**Yelp Review Analysis and Rating Prediction**

**Background:**

In this project, we have used the [Yelp Restaurants](https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset) dataset which comprises of the US restaurants details along with the customer reviews. Performed data profiling, merged the restaurants data with the reviews, and finally segregated the restaurants based on Philadelphia city. Among the list of restaurants in Philadelphia, we have considered only one restaurant which has the highest customer reviews count.

**Problem Statement:**

We have considered the yelp reviews of an individual restaurant and analyzing their customer reviews to find out their overall emotion factors that affects the customer reviews. Figuring out the positive and negative review tokens which impact the ratings by implementing ML models for enhancing the customer service in the future.

**Obtained Results:**

* Performed contextual mining of text by identifying and extracting the customers reviews based on the sentiment analysis
* Analyzed the sentiment analysis of all the customer reviews for a particular restaurant and plotted the emotion bar chart
* Removed stop words, infrequent words and performed lemmatization to identify the polarity scores of the individual reviews
* Plotted word cloud based on the positive and negative review tokens
* Once performed sentiment analysis, we have further proceeded with the individual text token analysis by which we can be able to find the best or worst ratings for the individual review tokens
* Used the Count Vectorizer to create document term matrices of the review tokens and implemented the model by using Multinomial Naive Bayes
* As part of this project, we have trained and tested our data with three machine learning models
  + Random Forest Classifier
  + Stochastic Gradient Descent (SGD) Classifier
  + Support Vector Classifier
* Calculated the Accuracy, F1 score and RSME
* Plotted the confusion matrix and ROC curve for all the respective models