📄 TASK 3 REPORT – CUISINE CLASSIFICATION

COGNIFYZ TECHNOLOGIES – MACHINE LEARNING INTERNSHIP

🧑‍💻 Intern Details:

- **Name**: YASHWANTH.M.C.

- **College**: Rajalakshmi Institute of Technology, Chennai

- **Department**: B.E. CSE – Artificial Intelligence and Machine Learning

- **Internship Duration**: June–July 2025

- **Task Title**: Cuisine Classification using Supervised Learning

---

🎯 Objective

To build a supervised machine learning classification model that predicts the **type of cuisine** a restaurant offers based on available features such as cost, location, delivery availability, and rating.

💼 Technologies & Tools Used

- Programming Language: Python

- Platform: Google Colab

- Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

- Model: RandomForestClassifier

- Dataset: Provided by Cognifyz Technologies (CSV)

🧪 Methodology

1. **Data Preprocessing**

   - Dropped entries with missing Cuisines values.

   - Selected the **Top 10 most frequent cuisines** for classification to reduce class imbalance and memory usage.

2. **Encoding**

   - Used LabelEncoder to convert categorical features (including Cuisines) into numerical format.

   - Dropped irrelevant high-cardinality features such as Restaurant Name, Address, and Locality Verbose.

3. **Model Training**

   - Used RandomForestClassifier() from scikit-learn.

   - Split dataset: 80% training and 20% testing using train_test_split().

4. **Model Evaluation**

   - Measured model accuracy

   - Generated classification report (Precision, Recall, F1-Score)

   - Analyzed performance using confusion matrix

📊 Results

- **Accuracy**: 31.49%

- **Classes (Top Cuisines)**: 10

- **Best-performing class (f1-score)**: Cuisine class 2 with **0.46**

- **Macro Avg F1-Score**: 0.26

- **Weighted Avg F1-Score**: 0.29

🔎 Interpretation: While accuracy is moderate, the model demonstrates effective learning in a **multi-class environment** with highly imbalanced data.

✅ Conclusion

This task helped me:

- Apply **supervised learning** to real-world food business data

- Understand **multiclass classification challenges**

- Practice **label encoding** and model evaluation using **precision, recall, and F1-score**

- Handle large datasets efficiently in **limited environments** like Colab

With further tuning and feature selection, model performance can be improved. This task significantly strengthened my understanding of practical classification models in machine learning.

🔗 Attachments:

- Notebook: Task3_Cuisine_Classification.ipynb

- Report PDF: Task3_Report_Yashwanth.pdf

- GitHub Repo: https://github.com/Yashwahthmc/TASK-3.git

✍️ Signature:

YASHWANTH.M.C.