machine learning.
- Overall, my internship provided a solid foundation for launching a career in data science, which I am excited to pursue further.



TOOLS AND TECHNOLOGIES

- Python is a popular programming language for data science due to its simplicity, flexibility, and vast array of libraries and tools designed specifically for data analysis and visualization.
- Python can be used for various data science tasks such as data wrangling, statistical analysis, machine learning, and deep learning.
- It has become a go-to language for data scientists due to its versatility and its ability to handle both structured and unstructured data.

- NLP is a subfield of artificial intelligence that involves developing algorithms and models that can understand, analyze, and generate human language.
- It has become a critical tool in the data science field, enabling data scientists to extract insights from large amounts of unstructured text data.

TOOLS AND TECHNOLOGIES

- Machine learning is a subfield of artificial intelligence that allows computers to learn from data and make predictions or decisions without being explicitly programmed.
- It has become an essential tool in the data science field, enabling data scientists to extract insights and value from large amounts of data.

- Flask is lightweight web application framework for Python. It is designed to be simple and
- easy to use, making it a popular choice for building web applications and APIs. Flask
- provides a range of features, including URL routing, templating, and support for database
- integration, that allow developers to quickly build and deploy web applications

OVERVIEW OF

PROJECT

What is sentiment analysis

Sentiment analysis is a type of natural language processing that involves analyzing text to determine the emotional tone and underlying attitudes of the writer. It involves using algorithms and machine learning techniques to classify text as either positive, negative, or neutral in terms of sentiment.

Sentiment analysis is important for a number of reasons. For one, it can help businesses and organizations better understand their customers and their needs by analyzing social media posts, customer feedback, and other forms of text data.

This information can be used to improve customer service, product development, and marketing strategies. Sentiment analysis can also be used to monitor brand reputation, identify emerging trends and issues, and gauge public opinion on various topics.

OVERVIEW OF PROJECT

How sentiment extraction tool works

- The system uses a random forest algorithm to train the sentiment analysis model.
- Random forest is a powerful machine learning algorithm that combines multiple decision trees to improve accuracy and reduce overfitting.
- The system is integrated with the Flask framework, which provides a user-friendly interface to access the sentiment analysis API.
- The API can be used to analyze text inputs, files, and reviews.
- The sentiment analysis system was evaluated using a test dataset and achieved an accuracy of 70% precision of 70%, and recall of 69%.
- The system performed well in classifying positive, negative, and neutral sentiment for different types of text inputs.

Data

- The data used in the machine learning model was obtained from Kaggle and consists of approximately 30,000 text samples along with their corresponding sentiment categories.
- The dataset is useful for training natural language processing models to accurately classify text based on sentiment.

