

Model Optimization and Tuning Phase

Date	19 Feb 2026
Team ID	LTVIP2026TMIDS80731
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	<pre>param_dist = { 'n_estimators': [50, 100], 'max_depth': [None, 10], 'class_weight': ['balanced'] } rf_clf = RandomForestClassifier(random_state=42)</pre>	<pre>Best Parameters: {'n_estimators': 50, 'max_depth': None, 'class_weight': 'balanced'} Best Cross-Validation Score: 0.8434</pre>
Decision Tree	<pre>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)</pre>	<pre>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: 0.8268</pre>

KNN	<pre>knn_param_dist = { 'n_neighbors': [3, 5, 7], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan'] } knn_clf = KNeighborsClassifier()</pre>	<pre>Best Parameters: {'weights': 'distance', 'n_neighbors': 5, 'metric': 'manhattan'} Best Cross-Validation Score: 0.7024</pre>
Gradient Boosting	<pre>gb_param_dist = { 'n_estimators': [50, 100], 'learning_rate': [0.05, 0.1], 'max_depth': [3, 5], 'subsample': [0.8, 1.0] } gb_clf = GradientBoostingClassifier(random_state=42)</pre>	<pre>Best Parameters: {'subsample': 1.0, 'n_estimators': 50, 'max_depth': 3, 'learning_rate': 0.05} Best Cross-Validation Score: 0.6604</pre>

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Random Forest	<pre>Optimized Random Forest Evaluation: Classification Report: precision recall f1-score support 0 1.00 1.00 1.00 1270881 1 0.98 0.76 0.85 1643 accuracy 0.99 macro avg 0.99 weighted avg 1.00 Accuracy: 0.9997 Confusion Matrix: [[1270856 25] [400 1243]]</pre>

Decision Tree	<div>Optimized Decision Tree Evaluation: Classification Report:<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>1.00</td><td>1.00</td><td>1.00</td><td>1270881</td></tr><tr><td>1</td><td>0.87</td><td>0.83</td><td>0.85</td><td>1643</td></tr><tr><td>accuracy</td><td></td><td></td><td>1.00</td><td>1272524</td></tr><tr><td>macro avg</td><td>0.93</td><td>0.91</td><td>0.92</td><td>1272524</td></tr><tr><td>weighted avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>1272524</td></tr></tbody></table> Accuracy: 0.9996 Confusion Matrix: [[1270669212] [2841359]]</div>		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
	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																											
KNN	<div>Optimized KNN Evaluation: Classification Report:<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>1.00</td><td>1.00</td><td>1.00</td><td>1270881</td></tr><tr><td>1</td><td>0.96</td><td>0.60</td><td>0.74</td><td>1643</td></tr><tr><td>accuracy</td><td></td><td></td><td>1.00</td><td>1272524</td></tr><tr><td>macro avg</td><td>0.98</td><td>0.80</td><td>0.87</td><td>1272524</td></tr><tr><td>weighted avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>1272524</td></tr></tbody></table> Accuracy: 0.9995 Confusion Matrix: [[127083843] [651992]]</div>		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
	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																											

Gradient Boosting	<pre> 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]] </pre>
-------------------	--

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	<p>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.</p>