Model Development Phase

Date	19 June 2025
Team ID	xxxxxx
Project Title	sloan digital sky survey (sdss) galaxy classification using machine learning
Maximum Marks	4 Marks

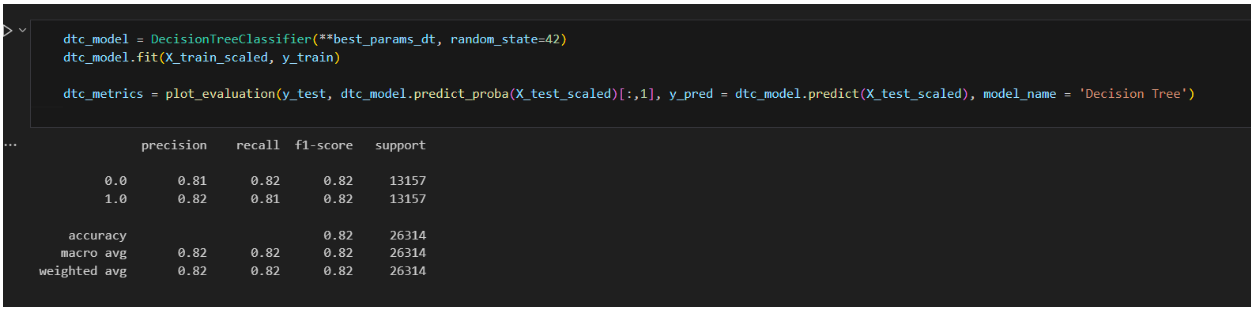
Initial Model Training Code, Model Validation and Evaluation Report

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:

Scenario 1:

Decision Tree Classifier:



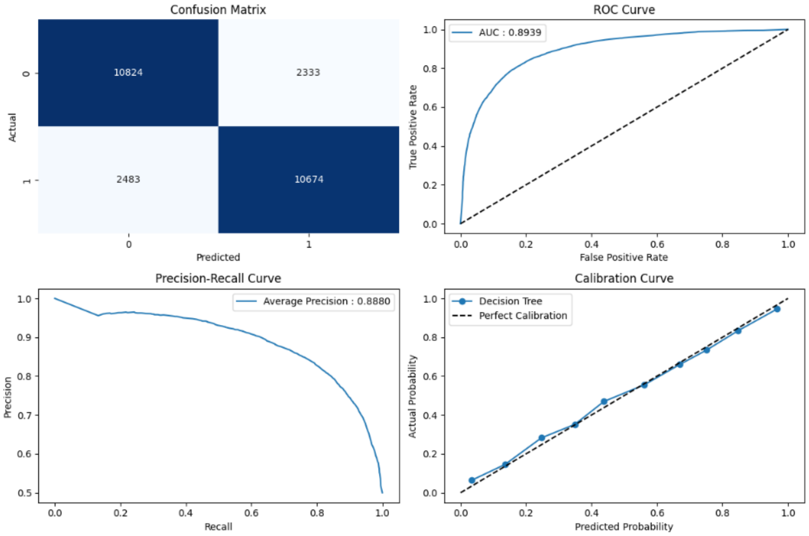
```
dtc_model = DecisionTreeClassifier(**best_params_dt, random_state=42)
dtc_model.fit(X_train_scaled, y_train)

dtc_metrics = plot_evaluation(y_test, dtc_model.predict_proba(X_test_scaled)[:,1], y_pred = dtc_model.predict(X_test_scaled), model_name = 'Decision Tree')
```

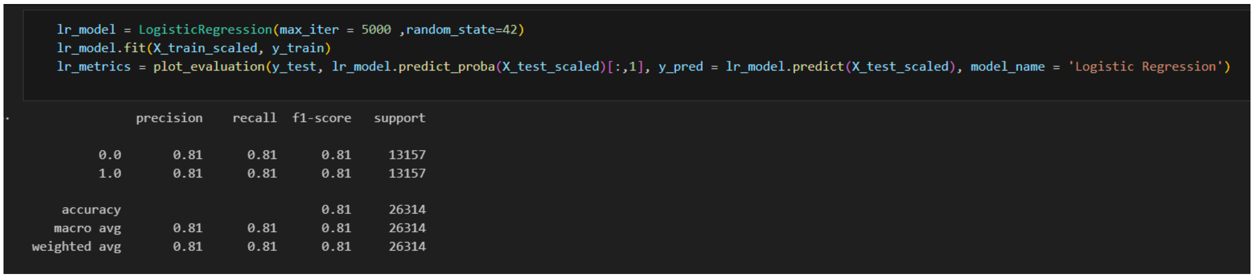
```
              precision    recall  f1-score   support

         0.0       0.81      0.82      0.82     13157
         1.0       0.82      0.81      0.82     13157

    accuracy                           0.82     26314
   macro avg       0.82      0.82      0.82     26314
weighted avg       0.82      0.82      0.82     26314
```



