Task 3: Cuisine Classification

Company: Cognizify Technologies

Project Title: Cuisine Classification using Machine Learning

1. Introduction

This project focuses on classifying restaurants based on their cuisine type. It forms part of Task 3 in the Machine Learning Internship with Cognizify Technologies. The objective is to build a machine learning model that can accurately predict a restaurant's cuisine category using structured features such as city, price range, online delivery availability, and votes.

2. Objective

To develop a classification model that identifies the cuisine of a restaurant using relevant attributes from the dataset. This involves preprocessing, model building, and evaluation using standard ML practices.

3. Dataset Overview

The dataset contains information about restaurants, including:
- Cuisines (Target variable)
- City
- Price Range (1 to 4)
- Has Online Delivery (Yes/No)
- Number of Votes

4. Data Preprocessing

The preprocessing steps involved:
- Dropping rows with missing values.
- Extracting the primary cuisine from multi-label cuisine fields.
- Encoding categorical features like City and Has Online Delivery using Label Encoding.
- Encoding the target cuisine labels using LabelEncoder.
- Splitting the dataset into training and testing sets using stratified sampling to maintain class balance.

5. Model Used

The model chosen was the Random Forest Classifier. It is an ensemble method known for handling tabular data effectively and providing good accuracy, especially with categorical features.

6. Evaluation Metrics

The model was evaluated using the following metrics:
- Accuracy: Proportion of total predictions that were correct.
- Precision: Of all predicted cuisines, how many were correctly predicted.
- Recall: Of all actual cuisines, how many were correctly identified.
- Classification Report: Detailed performance report for each cuisine present in the test set.

7. Results

The model achieved moderate accuracy and highlighted that the most common cuisines were predicted more accurately. Some classes with fewer examples performed poorly due to data imbalance. Using stratified sampling helped improve class representation in both training and test sets.

8. Conclusion

This project successfully implemented a classification pipeline to predict restaurant cuisines. It followed proper ML workflows including preprocessing, model training, and evaluation. All requirements from Task 3 of the internship were fulfilled.

9. Future Enhancements

- Use advanced NLP on reviews or menu descriptions for better cuisine prediction.
- Apply multi-label classification to support restaurants with multiple cuisines.
- Use sampling techniques to handle data imbalance (e.g., SMOTE, class weights).
- Explore deep learning models or embeddings for richer feature representation.