Model Development Phase Template

Date	13th June 2024
Team ID	SWTID1720109498
Project Title	Blueberry Yield Predictor
Maximum Marks	5 Marks

Feature Selection Report

This report outlines the process and rationale behind the selection and removal of features for the Wild Blueberry Pollination Simulation dataset. The goal of feature selection is to improve model performance and interpretability by focusing on the most impactful variables. Through data visualization and analysis, we identified key features that significantly contribute to predicting the target outcomes while removing redundant or non-informative variables. The following sections detail the reasons for selecting or removing each feature, ensuring a robust and efficient model.

Feature	Description	Selected (Yes/No)	Reasoning
Row#	Unique identifier for each row of the dataset	No	For predicting the yield value, Row number is not required.
clonesize	The size of the blueberry clone being studied.	Yes	Influences yield and fruit production due to varying pollination dynamics in different clone sizes.
honeybee	The density or activity level of honeybee pollinators.	Yes	Key pollinator affecting yield and seed production.
bumbles	The density or activity level of bumblebee pollinators.	No	Likely redundant with other bee density variables or showed little correlation with the target variable.

andrena	The density or activity level of Andrena bee pollinators.	Yes	Significant pollinator, adding diversity in understanding pollination effects.
osmia	The density or activity level of Osmia bee pollinators.	Yes	Important pollinator, contributing to overall pollination success.
MaxOfUpperTRange	The maximum temperature of the upper temperature range recorded.	No	Little variation or weak correlation with outcomes; redundant with other temperature measures.
MinOfUpperTRange	The minimum temperature of the upper temperature range recorded.	Yes	Influences pollinator activity and plant growth under extreme temperatures.
AverageOfUpperTRange	The average temperature of the upper temperature range.	Yes	Impacts pollinator behavior and plant physiology under consistent temperatures.
MaxOfLowerTRange	The maximum temperature of the lower temperature range recorded.	No	Limited variation or significance; possibly redundant with other temperature features.
MinOfLowerTRange	The minimum temperature of the lower temperature range recorded.	No	Showed little correlation with outcomes; redundancy with other temperature variables.
AverageOfLowerTRange	The average temperature of the lower temperature range.	Yes	Important for understanding cooler conditions affecting plant stress and pollination.

RainingDays	The total number of raining days.	Yes	Affects pollinator activity and plant health; significant for crop success.
AverageRainingDays	The average number of raining days.	No	Non-informative or redundant with total raining days (RainingDays).
fruitset	The proportion of flowers that set fruit.	No	Redundant with other yield-related measures.
fruitmass	The mass of the fruit produced.	No	Highly correlated with yield, making it redundant.
seeds	The number of seeds per fruit.	Yes	Direct measure of successful pollination and fruit development.