Logistic Regression:

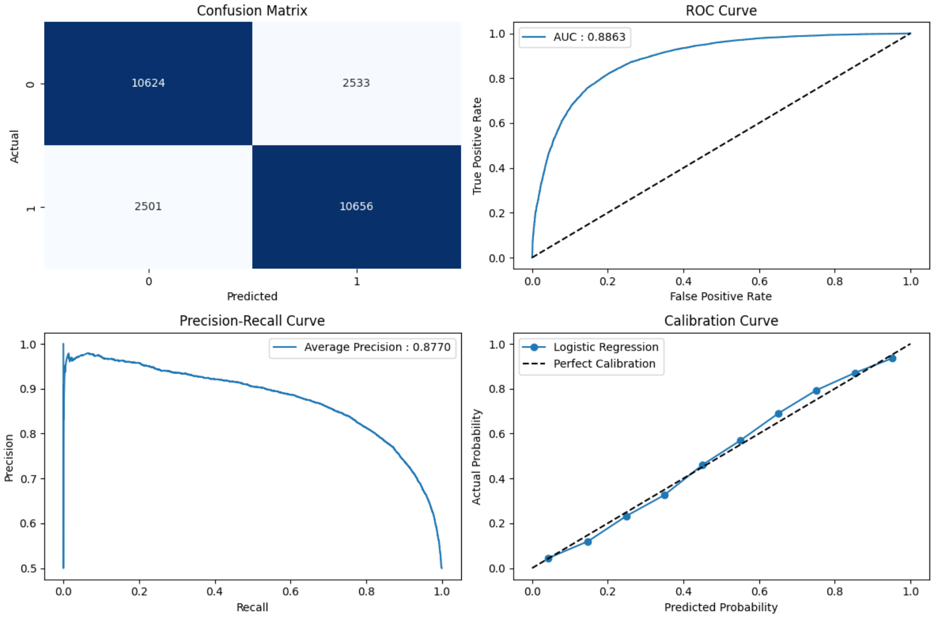


```
lr_model = LogisticRegression(max_iter = 5000 ,random_state=42)
lr_model.fit(X_train_scaled, y_train)
lr_metrics = plot_evaluation(y_test, lr_model.predict_proba(X_test_scaled)[:,1], y_pred = lr_model.predict(X_test_scaled), model_name = 'Logistic Regression')
```

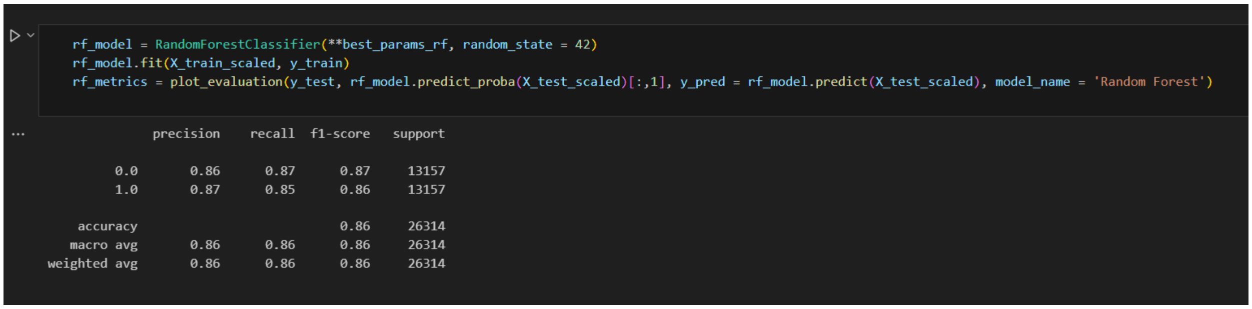
```
              precision    recall  f1-score   support

         0.0       0.81      0.81      0.81     13157
         1.0       0.81      0.81      0.81     13157

    accuracy                           0.81     26314
   macro avg       0.81      0.81      0.81     26314
weighted avg       0.81      0.81      0.81     26314
```



Random Forest Classifier:

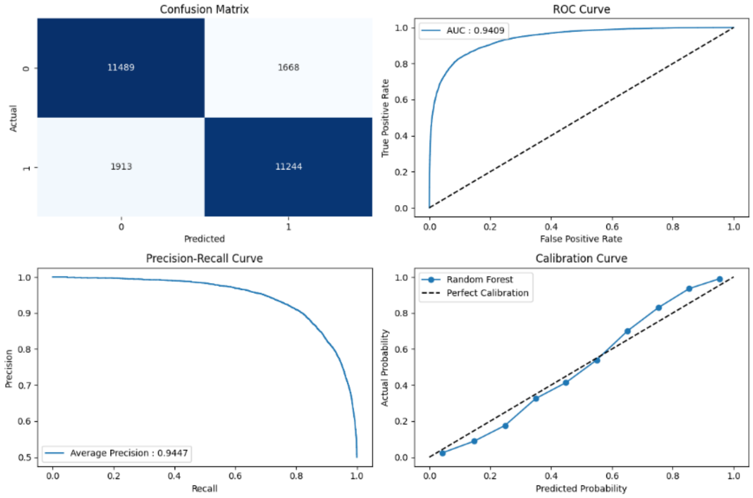


```
rf_model = RandomForestClassifier(**best_params_rf, random_state = 42)
rf_model.fit(X_train_scaled, y_train)
rf_metrics = plot_evaluation(y_test, rf_model.predict_proba(X_test_scaled)[:,1], y_pred = rf_model.predict(X_test_scaled), model_name = 'Random Forest')
```

