

# COGNIFYZ TECHNOLOGIES MACHINE LEARNING TASKS - DOCUMENTATION

## Problem Statement and Objective

Problem Statement:- There are four problem statements (tasks) given as per the internship task PDF document:

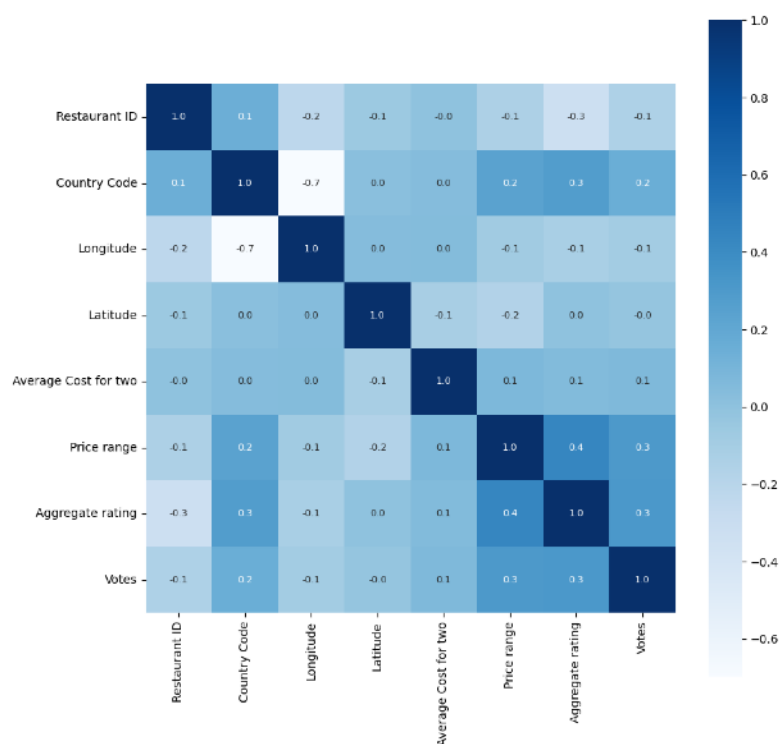
1. Build a machine learning model to predict the aggregate rating of a restaurant based on other features.
2. Create a restaurant recommendation system based on user preferences.
3. Develop a machine learning model to classify restaurants based on their cuisines
4. Perform a geographical analysis of the restaurants in the dataset.

The objective is to build a Prediction model for Task 1, Recommendation model for Task 2, Classification model for Task 3 and Clustering model for Task 4.

## Data Analysis, Visualization and Pre-Processing

The restaurant dataset was already provided, which contains of several features such as Restaurant ID, Name, City, Address, Longitude, Latitude, Cuisines, Average cost for two, Price Range, Aggregate Rating, etc. I have used data pre processing techniques such as:

1. Handling missing values in the dataset to ensure fairness in model training.
2. Performed data analysis such as Mean, Standard Deviation, etc
3. One-hot encoding "Cuisine" values from string to numerical attributes to train the model
4. Converting "Cuisines" String attributes into numerical attributes using Text Vectorization for Recommendation model.
5. Performed visualistic correlation between the parameters of the dataset:



## Model Selection and Development

The models that are selected for implementation are:

### 1. Random Forest Regressor Model for Rating Prediction

- i. Split the Dataset into two parts:- Training the Model(80% dataset) and Testing the Model(20% dataset). Again we separate them into X\_train, y\_train for training our model and X\_test and y\_test for the actual performance of the model.
- ii. Load the Random Forest Regressor Model, which is an ensemble of multiple decision trees, used especially for making predictions on continuous variables, and fit the X\_train and y\_train into the model.
- iii. Compare the predictions of both Training Data and Testing Data and generated the R-Square scores of each.
- iv. Built a predictive system, which takes the name of the restaurant to predict the aggregate rating of the restaurant.

