Data Collection and Preprocessing Phase

Date	6th June 2024
Team ID	SWTID1720109498
Project Title	Blueberry Yield Prediction
Maximum Marks	2 Marks

Data Collection Plan

Section	Description
Project Overview	This project focuses on developing an ML solution to predict blueberry yields accurately. It involves collecting and analyzing historical data on blueberry yields. Four different machine learning models were trained and evaluated, with linear regression identified as the most effective. The goal is to provide blueberry farmers with reliable predictions to optimize farming practices, enhance productivity, and support sustainable agriculture.
Data Collection Plan	Data will be collected from a vast platform on Internet named Kaggle.
Raw Data Sources Identified	The data has been collected over a period of 30 years from a wild blueberry plantation in Maine, which is situated in The United States of America.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The dataset being used to train the model in our project is **WildBlueberryPollinationSimulationData.csv** . The data here is experimental and it was collected in Maine, USA during the last 30 years.	https://www.kaggle.com/datasets/saurabhshahane/wild-blueberry-yield-prediction	CSV	85.26kB	Public