Data Collection and Preprocessing Phase

Date	18 July 2024
Team ID	SWTID1721319573
Project Title	Blueberry Yield Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	This project aims to develop a machine learning model to predict the yield of wild blueberries based on various factors such as weather conditions, soil quality, and pest pressure. Accurate yield predictions will help farmers optimize their resources, reduce waste, and improve profitability.
Data Collection Plan	Data has been collected from Kaggle, a well-known platform for datasets. The selected dataset includes comprehensive information necessary for building a robust predictive model.
Raw Data Sources Identified	The raw data source is publicly accessible and provides a diverse and comprehensive dataset for training the model.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Wild Blueberries Yield Data: Contains historical data on blueberry yields, weather conditions, soil quality, and pest pressure.	https://www.kaggle.com/datasets/saurabhshahane/wild-blueberry-yield-prediction	CSV	361 kB	Public