Model Development Phase Template

Date	24 April 2024
Team ID	739942
Project Title	Freedom of the World Classification
Maximum Marks	10 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include a summary and training and validation performance metrics for multiple models, presented through respective screenshots.

Initial Model Training Code (5 marks):

Paste the screenshot of the model training code

Model Validation and Evaluation Report (5 marks):

Model	Summary	Training and Validation Performance Metrics
Model 1	Random forest classifier model often encompass accuracy, precision, recall, F1 score to measure its prediction quality and robustness.	
Model 2	Decision tree classifier model commonly include accuracy, precision, recall, F1 score which help assess the model's prediction accuracy and generalizability	

For Model 1:

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score

# Separate features and target
X = filtered_data.drop(columns=['booking_status'])
y = filtered_data['booking_status']

# Split data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Train a Random Forest classifier
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict probabilities for validation set
val_preds = model.predict_proba(X_val)[:, 1]

# Calculate ROC AUC score
auc_score = roc_auc_score(y_val, val_preds)
print("Validation ROC AUC Score:", auc_score)

# Make predictions on test set
test_preds = model.predict_proba(test_data)[:, 1]
```

Validation ROC AUC Score: 0.9337955521191528

For Model 2:

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import roc_auc_score

# Separate features and target
X = filtered_data.drop(columns=['booking_status'])
y = filtered_data['booking_status']

# Split data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Train a Decision Tree classifier
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predict probabilities for validation set
val_preds_dt = dt_model.predict_proba(X_val)[:, 1]

# Calculate ROC AUC score
auc_score_dt = roc_auc_score(y_val, val_preds_dt)
print("Validation ROC AUC Score for Decision Tree:", auc_score_dt)

# Make predictions on test set
test_preds_dt = dt_model.predict_proba(test_data)[:, 1]
```

Validation ROC AUC Score for Decision Tree: 0.8360734124158138