OVERVIEW OF

PROJECT

Limitations of tool

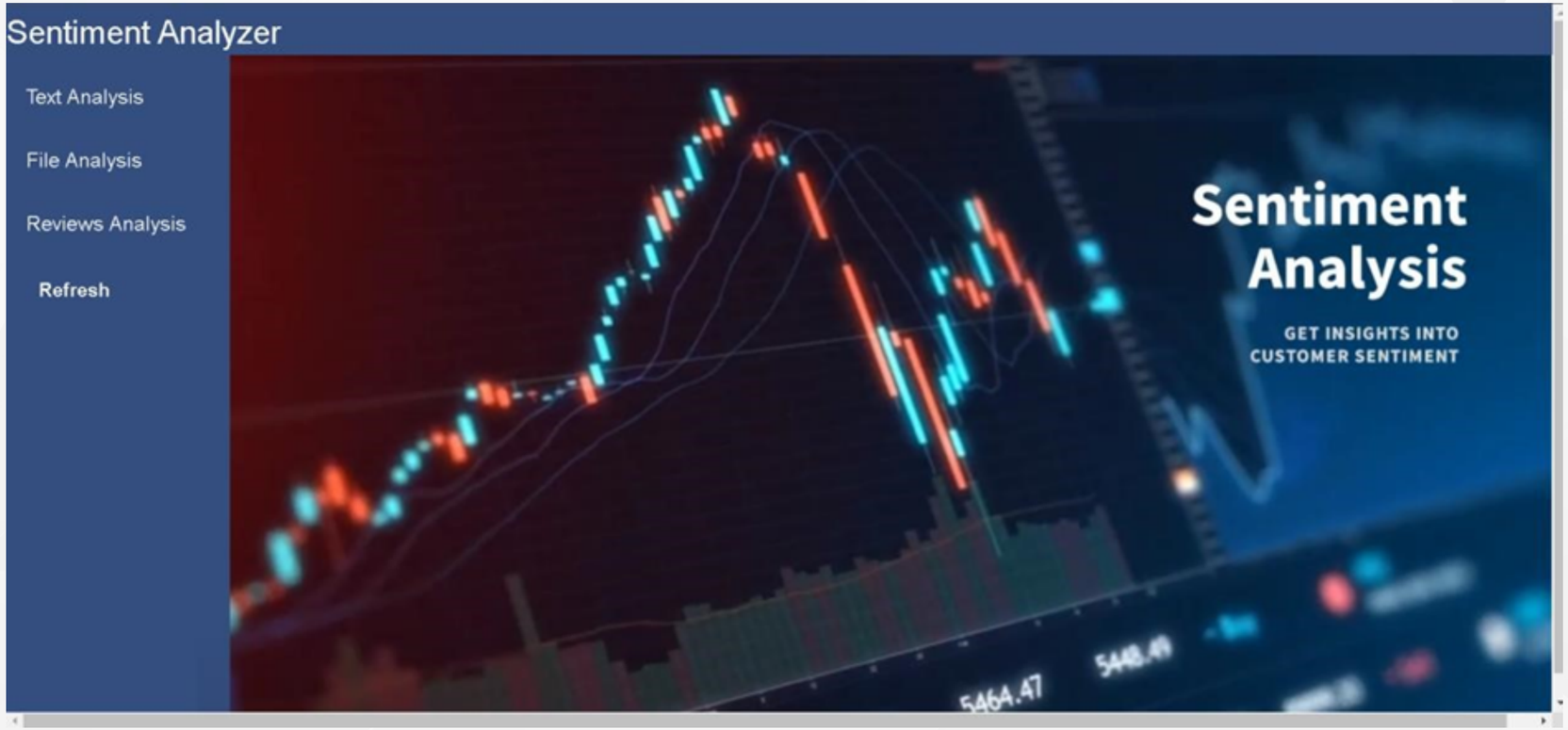
1.Language Limitation:
The sentiment analysis model is limited to one language, which means it can only classify the sentiment of text inputs written in that specific language.

2.Accuracy Limitation
The model is not 100% accurate and may not always provide correct sentiment classification for
inputs that contain slang or sarcasm. The accuracy of the model may be affected by the complexity of the text inputs and the level of ambiguity in the language used.

3.Reviews Division Limitation:
The reviews analysis division of the model is currently limited to Amazon reviews only. This means that the model may not be suitable for analyzing reviews from other sources.

