

Model Optimization and Tuning Phase

Date	28 July 2025
Project Title	Flight Delays Prediction Using Machine Learning
Maximum Marks	10 Marks

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Decision Tree	<ul style="list-style-type: none"> - Criterion: ['gini', 'entropy'] - Splitter: ['best', 'random'] - Max Depth: [10, 20, 30, 40] - Min Samples Split: [2, 5, 10] - Min Samples Leaf: [1, 2, 4] 	Criterion: gini Max Depth: 20 Min Samples Split: 2 Min Samples Leaf: 1
Random Forest	<ul style="list-style-type: none"> - n_estimators: [50, 100, 200] - Criterion: ['gini', 'entropy'] - Max Depth: [10, 20, 30] - Min Samples Split: [2, 5] - Min Samples Leaf: [1, 2] 	n_estimators: 100 Criterion: gini Max Depth: 20 Min Samples Split: 2
SVM	<ul style="list-style-type: none"> - Kernel: ['linear', 'rbf'] - C: [0.1, 1, 10] - Gamma: ['scale', 'auto'] 	Kernel: linear C: 1.0 Gamma: scale
LightGBM	<ul style="list-style-type: none"> - n_estimators: [50, 100, 200] - Learning Rate: [0.05, 0.1, 0.2] - Num Leaves: [20, 31, 40] 	n_estimators: 100 Learning Rate: 0.1 Num Leaves: 31

Performance Metrics Comparison Report (2 Marks):

Model	Metrics
Decision Tree	<pre> Classification Report: precision recall f1-score support False 0.99 0.99 0.99 1802 True 0.97 0.97 0.97 445 accuracy 0.98 macro avg 0.98 weighted avg 0.99 Confusion Matrix: [[1790 12] [12 433]] </pre>
Random Forest	<pre> Classification Report: precision recall f1-score support False 0.90 0.99 0.94 1802 True 0.93 0.58 0.71 445 accuracy 0.91 macro avg 0.92 weighted avg 0.91 Confusion Matrix: [[1782 20] [188 257]] </pre>
SVM	<pre> Classification Report: precision recall f1-score support False 0.80 1.00 0.89 1802 True 0.00 0.00 0.00 445 accuracy 0.40 macro avg 0.40 weighted avg 0.64 Confusion Matrix: [[1802 0] [445 0]] </pre>

LightGBM	<pre> Classification Report: precision recall f1-score support False 0.98 0.99 0.98 1802 True 0.95 0.93 0.94 445 accuracy 0.98 2247 macro avg 0.97 0.96 0.96 2247 weighted avg 0.98 0.98 0.98 2247 Confusion Matrix: [[1780 22] [33 412]] </pre>
----------	---

Final Model Selection Justification Report (2 Marks):

Final Model	Justification
Decision Tree	<p>Highest Performance: The Decision Tree achieved the highest accuracy (98.9%) and F1-score (0.97) among all tested models (Random Forest, SVM, and LightGBM).</p> <p>Dataset Size and Simplicity: With a dataset of 11,231 rows, a Decision Tree is highly efficient, interpretable, and sufficient for capturing the key decision patterns without overfitting.</p> <p>Low Complexity & Fast Predictions: Unlike SVM or LightGBM, the Decision Tree is computationally lightweight, making it ideal for real-time predictions in the Flask web app.</p> <p>Interpretability: Decision Trees provide clear decision rules,</p>

	<p>which makes it easy to explain the reasoning behind predictions, useful for operational deployment in airports/airlines.</p> <p>No Hyperparameter Dependency: The model performed extremely well even with default hyperparameters, indicating that the dataset features are naturally separable for this prediction task.</p>
--	---