**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 03 February 2025 |
| Team ID | xxxxxx |
| Project Title | Implementing AgriPrediction using Machine Learning |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Summarizes the dataset, including the number of records, features, data types, and overall structure to understand the scope and composition of agricultural data. |
| Univariate Analysis | Examines individual variables such as soil pH, temperature, rainfall, and crop yield using measures like mean, median, mode, and distribution plots to identify trends and inconsistencies. |
| Bivariate Analysis | Analyzes relationships between two variables, such as the correlation between rainfall and yield or soil nutrients and crop health, using correlation coefficients, scatter plots, and heatmaps. |
| Multivariate Analysis | Explores interactions between multiple features, such as how temperature, humidity, and soil moisture collectively impact crop growth, using techniques like PCA and clustering. |
| Outliers and Anomalies | Detects and handles anomalies in the data using statistical methods (e.g., Z-score, IQR) and machine learning techniques to ensure robust and reliable predictions. |
| **Data Preprocessing Code Screenshots** | | |
| Loading Data | Code to load the dataset into the preferred environment (e.g., Python, R). |
| Handling Missing Data | Code for identifying and handling missing values. |
| Data Transformation | Code for transforming variables (scaling, normalization). |
| Feature Engineering | Code for creating new features or modifying existing ones. |
| Save Processed Data | Code to save the cleaned and processed data for future use. |