Model Optimization and Tuning Phase Report

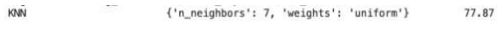
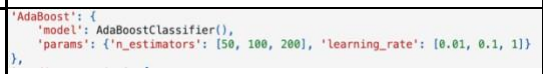
Date	20 June 2025
Team ID	SWTID1751617613
Project Title	Auto Insurance Fraud
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

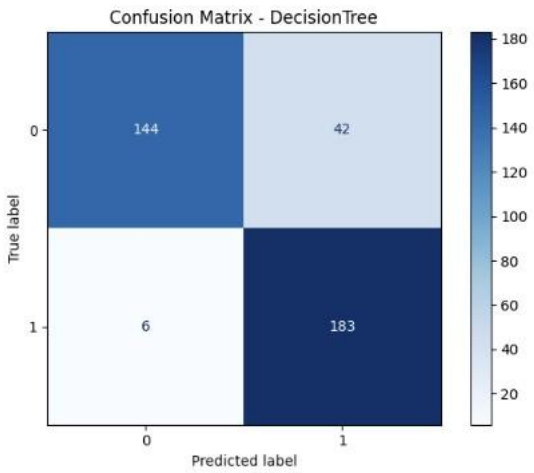
The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

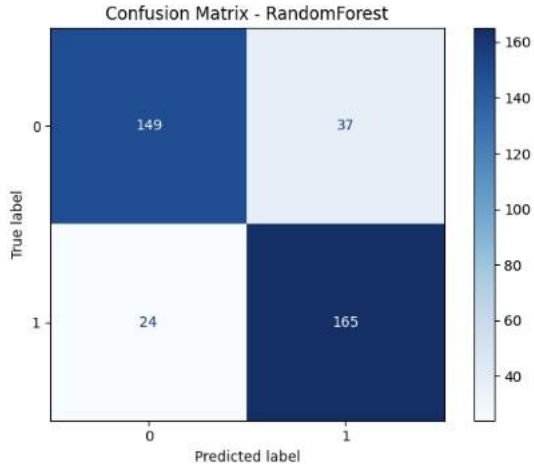
Model	Tuned Hyperparameters	Optimal Values
Decision Tree	`'DecisionTree': { 'model': DecisionTreeClassifier(), 'params': {'max_depth': [3, 5, 10], 'min_samples_split': [2, 5, 10]} },`	
Random Forest	`'RandomForest': { 'model': RandomForestClassifier(), 'params': {'n_estimators': [100, 200], 'max_depth': [5, 10], 'max_features': ['sqrt']} },`	`RandomForest {'max_depth': 10, 'max_features': 'sqrt', 'n_estimators': 100} 83.73`

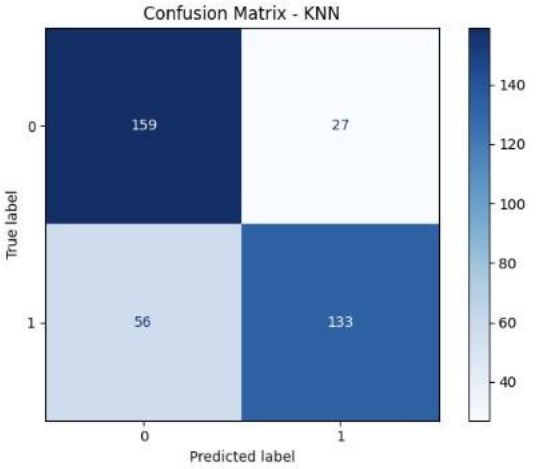
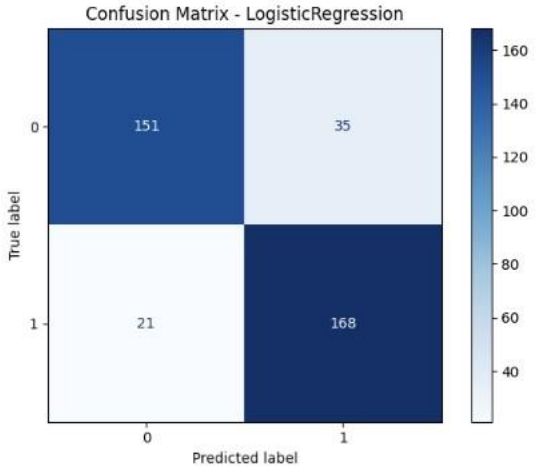
KNN		
Logistic Regression		
Gaussian NB		
SVC		
AdaBoost		
Gradient Boosting		
XG Boost		

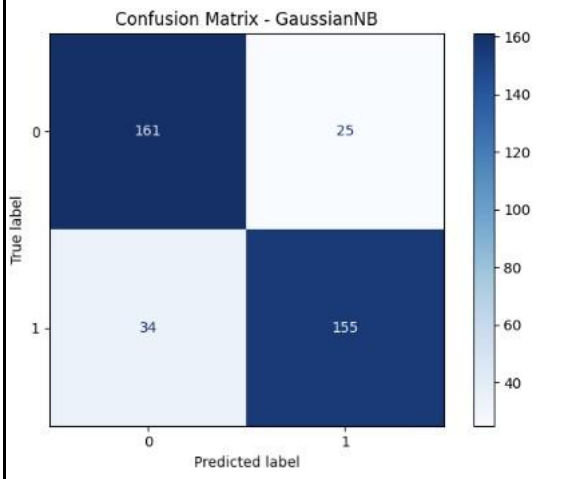
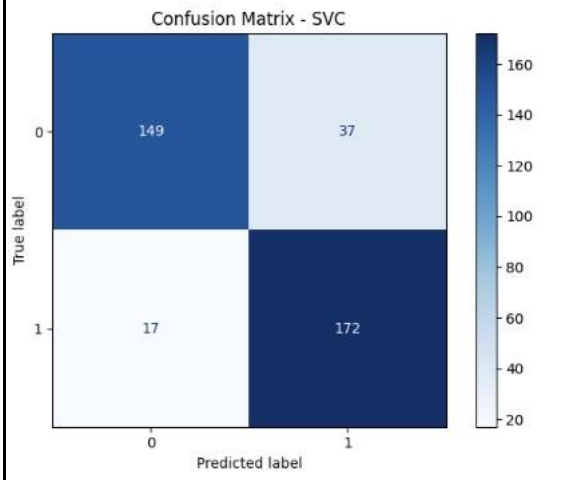
Model	Optimized Metric

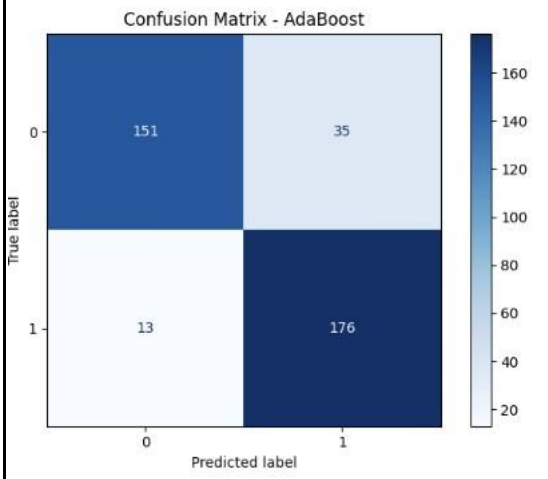
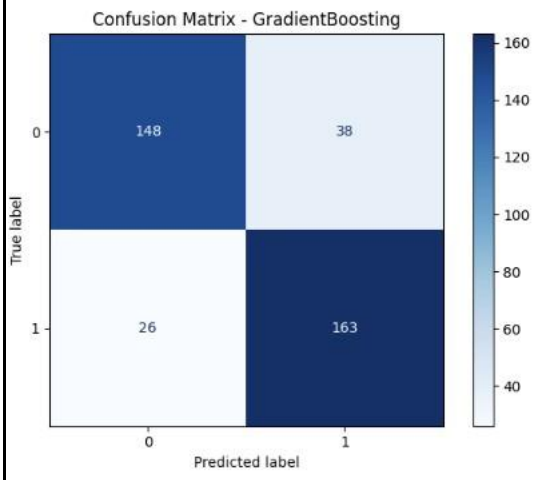
Decision Tree	Classification Report for DecisionTree: precision recall f1-score support 0 0.96 0.77 0.86 186 1 0.81 0.97 0.88 189 accuracy 0.87 375 macro avg 0.89 0.87 0.87 375 weighted avg 0.89 0.87 0.87 375 