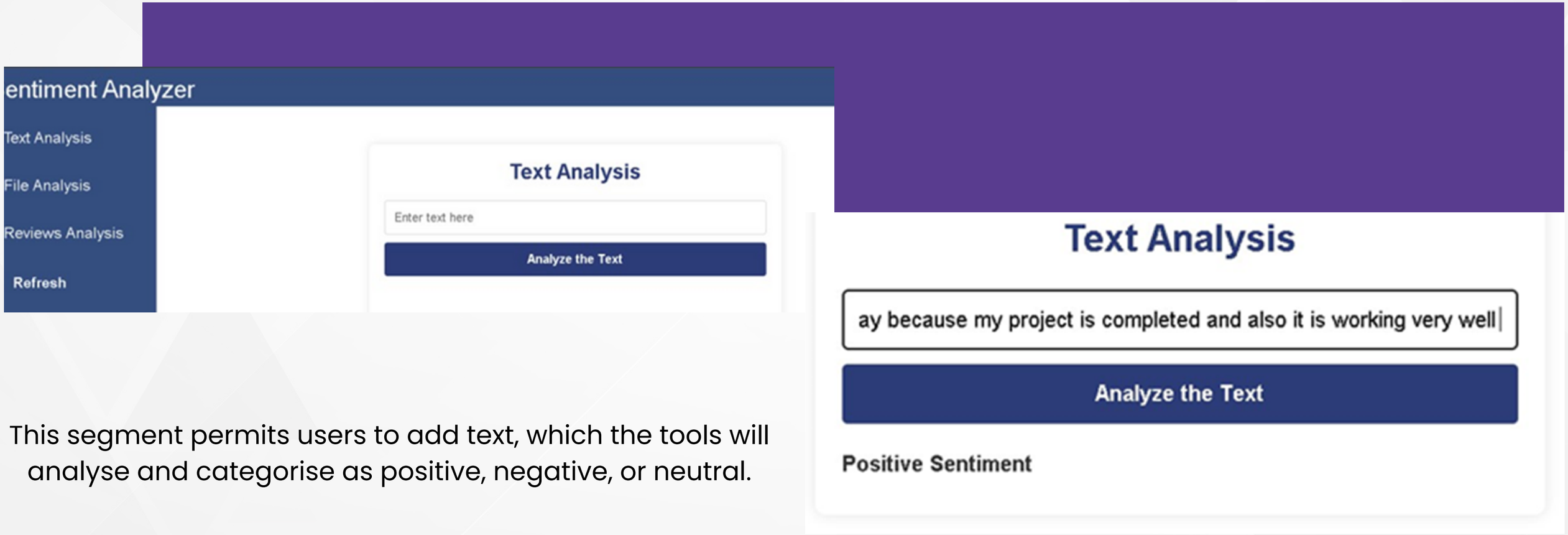
FEATURES OF SYSTEM

LANDING PAGE



FEATURES OF
SYSTEM

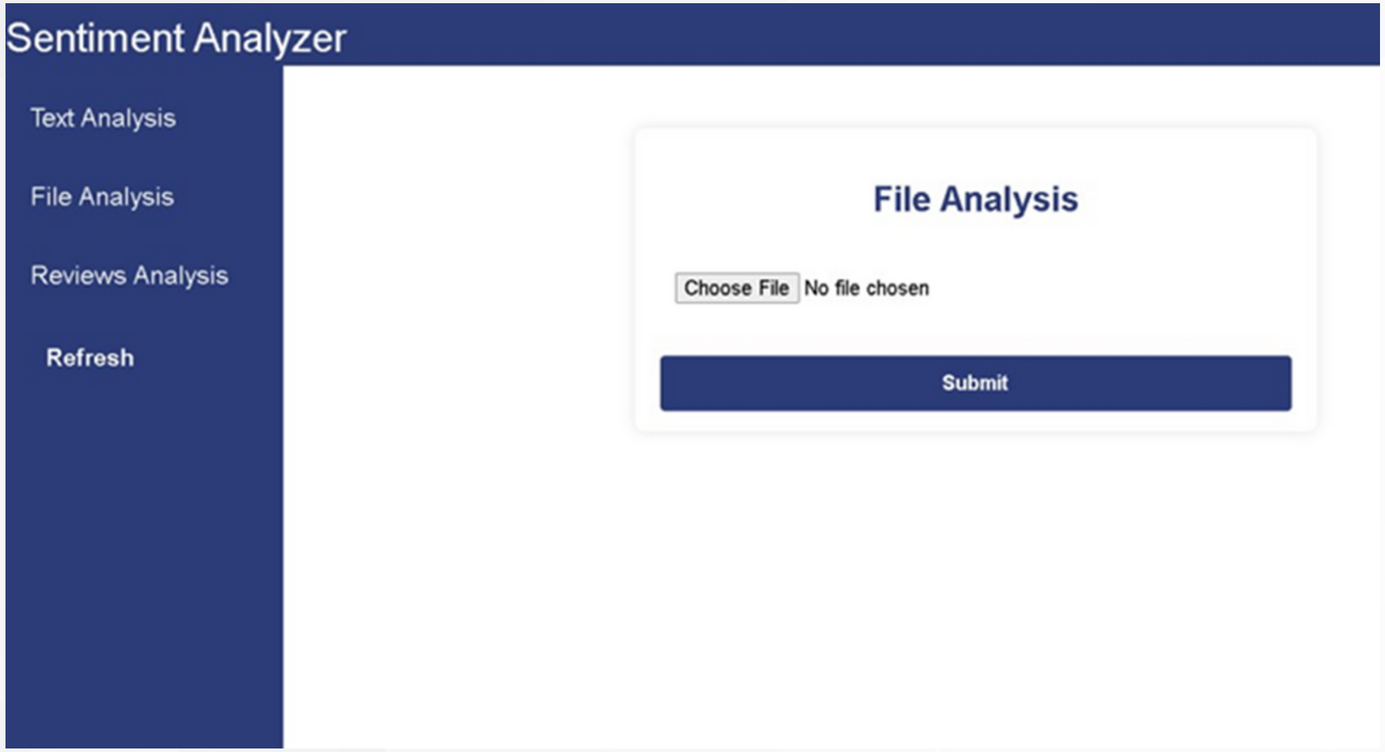
Text analysis

