Model Optimization and Tuning Phase

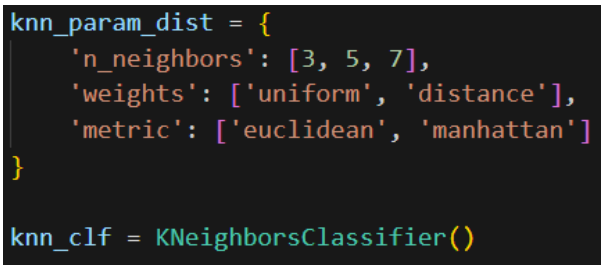
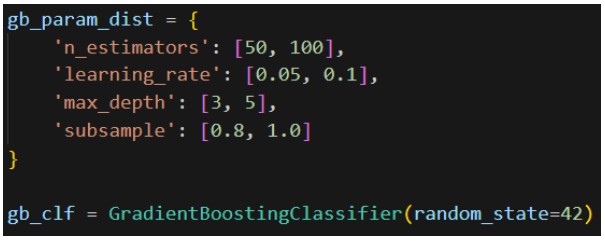
Date	25 June 2025
Team ID	SWTID1749710444
Project Title	Online Payment Fraud Detection using ML
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

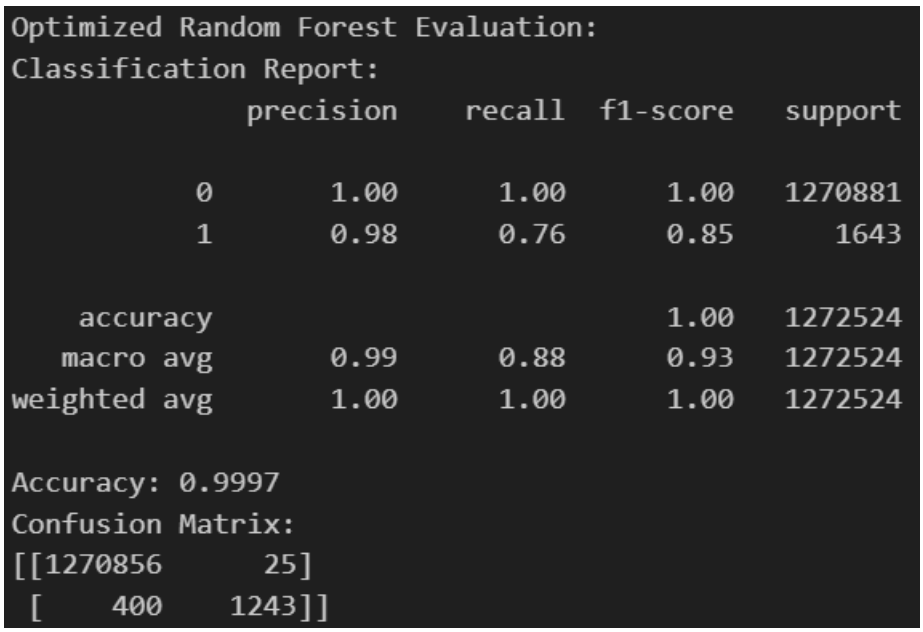
The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Random Forest	`param_dist = { 'n_estimators': [50, 100], 'max_depth': [None, 10], 'class_weight': ['balanced'] }` `rf_clf = RandomForestClassifier(random_state=42)`	`Best Parameters: {'n_estimators': 50, 'max_depth': None, 'class_weight': 'balanced'}` `Best Cross-Validation Score: 0.8434`
Decision Tree	`dt_param_dist = { 'max_depth': [None, 10, 15], 'min_samples_split': [2, 10], 'min_samples_leaf': [1, 5], 'criterion': ['gini', 'entropy'], 'class_weight': ['balanced', {0: 1, 1: 10}] }` `dt_clf = DecisionTreeClassifier(random_state=42)`	`Best Parameters: {'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None, 'criterion': 'gini', 'class_weight': {0: 1, 1: 10}}` `Best Cross-Validation Score: 0.8260`

KNN		
Gradient Boosting		

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Random Forest	

Decision Tree	Optimized Decision Tree Evaluation: Classification Report: precision recall f1-score support 0 1.00 1.00 1.00 1270881 1 0.87 0.83 0.85 1643 accuracy 1.00 1272524 macro avg 0.93 0.91 0.92 1272524 weighted avg 1.00 1.00 1.00 1272524 Accuracy: 0.9996 Confusion Matrix: [[1270669 212] [284 1359]]
KNN	Optimized KNN Evaluation: Classification Report: precision recall f1-score support 0 1.00 1.00 1.00 1270881 1 0.96 0.60 0.74 1643 accuracy 1.00 1272524 macro avg 0.98 0.80 0.87 1272524 weighted avg 1.00 1.00 1.00 1272524 Accuracy: 0.9995 Confusion Matrix: [[1270838 43] [651 992]]

Gradient Boosting	Optimized Gradient Boosting Evaluation: Classification Report: precision recall f1-score support 0 1.00 1.00 1.00 1270881 1 0.84 0.61 0.71 1643 accuracy 1.00 1272524 macro avg 0.92 0.81 0.85 1272524 weighted avg 1.00 1.00 1.00 1272524 Accuracy: 0.9993 Confusion Matrix: [[1270684 197] [633 1010]]

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	Random Forest emerged as the recommended model. It provided the best balance between high fraud precision (minimizing false positives) and reasonable recall (catching most frauds), which is crucial for business operations. Gradient Boosting and Decision Tree models either produced excessive false positives or missed too many frauds, while KNN struggled with the high dimensionality and imbalance of the data.