Title:

Classification of Apple Quality Based on Physical and Sensory Features

Objective:

To classify apples as "good" or "bad" based on a set of physical and sensory attributes. The goal is to develop a predictive model that can accurately assess apple quality using features such as size, weight, sweetness, crunchiness, juiciness, ripeness, and acidity.

Data Analysis Summary:

1. **Data Exploration:**
    - Loaded the dataset containing various attributes of apples, including size, weight, sweetness, crunchiness, juiciness, ripeness, acidity, and overall quality.
    - Explored the dataset to understand its structure, including checking the number of rows, columns, and data types.
2. **Missing Values Check:**
    - There were no missing values in the dataset which was taken from kaggle
3. **Target Variable Balance:**
    - Assessed the balance of the target variable, "Quality," which indicates whether an apple is classified as "good" or "bad."
    - Found that the dataset was well-balanced, with an approximately equal number of "good" and "bad" apples.
4. **Feature Selection and Outlier Detection:**
    - Analyzed features to detect and handle outliers by calculating the Interquartile Range (IQR) and applying appropriate filtering methods.
    - Removed outliers to ensure that the model training was based on representative data.
5. **Data Scaling:**
    - Applied standardization using `StandardScaler` to normalize the feature set, enhancing model performance and convergence.
6. **Model Training and Evaluation:**
    - Trained various classification models, including K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Random Forest, Logistic Regression, XGBoost, and LightGBM.
    - Evaluated model performance using accuracy, confusion matrix, and classification report metrics.
    - The best models were KNN and LightGBM.
7. **Model Explainability with LIME:**
    - Applied LIME (Local Interpretable Model-agnostic Explanations) to interpret the KNN and LightGBM models.
    - Analyzed the impact of features on the prediction results, highlighting key features such as sweetness, ripeness, and juiciness for both models.

8. **Key Findings:**
    - Both KNN and LightGBM models predicted apple quality with high accuracy, influenced predominantly by sweetness and ripeness.
    - **Sweetness** was identified as the most significant factor in determining apple quality.
    - The LightGBM model also considered additional features such as acidity, while the KNN model included crunchiness.
9. **Conclusion:**
    - The analysis demonstrated the effectiveness of machine learning models in classifying apple quality based on physical and sensory features.
    - The findings provide insights into which attributes are most critical for determining apple quality, supporting better quality control and selection processes in the industry.