**CuisineCamouflage-ClassificationModel for Meal Origins**

This project uses machine learning to classify dishes as homemade or restaurant-prepared, starting with a Basic Classifier for rice and chips, and advancing to an Advanced Classifier for detailed origin analysis, providing valuable insights for food businesses and enhancing consumer culinary experiences.

**Basic Classifier**

**Machine Learning Problem:** Build a classifier that differentiates images of dishes containing either rice or chips.

**Interest Point:** The ability to distinguish between rice and chips, despite their similar coloration, showcases the model's sensitivity to subtle textural and shape differences, which is critical in real-world food classification tasks.

In the Basic Classifier project, I processed the MLEnd Yummy Dataset images for uniformity, split them into training and testing sets, and extracted key yellow color and texture features. After normalizing these features, I trained **LinearSVC** and **RandomForestClassifier** models and evaluated their performance, selecting the superior model based on confusion matrix results.

**Advanced Classifier**

**Machine Learning Problem:** Develop a classifier to predict whether a dish in an image is homemade or Non-Homemade.

**Interest Point:** This model explores the nuances of food presentation, highlighting how homemade dishes can often mirror the aesthetic appeal of restaurant dishes, challenging the classifier to discern fine details beyond mere appearance.

In my Advanced Classifier project, I began with data preparation, collecting and standardizing images of homemade and non-homemade food. Following this, I extracted texture features using **Local Binary Patterns (LBP),** a crucial step for capturing the essence of each dish. I then defined the model using a **Random Forest Classifier,** valuing its robustness with complex, high-dimensional data and its insightful feature importance. After training the model on these features, I evaluated its performance, employing confusion matrices to gauge accuracy and fine-tune the classifier's ability to differentiate dishes based on their origin.

**Project Outcomes:**

Basic Model excels in chip identification but suggests dataset biases; Advanced Model accurately recognizes homemade dishes amidst data imbalances, needing rice identification refinement and showing potential in diverse dish presentation analysis.

**NOTE : MLEnd Yummy dataset**

The Yummy dataset, curated under the expert guidance of Professor Jesus Requena-Carrion at Queen Mary University of London, features a comprehensive collection of 3,250 food images, each meticulously annotated in a metadata CSV file with insights on food preferences and vegetarian categories, as per the MLEnd dataset documentation. This rich dataset, gathered by students including myself, serves as an excellent platform for developing skills in image preprocessing and feature extraction.

**Skills I learned:**

* **Machine Learning Pipeline**
* **Texture Feature exaction method (LBP- Local Binary Patterns )**
* **Random Forest Classifier**
* **Linear SVC (Support Vector Machine)**

**Q: How do I choose between LinearSVC and RandomForestClassifier for my classification tasks?**

**A: When deciding which model to use, consider these two key factors:**

* **Data Characteristics:**
  + **LinearSVC** is best suited for datasets with linear relationships. It’s efficient and fast, making it ideal for high-dimensional spaces where the number of features exceeds the number of samples.
  + **RandomForestClassifier** excels in datasets with non-linear relationships and complex patterns. It’s robust against overfitting and provides insights into feature importance, making it useful for more intricate datasets.
* **Computational Resources and Performance Needs:**
  + **LinearSVC** is less resource-intensive, which is advantageous in environments where computational resources are limited or where high speed is necessary for large datasets.
  + **RandomForestClassifier** requires more computational power but generally offers superior performance, especially in accuracy, making it suitable for handling larger and more diverse datasets.