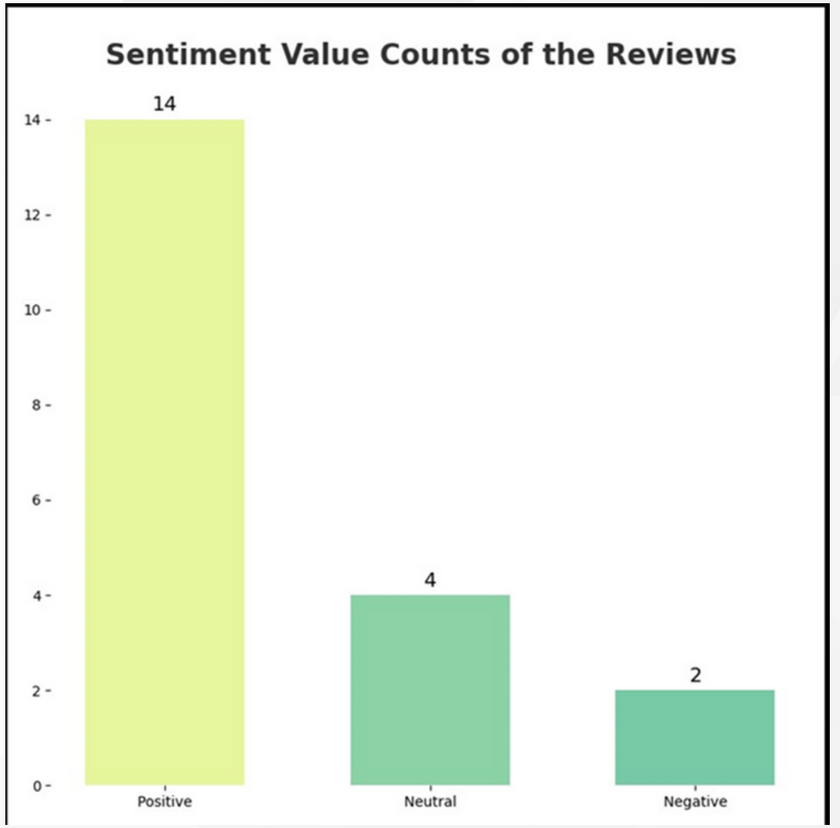


This segment permits users to add text, which the tools will analyse and categorise as positive, negative, or neutral.

