Model Development Phase Template

Date	23 September 2024
Team ID	LTVIP2024TMID24986
Project Title	Movie Box Office Gross Prediction using Machine Learning
Maximum Marks	4 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 classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

```python
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

def linear_regression(X_train, X_test, y_train, y_test):
    lr = LinearRegression()
    lr.fit(X_train, y_train)
    y_pred = lr.predict(X_test)
    print("**Linear Regression**")
    print("Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
    print("Mean Squared Error: ", mean_squared_error(y_test, y_pred))
    print("R2 Score: ", r2_score(y_test, y_pred))
```

```python
from sklearn.tree import DecisionTreeRegressor

def decision_tree_regression(X_train, X_test, y_train, y_test):
    dt = DecisionTreeRegressor()
    dt.fit(X_train, y_train)
    y_pred = dt.predict(X_test)
    print("**Decision Tree Regression**")
    print("Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
    print("Mean Squared Error: ", mean_squared_error(y_test, y_pred))
    print("R2 Score: ", r2_score(y_test, y_pred))
```

```python
from sklearn.svm import SVR

def svm_regression(X_train, X_test, y_train, y_test):
    svm = SVR()
    svm.fit(X_train, y_train)
    y_pred = svm.predict(X_test)
    print("**Support Vector Machine Regression**")
    print("Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
    print("Mean Squared Error: ", mean_squared_error(y_test, y_pred))
    print("R2 Score: ", r2_score(y_test, y_pred))
```

```python
from sklearn.ensemble import RandomForestRegressor

def random_forest_regression(X_train, X_test, y_train, y_test):
    rf = RandomForestRegressor()
    rf.fit(X_train, y_train)
    y_pred = rf.predict(X_test)
    print("**Random Forest Regression**")
    print("Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
    print("Mean Squared Error: ", mean_squared_error(y_test, y_pred))
    print("R2 Score: ", r2_score(y_test, y_pred))
```

Model Validation and Evaluation Report:

Model	Evaluation Metric	R^2, MSE, MAE	Confusion Matrix
Linear Regression	R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE)	0.7758, 8649.14, 54.377	N/A
SVM Regression	R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE)	0.1528, 32681.37, 77.861	N/A

Decision Tree Regressor	R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE)	0.5318, 18060.04, 73.851	N/A
Random Forest Regressor	R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE)	0.7617, 9190.28, 50.225	N/A