**PERSONALIZED RESTAURANT RECOMMENDER SYSTEM**

****

Group Members:

1. Harris Lukundi,
2. Brian Muthama
3. Lynette Wangari
4. Chris Laaria
5. Beryl Agai
6. Henry Rono

# Business Understanding

## Business Overview

The dining industry in various states across the United States is a dynamic and diverse landscape, offering a wide array of options from local eateries to upscale restaurants. However, both locals and tourists often struggle to find restaurants that align with their specific preferences in terms of location, cuisine, and quality. The sheer number of choices, coupled with the lack of a centralized platform for personalized recommendations, makes it challenging for users to make informed dining decisions.

Traditional methods of discovering restaurants, such as relying on word-of-mouth or general review platforms, often fall short in delivering tailored suggestions that cater to individual tastes. These solutions tend to be too broad and do not provide real-time, location-based recommendations, leaving users—especially those in unfamiliar areas—frustrated in their search for the perfect dining spot. The need for a more intelligent, user-centric solution is increasingly evident as the dining scene continues to grow.

The personalized restaurant recommender system addresses this gap by offering users customized dining suggestions based on their preferences and historical data. Utilizing advanced technologies like machine learning, the system provides real-time recommendations that are highly relevant and location-specific. The platform’s database includes detailed information about a vast range of restaurants across various states, capturing essential attributes like business details, reviews, ratings, and operational hours.

By analyzing this comprehensive dataset and integrating it with user profiles, the system can deliver personalized recommendations that match user preferences for cuisine, ambiance, proximity, and more. Whether a user is looking for a cozy Italian brunch spot in California or an authentic Tex-Mex BBQ joint in Texas, the system can quickly suggest top-rated options that meet their criteria. This tailored approach not only enhances the dining experience but also supports local businesses by driving targeted traffic to their establishments.

The platform's intuitive interface ensures that users can easily refine their search and explore new dining options with confidence. This personalized restaurant recommender system stands out as a valuable tool, helping users navigate the rich culinary landscape with ease and satisfaction.

## Problem Statement

With the dining industry expanding, consumers are overwhelmed by numerous restaurant options, struggling to find places that match their specific preferences for location, cuisine, and quality. Current solutions often offer generic recommendations and lack real-time, location-based suggestions, leaving users, especially tourists, at a disadvantage. There is a pressing need for an intelligent, user-friendly recommender system that provides personalized, real-time dining options to help users efficiently navigate the diverse culinary landscape and enhance their dining experience.

## Proposed Solution

We propose developing a Restaurant Recommender System specifically for the United States. The system will:

* **Comprehensive Mapping:** Systematically map and categorize restaurants in the United States, grouping them by cuisine type, location, and other relevant factors.
* **Personalized Recommendations**: Enable users to input their current location and preferred cuisine type, delivering a ranked list of nearby restaurants based on their ratings and proximity.
* **Enhanced User Experience:** Offer an intuitive, user-friendly interface with integrated map features to simplify navigation and help users easily explore and select dining options.

## Metrics of Success

## Objectives

### Main Objective

Develop an intelligent, user-friendly restaurant recommender system that provides personalized recommendations based on user location and cuisine preferences, and additional data sources to ensure accuracy and relevance.

### Specific Objectives

1. **To evaluate business performance (**For example, you could investigate if businesses in certain categories or locations tend to receive higher or lower average ratings, or if higher review counts correlate with more positive reviews.)
2. Establish a comprehensive database of restaurants across the United States.
3. Provide content about different cuisines and dining etiquette to enrich the user experience.
4. Create machine learning algorithms to rank restaurants based on user location, cuisine preferences, and ratings.
5. Develop an intuitive, responsive web/mobile application.

## Challenges

1. Ensuring accurate, complete, and up-to-date data from various sources will require continuous validation and updates.
2. Managing the rate limits and costs associated with using APIs.
3. Designing an intuitive, user-friendly interface that caters to users with varying technical capability may require extensive testing and iteration.
4. Computational capabilities to manage the vast database of restaurants across various states

# Data Understanding

The data used to develop this project was collected from a vast Yelp database where reviews and business data was utilized. The business data contained a wide variety of businesses and was filtered to contain only the data from restaurants, coming to a total of 52,286 rows and 14 columns and stored as restaurant data. The reviews data was filtered to only include restaurant reviews and only for one year. This brought the total number of entries to 429,771 rows and 9 columns and was stored as users’ data.

The data was loaded using a dataloader class and a datainfo class to gather all the information together. A datacheck class was employed on the datasets with the following results: The restaurant data contained no duplicates or outliers but had some missing values. There were no duplicates, outliers or missing values in the users’ data.

Exploratory data analysis (EDA) was carried out on the restaurant data where the following was explored, counts of restaurants per state, frequency counts of ratings, frequency of restaurant names, distribution of star rating per state, and review counts vs stars rating.

EDA was also carried out for the users’ reviews data which showed seasonal trends with peaks around April and July and decline in reviews in between September and December. A word cloud was explored for the reviews and most common words included food, place, restaurant, order, service, good.

# Data Preparation

The users’ dataset was first prepared. There were no missing values, outliers or duplicates to handle. The columns useful, funny, cool were dropped as they did not contain any useful information for the recommendation system.

For the restaurant data, the attributes column was missing 0.8% of its dataset. These rows were dropped as attributes column is crucial for modeling. The other rows missing values i.e. address, postal code, hours were filled with Unknown as this information was not to be used in modelling and hence dropping then would result in losing out of data points. A value count of state column revealed 4 states with 1 restaurant each and these rows were dropped i.e. states containing NC, HI, XMS, MT and CO as recommending in a state with 1 restaurant will be flawed. The remaining 14 states were mapped with their full name for better understanding e.g. CA changed to California.

Feature Engineering was carried out on the categories column to explode the categories so as to filter each restaurant. The exploded column was then standardized to remove spacing between words. This was done so as to achieve a dataframe with only the specific cuisine/s in that column e.g. a row having Vietnamese, Food, Restaurants, and Food Trucks as a category was engineered to have Vietnamese as its category.

Further engineering was performed to combine city, state and address into one column called location which would be used to for modeling. The user data containing 429771 rows and 6 columns was saved in a cleaned user data csv, while the restaurant data containing 38,550 rows and 16 columns was pickled for ease of deployment.

# Modelling

The approach take to model the recommender system was split into 3:

1. Content Based Filtering
2. Collaborative Based Filtering
3. Deep Neural Networks