

## Model Optimization and Tuning Phase Template

Date	24 April 2024
Team ID	739739
Project Title	RESERVATION CANCELLATION PREDICTION
Maximum Marks	10 Marks

### Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining neural network 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 (8 Marks):

Model	Tuned Hyperparameters

Model	Tuned Hyperparameters																														
Random Forest	<p>The parameter grid (rfc_param_grid) for hyperparameter tuning. It specifies different values for the number of trees (n_estimators), maximum depth of trees (max_depth), and minimum number of features considered for splitting (min_samples_split).and minimum number of leafs(min_samples_leaf). GridSearchCV is employed with 5-fold cross-validation (cv=5), evaluating model performance based on accuracy (scoring="accuracy").</p> <pre># Hyperparameter tuning using GridSearchCV param_grid = {     'n_estimators': [100, 200, 300],     'max_depth': [None, 10, 20, 30],     'min_samples_split': [2, 5, 10],     'min_samples_leaf': [1, 2, 4] }  grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=5, n_jobs=-1, verbose=2)</pre> <p>🔄 Fitting 5 folds for each of 108 candidates, totalling 540 fits Best Parameters: {'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200} Accuracy Score: 0.8629655657062544 Confusion Matrix: [[772 61]  [134 456]] 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>0.85</td><td>0.93</td><td>0.89</td><td>833</td></tr><tr><td>1</td><td>0.88</td><td>0.77</td><td>0.82</td><td>590</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.86</td><td>1423</td></tr><tr><td>macro avg</td><td>0.87</td><td>0.85</td><td>0.86</td><td>1423</td></tr><tr><td>weighted avg</td><td>0.86</td><td>0.86</td><td>0.86</td><td>1423</td></tr></tbody></table></p>		precision	recall	f1-score	support	0	0.85	0.93	0.89	833	1	0.88	0.77	0.82	590	accuracy			0.86	1423	macro avg	0.87	0.85	0.86	1423	weighted avg	0.86	0.86	0.86	1423
	precision	recall	f1-score	support																											
0	0.85	0.93	0.89	833																											
1	0.88	0.77	0.82	590																											
accuracy			0.86	1423																											
macro avg	0.87	0.85	0.86	1423																											
weighted avg	0.86	0.86	0.86	1423																											

Model	Tuned Hyperparameters
Decision Tree	<p>The parameters (params) define a grid for hyperparameter tuning of the Decision Tree Classifier (DecisionTreeClassifier), including max_depth, min_samples_leaf, and criterion ('gini' or 'entropy'). GridSearchCV (dt_model) is used with 5-fold cross-validation (cv=5), evaluating model performance based on accuracy (scoring="accuracy")</p> <pre># Hyperparameter tuning using GridSearchCV param_grid = {     'criterion': ['gini', 'entropy'],     'splitter': ['best', 'random'],     'max_depth': [None, 10, 20, 30],     'min_samples_split': [2, 5, 10],     'min_samples_leaf': [1, 2, 4] }  grid_search = GridSearchCV(estimator=dt_model, param_grid=param_grid, cv=5, n_jobs=-1, verbose=2)</pre> <pre>Fitting 5 folds for each of 144 candidates, totalling 720 fits Validation ROC AUC Score for Decision Tree: 0.9182462378935301 Best Parameters: {'criterion': 'entropy', 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5, 'splitter': 'best'} Accuracy Score: 0.86742006615215 Confusion Matrix: [[2222  213]  [ 268  925]] Classification Report:       precision    recall  f1-score   support        0       0.89       0.91       0.90       2435       1       0.81       0.78       0.79       1193     accuracy       0.87   macro avg       0.85 weighted avg       0.87</pre> <p>Test Predictions: [0. 0. 0.16058394 ... 1. 0.18963415 0.97826087]</p>

**Final Model Selection Justification (2 Marks):**

Final Model	Reasoning
Random Forest	<p>Random Forest model is chosen for its robustness in handling complex datasets and its ability to mitigate overfitting while providing high predictive accuracy.</p> <pre> Fitting 5 folds for each of 108 candidates, totalling 540 fits Best Parameters: {'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200} Accuracy Score: 0.8629655657062544 Confusion Matrix: [[772  61]  [134 456]] Classification Report:               precision    recall  f1-score   support        0       0.85         0.93         0.89         833       1       0.88         0.77         0.82         590   accuracy          0.86         0.86         0.86         1423  macro avg          0.87         0.85         0.86         1423  weighted avg       0.86         0.86         0.86         1423 </pre> <p>Above two models Random Forest model have the highest accuracy among the models.</p>