Model Optimization and Tuning Phase Template

Date	18 July 2024
Team ID	-
Project Title	Golden Harvesting: A Predictive Model For Apple Quality Assurance
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values

XGBoost	`param_grid = { 'n_estimators': [100, 200, 300], 'learning_rate': [0.01, 0.1, 0.2], 'max_depth': [3, 4, 5], 'min_child_weight': [1, 5, 10], 'subsample': [0.8, 0.9, 1.0], 'colsample_bytree': [0.8, 0.9, 1.0] }`	`best_model = grid_search.best_estimator_`
		`xgb_pred = best_model.predict(x_test)`
		`acc_score = accuracy_score(y_test, xgb_pred)`
		`print("Accuracy of XGBoost(best parameters) {:.2f}%" .format(acc_score * 100))`
		`print("Confusion matrix of XGBoost is as follows:")`
		`print(confusion_matrix(y_test, xgb_pred))`

Random forest	`param_grid = { 'n_estimators': [100, 200, 300], 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4], 'bootstrap': [True, False] }`	`best_rf_model = grid_search.best_estimator_`
		`rf_pred = best_rf_model.predict(x_test)`
		`acc_score = accuracy_score(y_test, rf_pred)`
		`print("Accuracy of Random Forest(best parameters) {:.2f}%" .format(acc_score * 100))`
		`print("Confusion matrix of Random Forest is as follows:")`
		`print(confusion_matrix(y_test, rf_pred))`

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
XGBoost	Accuracy of XGBoost is 0.91 Confusion matrix of XGBoost array([[363, 38], [37, 362]], dtype=int64)	Accuracy of XGBoost(best parameters) 90.50% Confusion matrix of XGBoost is as follows: [[359 42] [34 365]]
Random forest	Accuracy of Random forest is 0.91 Confusion matrix of Random forest array([[357, 44], [29, 370]], dtype=int64)	Accuracy of Random Forest(best parameters) 90.88% Confusion matrix of Random Forest is as follows: [[359 42] [31 368]]

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	Random Forest was chosen as the final optimized model because it demonstrated the highest accuracy of 90.88%. The confusion matrix of Random Forest also showed fewer misclassifications, with lower False Positive and False Negative counts, indicating better overall performance in predicting the correct classes. Additionally, Random Forest is known for its robustness and efficiency, especially with structured data, making it a more suitable and reliable choice for predicting apple quality based on the given dataset.