Model Optimization and Tuning Phase Template

Date	24 April 2024
Team ID	740663
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

Random Forest	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").
	
	`# 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)`
	
	`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 1423`
	`macro avg 0.87 0.85 0.86 1423`
	`weighted avg 0.86 0.86 0.86 1423`

<table>
<tr>
<td>Decision Tree</td>
<td>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")

```
# 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)
```

```
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      3628
   macro avg       0.85      0.84      0.85      3628
weighted avg       0.87      0.87      0.87      3628

Test Predictions: [0.         0.         0.16058394 ... 1.         0.18963415 0.97826087]
```
</td>
</tr>
</table>

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	Random Forest model is chosen for its robustness in handling complex datasets and its ability to mitigate overfitting while providing high predictive accuracy. 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 1423 macro avg 0.87 0.85 0.86 1423 weighted avg 0.86 0.86 0.86 1423 Above two models Random Forest model have the highest accuracy among the models.