Model Optimization and Tuning Phase

Date	12 July 2024
Team ID	SWTID1720077079
Project Title	Wild Blueberry Yield Prediction
Maximum Marks	10 Marks

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
| XGBoost - Grid Search Optimized | ```python
param_grid = {
    'n_estimators': [100, 200, 300],
    'learning_rate': [0.01, 0.1, 0.2],
    'max_depth': [3, 5, 7],
    'min_child_weight': [1, 3, 5],
    'subsample': [0.8, 1.0],
    'colsample_bytree': [0.8, 1.0]
}
``` | 'colsample_bytree': 0.8, 'learning_rate': 0.2, 'max_depth': 3, 'min_child_weight': 3, 'n_estimators': 300, 'subsample': 0.8 |
| XGBoost - Random Search Optimized | ```python
param_dist = {
    'n_estimators': [int(x) for x in np.linspace(start=100, stop=500, num=10)],
    'learning_rate': [0.001, 0.01, 0.1, 0.2],
    'max_depth': [int(x) for x in np.linspace(3, 10, num=8)],
    'min_child_weight': [1, 3, 5, 7],
    'subsample': [0.6, 0.8, 1.0],
    'colsample_bytree': [0.6, 0.8, 1.0]
}
``` | 'subsample': 1.0, 'n_estimators': 411, 'min_child_weight': 1, 'max_depth': 3, 'learning_rate': 0.2, 'colsample_bytree': 1.0 |

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
XGBoost - Grid Search Optimized	Mean Absolute Error: 180.54001388799836 Root Mean Squared Error: 260.16663946930686 R2: 96.60866093301176	Mean Absolute Error: 168.5955481947737 Root Mean Squared Error: 250.6726388510296 R2: 96.93485057578692
XGBoost - Random Search Optimized	Mean Absolute Error: 180.54001388799836 Root Mean Squared Error: 260.16663946930686 R2: 96.60866093301176	Mean Absolute Error: 164.7842086830549 Root Mean Squared Error: 236.57209070078258 R2: 97.24612545035339

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
XGBoost - Random Search Optimized	It was chosen because of: 1. Low Mean Absolute Error 2. Low Root Mean Squared Error 3. High R2 Score on testing set.