

Predictive Restaurant Recommender System

– Project Report

Introduction

In today's data-driven world, personalizing user experience has become a key differentiator. This project aims to build a **machine learning-based recommendation engine** that predicts which restaurant a customer is most likely to order from, given their **location, demographics, and historical order behavior**.

This model was developed as part of a data science internship assignment for **Soulpage IT Solutions** and involves working with anonymized order data, vendor information, and customer metadata.

Dataset Overview

The dataset is divided into training and testing segments:

Train Set:

train_customers.csv: 34,674 customers with profile data

train_locations.csv: customer locations with geospatial details

orders.csv: order history including vendor interactions

vendors.csv: restaurant details (tags, coordinates, ratings)

Test Set:

test_customers.csv, test_locations.csv: 9,768 customers needing vendor recommendations

Objective

To **predict the most likely restaurant (vendor)** a customer will order from, based on:

Past vendor interactions

Customer profile (age, gender, status, etc.)

Location behavior

Restaurant metadata (tag, ratings, location)

Methodology

1. Data Preprocessing

Handled missing values across datasets (e.g., DOB, ratings)

Normalized column types (string → numeric where needed)

Merged datasets using `customer_id`, `location_number`, and `vendor_id`

2. Feature Engineering

Encoded categorical variables like:

`language`, `gender`, `location_type`, `vendor_tag_name`

Extracted features from date columns (`created_at`, etc.)

Aggregated historical interactions per customer-vendor pair

3. Model Building

Model used: XGBoost Classifier

Target variable: `vendor_id` from `orders.csv`

Trained on combined features from customers, orders, vendors, and locations

4. Prediction

Generated probabilities for each vendor per test customer

Selected top vendor with highest probability using `np.argmax(probs, axis=1)`

5. Output

Final output file: `predicted_vendors.csv`

Format:

`customer_id, predicted_vendor`

4958, 33

4302, 33

5940, 33

...

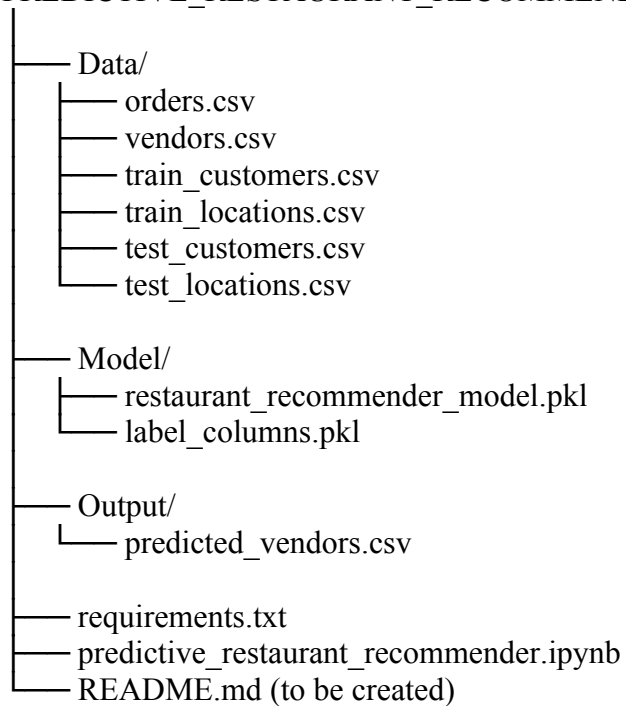
Results

The model successfully generates predictions for all customers in the test dataset. A sample output is shown below:

	customer_id	predicted_vendor
0	4958	33
1	4302	33
2	5940	33
3	3571	33
4	2194	33

Folder Structure

PREDICTIVE_RESTAURANT_RECOMMENDER/



Tools & Libraries Used

Python 3.11

Pandas, NumPy

Scikit-learn

XGBoost

Joblib

Matplotlib (for any visualizations)

Conclusion

This project demonstrates how historical behavior and contextual metadata can power personalized recommendations. It's scalable to real-world food delivery systems like Zomato, Swiggy, or Uber Eats. With further enhancements like:

Collaborative filtering

Real-time user behavior signals

Deep learning architectures (e.g., neural collaborative filtering)

...the system can be production-ready.