**Sentiment Analysis on Trip Advisor Hotel Reviews**

**Introduction:**

This project focuses on sentiment analysis conducted on hotel reviews sourced from Trip Advisor, a prominent platform for travel-related content. The analysis aims to extract insights from customer feedback, which can be instrumental in enhancing hotel services and customer satisfaction levels.

**Dataset:**

The dataset consists of 20,491 hotel reviews collected from Kaggle. It comprises two columns:

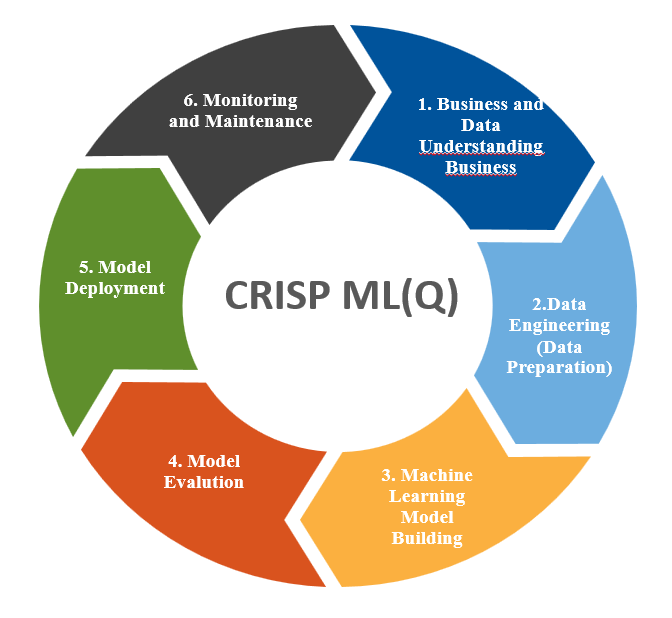
**Text Content:** This column contains the textual content of the reviews provided by customers.

**Star Ratings:** This column includes the corresponding star ratings given by customers, ranging from 1 to 5, reflecting their overall satisfaction level with the hotel experience.

The dataset has been thoroughly inspected, and no missing values were found in either column.

**Methodology: CRISP MLQ**

The sentiment analysis approach employed in this project utilizes the Crisp MLQ methodology



**Preprocessing Steps:**

* **Text Cleaning:** The reviews are cleaned to remove any unwanted characters, symbols, and punctuation marks.
* **Tokenization:** The cleaned text is tokenized, breaking down the text into individual words or tokens.
* **Stop Words Removal:** Commonly used words (stop words) that do not contribute to the sentiment are removed.
* **Lemmatization:** Words are lemmatized to their base or root form.
* **Vectorization:** The text data is transformed into numerical features using the CountVectorizer method, which converts the text into a matrix of token counts.

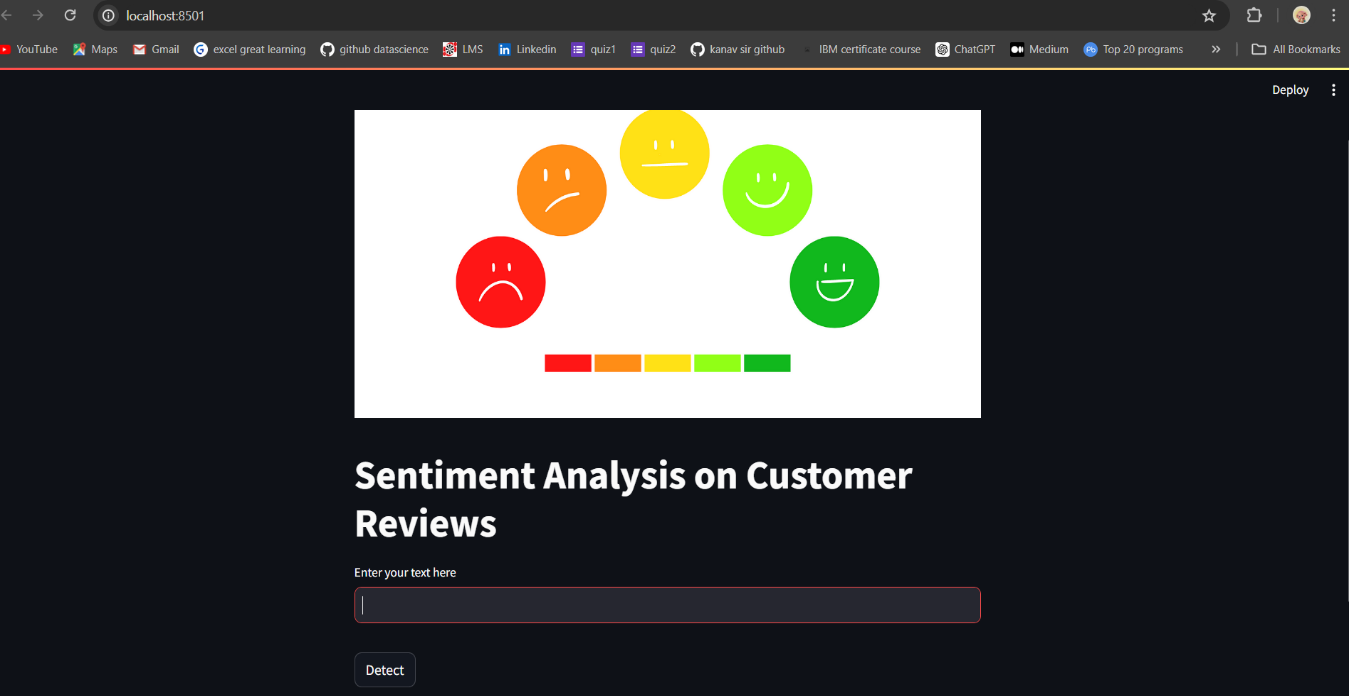
**Model Training and Evaluation:**

The dataset is split into training and testing sets using a 75-25 split ratio to train the model on a portion of the data and evaluate its performance on unseen data.

The models are trained and their hyperparameters are optimized using GridSearchCV and build models with Multinomial Naive Bayes, Logistic Regression, and XGBoost algorithms, which are commonly used techniques for text classification tasks.

**Results:**

The sentiment analysis model achieved an accuracy of 80% on the testing set, demonstrating its effectiveness in predicting sentiment from hotel reviews. Accuracy score metrics were computed for each sentiment class (positive, negative, neutral), providing insights into the model's performance across different sentiment categories.



**Discussion:**

* Analysis of the results revealed that a majority of the reviews were positive, with a significant portion corresponding to 4- and 5-star ratings.
* Negative reviews were less prevalent but highlighted areas for improvement, such as service quality, cleanliness, and amenities.
* Neutral reviews often contained constructive feedback or comments that did not strongly express either positive or negative sentiment.

**Applications:**

Insights derived from sentiment analysis can be leveraged by hotel management to identify strengths and weaknesses in their services, prioritize areas for improvement, and tailor marketing strategies to target specific customer segments.

Additionally, sentiment analysis can facilitate reputation management efforts by monitoring online reviews and responding promptly to customer feedback.

**Conclusion:**

In conclusion, sentiment analysis of Trip Advisor hotel reviews offers valuable insights into customer sentiments and perceptions. By harnessing machine learning techniques, hotels can optimize their service offerings, enhance customer experiences, and ultimately drive business success.

**References:**

**Trip Advisor dataset:** https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews