

Model Development Phase

Date	24 June 2025
Team ID	SWTID1749705685
Project Title	Movie Box Office Gross Prediction using Machine Learning
Maximum Marks	4 Marks

Initial Model Training Code, Model Validation and Evaluation Report

Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions.

Root Mean Squared Error (RMSE): Penalizes larger errors more significantly.

R² Score (Coefficient of Determination): Indicates how well the model explains variance in the target variable.

Initial Model Training Code:

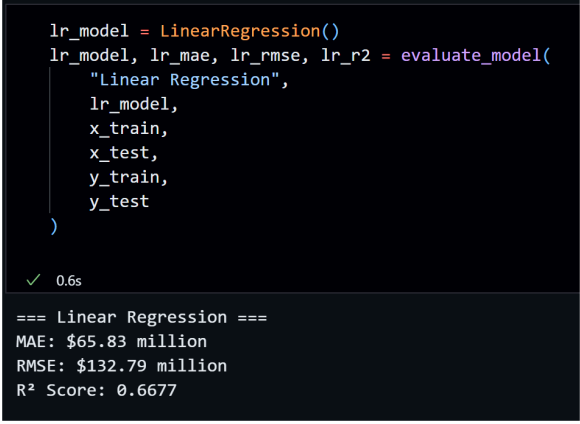
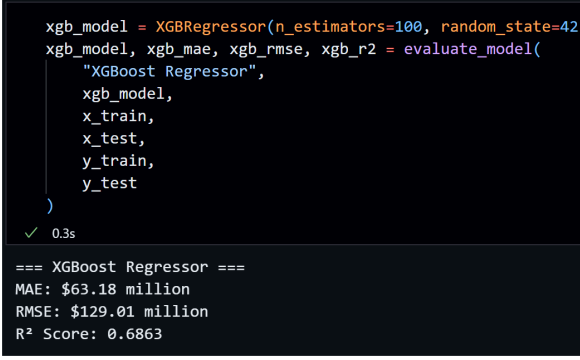
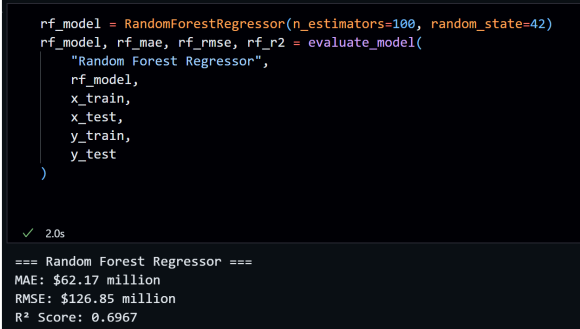
```
x = movies_box.drop('revenue', axis=1)
y = movies_box['revenue']

scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)

x_train, x_test, y_train, y_test = train_test_split(x_scaled, y, test_size=0.1, random_state=0)

mr = LinearRegression()
mr.fit(x_train, y_train)
```

Model Validation and Evaluation Report:

Model	MAE	RMSE	R ²	Screenshot
Linear Regression	\$65.83 million	\$132.79 million	0.6677	 <pre> lr_model = LinearRegression() lr_model, lr_mae, lr_rmse, lr_r2 = evaluate_model("Linear Regression", lr_model, x_train, x_test, y_train, y_test) ✓ 0.6s === Linear Regression === MAE: \$65.83 million RMSE: \$132.79 million R² Score: 0.6677 </pre>
XGBoost Regressor	\$63.18 million	\$129.01 million	0.6863	 <pre> xgb_model = XGBRegressor(n_estimators=100, random_state=42) xgb_model, xgb_mae, xgb_rmse, xgb_r2 = evaluate_model("XGBoost Regressor", xgb_model, x_train, x_test, y_train, y_test) ✓ 0.3s === XGBoost Regressor === MAE: \$63.18 million RMSE: \$129.01 million R² Score: 0.6863 </pre>
Random Forest Regressor	\$62.17 million	\$126.85 million	0.6967	 <pre> rf_model = RandomForestRegressor(n_estimators=100, random_state=42) rf_model, rf_mae, rf_rmse, rf_r2 = evaluate_model("Random Forest Regressor", rf_model, x_train, x_test, y_train, y_test) ✓ 2.0s === Random Forest Regressor === MAE: \$62.17 million RMSE: \$126.85 million R² Score: 0.6967 </pre>