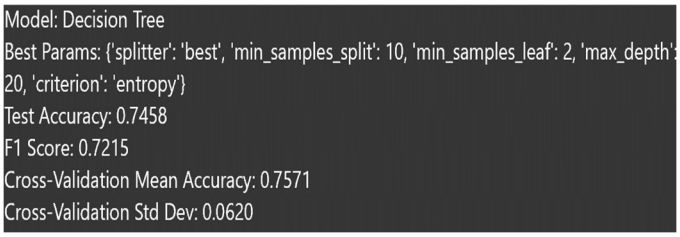
Model Optimization and Tuning Phase Report

Date	19 June 2025
Team ID	SWTID1750050475
Project Title	SmartLender - Applicant Credibility Prediction for Loan Approval
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
Decision Tree		

Random Forest	`'Random Forest': { 'model': RandomForestClassifier(), 'params': { 'n_estimators': [50, 100, 200], 'criterion': ['gini', 'entropy'], 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] }`	Model: Random Forest Best Params: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 30, 'criterion': 'entropy'} Test Accuracy: 0.8333 F1 Score: 0.8374 Cross-Validation Mean Accuracy: 0.8357 Cross-Validation Std Dev: 0.0417

KNN	`'KNN': { 'model': KNeighborsClassifier(), 'params': { 'n_neighbors': [3, 5, 7, 9], 'weights': ['uniform', 'distance'], 'p': [1, 2] # p=1: Manhattan, p=2: Euclidean } },`	Model: KNN Best Params: {'weights': 'distance', 'p': 1, 'n_neighbors': 5} Test Accuracy: 0.7958 F1 Score: 0.7860 Cross-Validation Mean Accuracy: 0.8286 Cross-Validation Std Dev: 0.0249
Gradient Boosting	`'Gradient Boosting': { 'model': GradientBoostingClassifier(), 'params': { 'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.2], 'max_depth': [3, 4, 5], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4], 'subsample': [0.8, 1.0] }`	Model: Gradient Boosting Best Params: {'subsample': 1.0, 'n_estimators': 100, 'min_samples_split': 10, 'min_samples_leaf': 2, 'max_depth': 5, 'learning_rate': 0.2} Test Accuracy: 0.8167 F1 Score: 0.8167 Cross-Validation Mean Accuracy: 0.8268 Cross-Validation Std Dev: 0.0394

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Decision Tree	Best Params: {'splitter': 'best', 'min_samples_split': 10, 'min_samples_leaf': 2, 'max_depth': 20, 'criterion': 'entropy'} Test Accuracy: 0.7458 F1 Score: 0.7215 Cross-Validation Mean Accuracy: 0.7571 Cross-Validation Std Dev: 0.0620 Confusion Matrix: [[100 29] [32 79]] Classification Report: precision recall f1-score support 0 0.76 0.78 0.77 129 1 0.73 0.71 0.72 111 accuracy 0.75 240 macro avg 0.74 0.74 0.74 240 weighted avg 0.75 0.75 0.75 240

Random Forest	Best Params: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 30, 'criterion': 'entropy'} Test Accuracy: 0.8333 F1 Score: 0.8374 Cross-Validation Mean Accuracy: 0.8357 Cross-Validation Std Dev: 0.0417 Confusion Matrix: [[97 32] [8 103]] Classification Report: precision recall f1-score support 0 0.92 0.75 0.83 129 1 0.76 0.93 0.84 111 accuracy 0.83 240 macro avg 0.84 0.84 0.83 240 weighted avg 0.85 0.83 0.83 240
KNN	Best Params: {'weights': 'distance', 'p': 1, 'n_neighbors': 5} Test Accuracy: 0.7958 F1 Score: 0.7860 Cross-Validation Mean Accuracy: 0.8286 Cross-Validation Std Dev: 0.0249 Confusion Matrix: [[101 28] [21 90]] Classification Report: precision recall f1-score support 0 0.83 0.78 0.80 129 1 0.76 0.81 0.79 111 accuracy 0.80 240 macro avg 0.80 0.80 0.80 240 weighted avg 0.80 0.80 0.80 240

Gradient Boosting	Best Params: {'subsample': 1.0, 'n_estimators': 100, 'min_samples_split': 10, 'min_samples_leaf': 2, 'max_depth': 5, 'learning_rate': 0.2} Test Accuracy: 0.8167 F1 Score: 0.8167 Cross-Validation Mean Accuracy: 0.8268 Cross-Validation Std Dev: 0.0394 Confusion Matrix: [[98 31] [13 98]] Classification Report: precision recall f1-score support 0 0.88 0.76 0.82 129 1 0.76 0.88 0.82 111 accuracy 0.82 240 macro avg 0.82 0.82 0.82 240 weighted avg 0.83 0.82 0.82 240

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
RANDOM FOREST CLASSIFIER	The Random Forest model was selected for Smart Lender due to its superior performance and practical advantages. It delivered the highest F1 Score and test accuracy, indicating strong predictive capability. Compared to other models, Random Forest is: Less prone to overfitting than a single Decision Tree Robust with mixed data types (categorical + numerical) Efficient in real-time inference, unlike Gradient Boosting Interpretable, offering feature importance insights Its consistent results in cross-validation and ability to handle real-world variability make it ideal for our loan approval prediction system.