Documentation

**Documentation**

**Data Preprocessing Steps**

**Handling Delimiters**

**Step 1: Converting Commas to Semicolons in the Dataset** **Objective:** Ensure the dataset is properly formatted by replacing all semicolons (;) with commas (,).

**Step 2: Checking for Outliers, Missing Values, and Special Characters** **Objective:** Ensure the data quality by identifying and handling the following issues:

* **Outliers:** Perform statistical analysis (such as interquartile range or z-score) to detect extreme data points that may affect model performance.
* **Missing Values:** Identify missing entries and handle them through appropriate methods, such as imputation (mean, median, or mode) or row removal when necessary.
* **Special Characters:** Detect and clean special characters such as the apostrophe in entries like "Cote d'Ivoire" to avoid parsing errors. These can be replaced by suitable alternatives (e.g., removing the apostrophe or using escape sequences).

**Steps Taken:**

* Applied functions to identify and visualize outliers.
* Used conditional checks to replace or remove missing values.
* Implemented a find-and-replace mechanism for problematic characters.

**Algorithm Implementation**

A step-by-step guide to the algorithms used, including any parameter tuning.

**Evaluation Metrics**

Detailed explanation of how each model was evaluated.

**Results and Insights**

Visualizations, performance scores, and any key insights gained from applying the algorithms.