Project Initialization and Planning Phase

Date	7 June 2024
Team ID	-
Project Title	Golden Harvesting: A Predictive Model For Apple Quality Assurance
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	Develop a machine learning model to predict apple quality based on factors like size, color, firmness, and sugar content, aiding farmers and distributors in making informed harvesting and storage decisions.
Scope	• Collect data from farms, agricultural databases, and studies. • Prepare and analyse data. • Build, evaluate, and tune multiple ML models. • Deploy the best model in a web framework for practical use.
Problem Statement	
Description	Address inconsistent apple quality due to varying growth conditions, which leads to financial losses and affects consumer satisfaction.
Impact	Accurate apple quality predictions can reduce waste, improve profitability, and ensure higher quality produce for consumers.
Proposed Solution	
Approach	• Data Collection and Preparation: Gather data on apple quality from various sources, preprocess the data for analysis. • Exploratory Data Analysis: Perform descriptive statistics and visual analysis to uncover patterns and insights.

	• Model Building: Train multiple machine learning models (decision trees, random forests, XGBoost, logistic regression, SVM, k-NN, naive Bayes) on the prepared data. • Hyperparameter Tuning: Optimize the models' performance by fine-tuning their hyperparameters. • Model Evaluation: Compare models using performance metrics to identify the best-performing model. • Model Deployment: Deploy the selected model by integrating it into a web framework for practical use by farmers and distributors.
Key Features	• Comprehensive data analysis. • Evaluation of multiple ML models. • Optimization through hyperparameter tuning. • User-friendly web integration for easy access.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	512GB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	numpy, pandas, matplotlib, seaborn, scikit-learn, pickle
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle , 174 kB, CSV