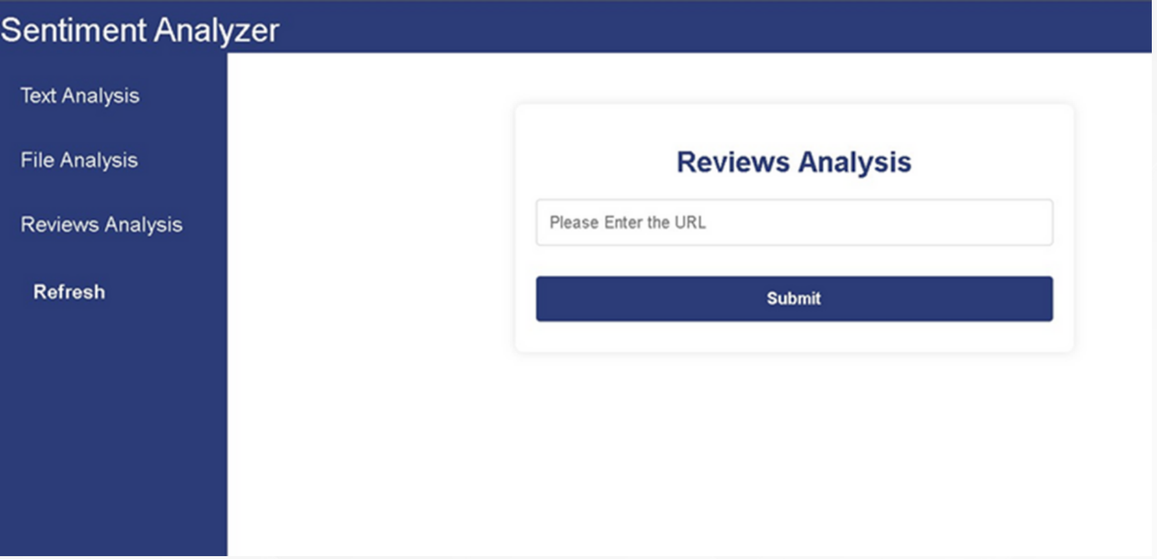
FEATURES OF SYSTEM

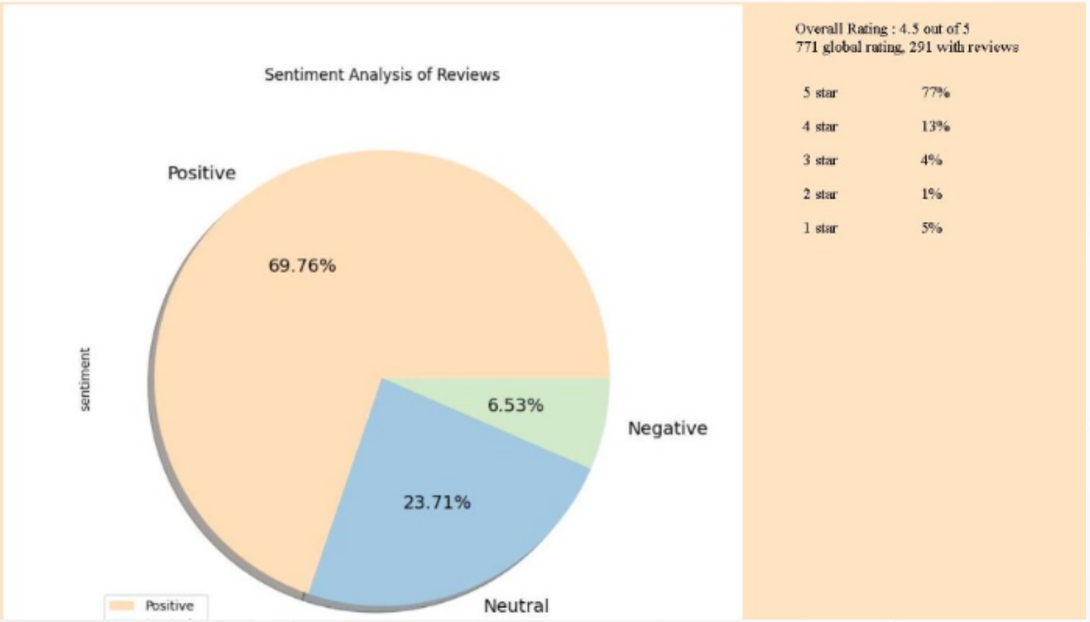
FILE ANALYSIS





The file analysis division is very helpful for business needs. Users can submit their csv files including reviews or any other writings they wish to analyse and gain insights from in this section. The tool will read the file's text and, after conducting sentiment analysis, produce a graph.