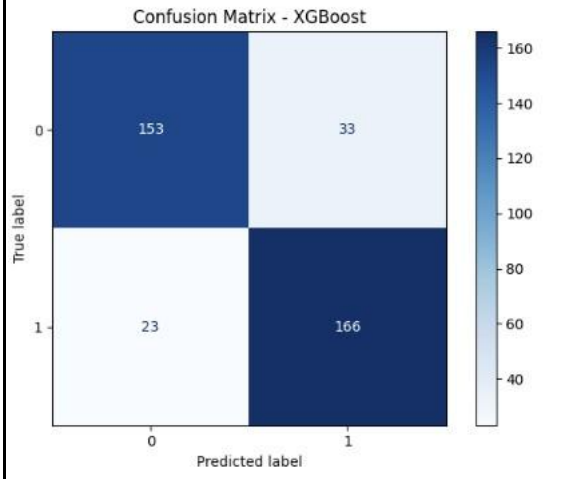
Performance Metrics Comparison Report (2 Marks):

Random Forest	Classification Report for RandomForest: precision recall f1-score support 0 0.86 0.80 0.83 186 1 0.82 0.87 0.84 189 accuracy 0.84 375 macro avg 0.84 0.84 0.84 375 weighted avg 0.84 0.84 0.84 375 

KNN	Classification Report for KNN: precision recall f1-score support 0 0.74 0.85 0.79 186 1 0.83 0.70 0.76 189 accuracy 0.78 375 macro avg 0.79 0.78 0.78 375 weighted avg 0.79 0.78 0.78 375 
Logistic Regression	Classification Report for LogisticRegression: precision recall f1-score support 0 0.88 0.81 0.84 186 1 0.83 0.89 0.86 189 accuracy 0.85 375 macro avg 0.85 0.85 0.85 375 weighted avg 0.85 0.85 0.85 375 

Gaussian NB	Classification Report for GaussianNB: precision recall f1-score support 0 0.83 0.87 0.85 186 1 0.86 0.82 0.84 189 accuracy 0.84 375 macro avg 0.84 0.84 0.84 375 weighted avg 0.84 0.84 0.84 375 
SVC	Classification Report for SVC: precision recall f1-score support 0 0.90 0.80 0.85 186 1 0.82 0.91 0.86 189 accuracy 0.86 375 macro avg 0.86 0.86 0.86 375 weighted avg 0.86 0.86 0.86 375 

Ada Boost	Classification Report for AdaBoost: precision recall f1-score support 0 0.92 0.81 0.86 186 1 0.83 0.93 0.88 189 accuracy 0.87 375 macro avg 0.88 0.87 0.87 375 weighted avg 0.88 0.87 0.87 375 
Gradient Boosting	Classification Report for GradientBoosting: precision recall f1-score support 0 0.85 0.80 0.82 186 1 0.81 0.86 0.84 189 accuracy 0.83 375 macro avg 0.83 0.83 0.83 375 weighted avg 0.83 0.83 0.83 375 

XG Boost	Classification Report for XGBoost: precision recall f1-score support 0 0.87 0.82 0.85 186 1 0.83 0.88 0.86 189 accuracy 0.85 375 macro avg 0.85 0.85 0.85 375 weighted avg 0.85 0.85 0.85 375 

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Decision Tree	The Decision Tree classifier is a strong choice for this fraud detection task due to its balanced and interpretable performance. It achieved the highest F1 score (0.5926) among all models tested, indicating a solid balance between precision and recall — both critical metrics in detecting fraud. With a recall of 0.6531, it successfully identifies a significant portion of fraudulent cases, helping minimize the risk of missed frauds. Additionally, Decision Trees are inherently easy to understand and visualize, making them highly suitable in domains like insurance and finance where model transparency and explainability are essential. Although it shows signs of overfitting (perfect training accuracy), its relatively high test accuracy (0.78) suggests it still generalizes well. Overall, the Decision Tree offers a practical trade-off between performance, interpretability, and implementation simplicity.