NutriClass: Food Classification Using Nutritional Data

Project Report

1. Approach

The goal of this project is to classify food items using their nutritional values such as calories, protein, fat, carbs, sugar, sodium, etc. The dataset was cleaned, analyzed, and prepared through preprocessing. A new target variable (Food_Type) was created based on macronutrient dominance. Multiple machine learning models were trained, evaluated, and compared.

2. Data Analysis (EDA)

Basic observations showed skewed distributions in several numerical features. Outliers were present, which is normal in nutrition datasets. Correlation analysis revealed strong relationships such as Calories with Fat and Carbs. Visualizations included histograms, boxplots, correlation heatmaps, and category count plots.

3. Feature Engineering & Preprocessing

- One-hot encoding for categorical attributes
- Label encoding for the target variable
- StandardScaler applied to numerical features
- Food_Type categories created: High Protein, High Carb, High Fat, Balanced

4. Models Used

The following models were trained:

- Logistic Regression
- SVM
- KNN
- Decision Tree
- Random Forest
- XGBoost
- Gradient Boosting

5. Model Evaluation

Performance summary:

Logistic Regression: Train Accuracy 99.76%, Test Accuracy 99.76%

SVM: Train Accuracy 99.05%, Test Accuracy 98.97%

XGBoost: Train Accuracy 100%, Test Accuracy 98.91%

Random Forest: Train Accuracy 100%, Test Accuracy 98.17%

Gradient Boosting: Train Accuracy 99.07%, Test Accuracy 98.16%

Decision Tree: Train Accuracy 100%, Test Accuracy 97.85%

KNN: Train Accuracy 95.43%, Test Accuracy 92.56%

Logistic Regression showed the highest overall performance. Confusion matrices confirmed correct separation of food categories.

6. Key Insights

- Nutritional values can effectively classify food types.

- Macronutrient-based categorization aligns well with model predictions.

- Scaling significantly improves performance for linear models.

- Tree models handle non-linear relationships and outliers effectively.

7. Recommendations

- Prefer Logistic Regression or SVM for deployment.

- Add more features such as ingredients, cuisine type, or name embeddings for advanced classification.

- The model can support diet planning systems, menu categorization, or nutrition recommendation tools.

8. Conclusion

This project successfully developed a food classification system using nutritional data. With appropriate preprocessing and feature engineering, models achieved over 98–99% accuracy. This confirms the potential of nutrition-driven machine learning for food categorization and health-related applications.