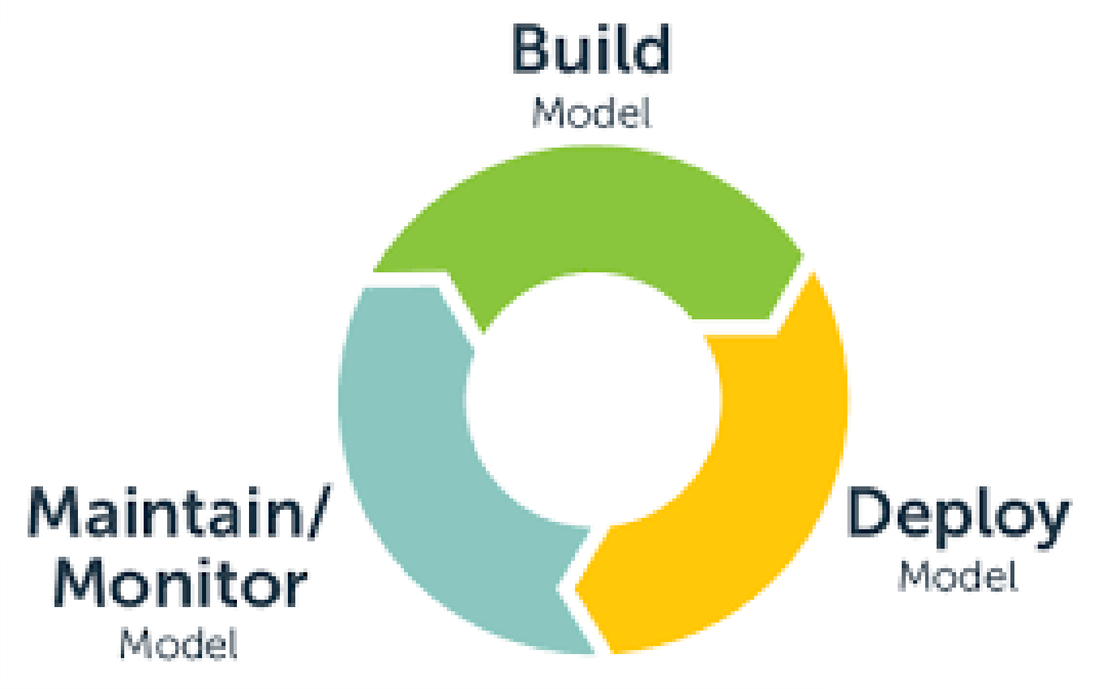
REVIEW ANALYSIS

FEATURES OF SYSTEM

- Customers' comments or reviews play a significant role in deciding how well a product is performing on the market, which is crucial information for businesses.
- This section will therefore enable you to learn more about the product.
- The model will first retrieve all the reviews using SELENIUM WEB DRIVER, then after applying the machine learning algorithm, it will provide a visualisation of what the customers tell about the product.
- The user only needs to enter the url of the product they want to analyse.

FEATURES OF SYSTEM

REFRESHING PREDICTIVE MODEL



- This feature is used to retrain the machine learning model .
- Retraining a predictive model with both the training data and past predictions is important
- Using past predictions as part of the training data can help to capture the model's previous errors and improve its accuracy going forward. By including past predictions in the training data, the model can learn from its past mistakes and avoid repeating them in the future.
- Retraining the model with past predictions can help to improve its ability to handle unexpected events or outliers that may not have been present in the original training data.
- By regularly retraining the model, data scientists can ensure that it remains accurate and effective, and that it continues to provide value to the organization.

REFERENCES

1.https://www.kaggle.com/competitions/tweet-sentiment-extraction/data

2. https://towardsdatascience.com

3. https://medium.com

4.https://stackoverflow.com/

5.https://www.w3schools.com

6.https://www.geeksforgeeks.org

THANK YOU