

Project Initialization and Planning Phase

Date	5 th June 2024
Team ID	SWTID1720109498
Project Title	Blueberry Yield Predictor
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	Implement a robust machine learning model to accurately predict blueberry yield based on historical data
Scope	This project aims to develop a ML solution to predict blueberry yields based on historical data and its scope includes data collection, feature selection and engineering, model development using regression or other suitable algorithms, model training and evaluation, and implementation of the predictive model for practical use by farmers. Continuous improvement and collaboration with stakeholders will ensure the model's accuracy and relevance in optimizing agricultural practices and maximizing blueberry productivity.
Problem Statement	
Description	Blueberry farmers struggle with unreliable yield predictions due to unpredictable weather, soil variations, and pest outbreaks. This project aims to develop a precise ML model using historical data accurately forecast blueberry yields.
Impact	This project will have a positive impact in a way that it will empower farmers with reliable insights, enabling them to make informed decisions regarding planting, resource allocation, and pest management. This capability has the potential to significantly enhance agricultural productivity, reduce uncertainty, and support sustainable farming practices, thereby improving overall profitability and environmental stewardship in blueberry cultivation.

Proposed Solution	
Approach	We utilized four different machine learning models, namely linear regression, random forest regressor, decision tree regressor and XGB regressor to train our predictive model for blueberry yield.
Key Features	Out of the four models used, we found that XGB regressor has the best outcome and we have proceeded with that to create a front end

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	GTX-1650 ti
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, seaborn, scipy, dabl
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset(777 rows)