### 2. Nearest Neighbours Algorithm with TF-IDF (Term Frequency-Inverse Document Frequency) for Restaurant Recommendation

- i. Create a TF-IDF vectorizer to transform the "Cuisines" column from string to numerical attributes and combined it with the other numerical features.
- ii. Create a Nearest Neighbours Model using Cosine Similarity, which measures the angle between the vectors and fit the required features into the model.
- iii. Built a recommendation system, which takes the important features such as Cuisines, Average cost for two and aggregate rating and recommends the top restaurants based on the user preferences.

### 3. Random Forest Classifier Model for Restaurant Classification

- i. Split the Dataset into two parts:- Training the Model(80% dataset) and Testing the Model(20% dataset). Again we separate them into X\_train, y\_train for training our model and X\_test and y\_test for the actual performance of the model.
- ii. Load the Random Forest Classifier Model, which is an ensemble of multiple decision trees used for categorical variables, and fit the X\_train and y\_train into the model.
- iii. Compare the predictions of both Training Data and Testing Data and generate the accuracy scores of each.
- iv. Built a classification-based system, which takes the name of the cuisine and classifies the restaurants based on the cuisine.

#### 4. K-Means Clustering for Geographical Restaurant Analysis

- i. Extract the values of Geographical Longitude and Latitude columns from the dataset.
- ii. Find the WCSS(Within Clusters Sum of Squares) value of the dataset.
- iii. Plot an Elbow Graph to find the value of minimum number of clusters that we can group into.
- iv. Train the Model according to the value of number of clusters we found from the elbow graph.
- v. Plot the clusters and their centroids using matplotlib.
- vi. Calculate statistics such as average ratings and price ranges by City.

### Visualizations and Insights

You can take a look at some of the output visuals and some statistical insights of the trained models:

```
[19] restaurant_name = "Vikings"
restaurant = dataset[dataset['Restaurant Name'].str.lower() == restaurant_name.lower()]
if not restaurant.empty:
    input_features = pd.get_dummies(restaurant.drop(columns=features_to_drop))
    input_features = input_features.reindex(columns=X.columns, fill_value=0)
    prediction = model.predict(input_features)[0]
    print(f"Predicted rating for {restaurant_name}: {prediction:.2f}")
else:
    print(f"Restaurant '{restaurant_name}' not found in the dataset.")

... Predicted rating for Vikings: 4.26
```

...

```
Top recommended restaurants based on your preferences:
Restaurant: Ooma, Cuisine: Japanese, Sushi
Restaurant: IZU, Cuisine: Japanese, Sushi
Restaurant: Sushi Haus, Cuisine: Japanese, Sushi
Restaurant: Aim Cafe And Restaurant, Cuisine: Japanese, Sushi
Restaurant: Kuuraku, Cuisine: Japanese
```

...

Restaurants serving 'Italian':

Restaurant Name	Cuisines
Cafe Arabelle	Cafe, American, Italian, Filipino
Nonna's Pasta & Pizzeria	Italian, Pizza
Wildflour Cafe + Bakery	Cafe, Bakery, American, Italian
The Food Hall by Todd English	American, Asian, Italian, Seafood
Villa Tevere	Italian
Gero	Italian
D.O.C Ristorante	Italian
Cantina Famiglia Mancini	Italian, Pizza
Paribar	Brazilian, Italian
Terra_o It@lia	Italian
Villa Gargano	Italian, Pizza
DePalma's Italian Cafe - Downtown	American, Italian, Pizza
DePalma's Italian Cafe - East Side	American, Italian, Pizza
La Dolce Vita Ristorante	Italian
Transmetropolitan	Italian, Pizza, Sandwich
Giuseppe's Pizza & Italian Specialities	Desserts, Italian, Pizza
Mellow Mushroom	Italian, Pizza, Sandwich
Flatbread Neapolitan Pizzeria	Italian, Pizza
Lucianos Italian Restaurant	Italian, Seafood, Vegetarian
Biaggi's Ristorante Italiano	Italian, Pizza
Monica's	American, Italian, Pizza
...	...
The Paddington	Cafe, Italian
The Manhattan Fish Market	Seafood, Italian
Cafemiz	World Cuisine, Mexican, Italian
Huqqa	Italian, World Cuisine