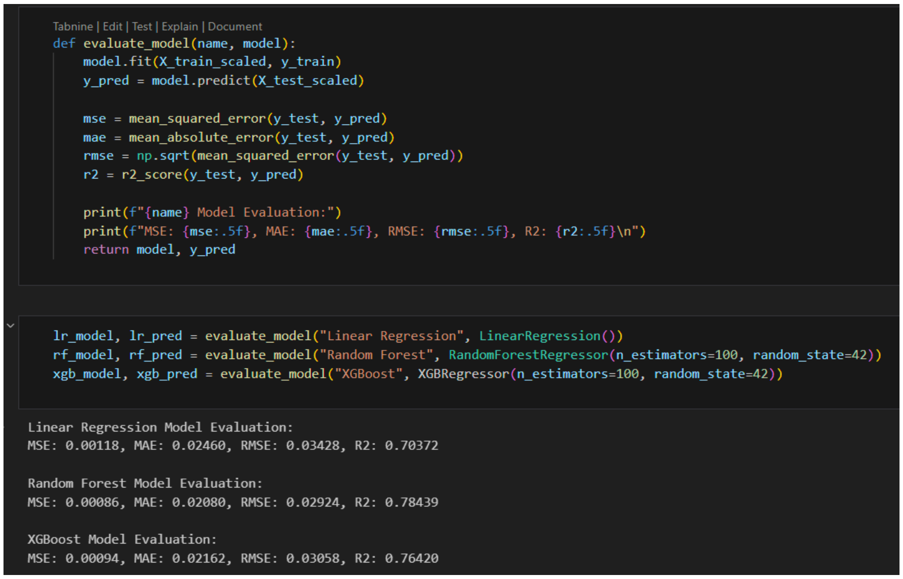
```
              precision    recall  f1-score   support

         0.0       0.86      0.87      0.87     13157
         1.0       0.87      0.85      0.86     13157

    accuracy                           0.86     26314
   macro avg       0.86      0.86      0.86     26314
weighted avg       0.86      0.86      0.86     26314
```



Scenario 2:



```
Tabnine | Edit | Test | Explain | Document
def evaluate_model(name, model):
    model.fit(X_train_scaled, y_train)
    y_pred = model.predict(X_test_scaled)

    mse = mean_squared_error(y_test, y_pred)
    mae = mean_absolute_error(y_test, y_pred)
    rmse = np.sqrt(mean_squared_error(y_test, y_pred))
    r2 = r2_score(y_test, y_pred)

    print(f"{name} Model Evaluation:")
    print(f"MSE: {mse:.5f}, MAE: {mae:.5f}, RMSE: {rmse:.5f}, R2: {r2:.5f}\n")
    return model, y_pred
```

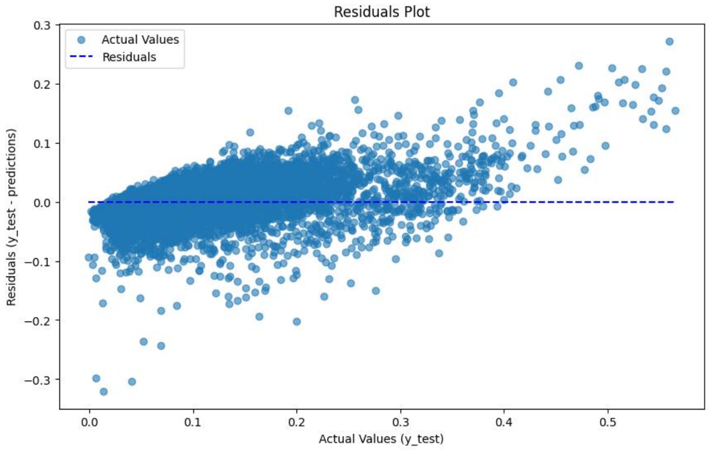
```
lr_model, lr_pred = evaluate_model("Linear Regression", LinearRegression())
rf_model, rf_pred = evaluate_model("Random Forest", RandomForestRegressor(n_estimators=100, random_state=42))
xgb_model, xgb_pred = evaluate_model("XGBoost", XGBRegressor(n_estimators=100, random_state=42))
```

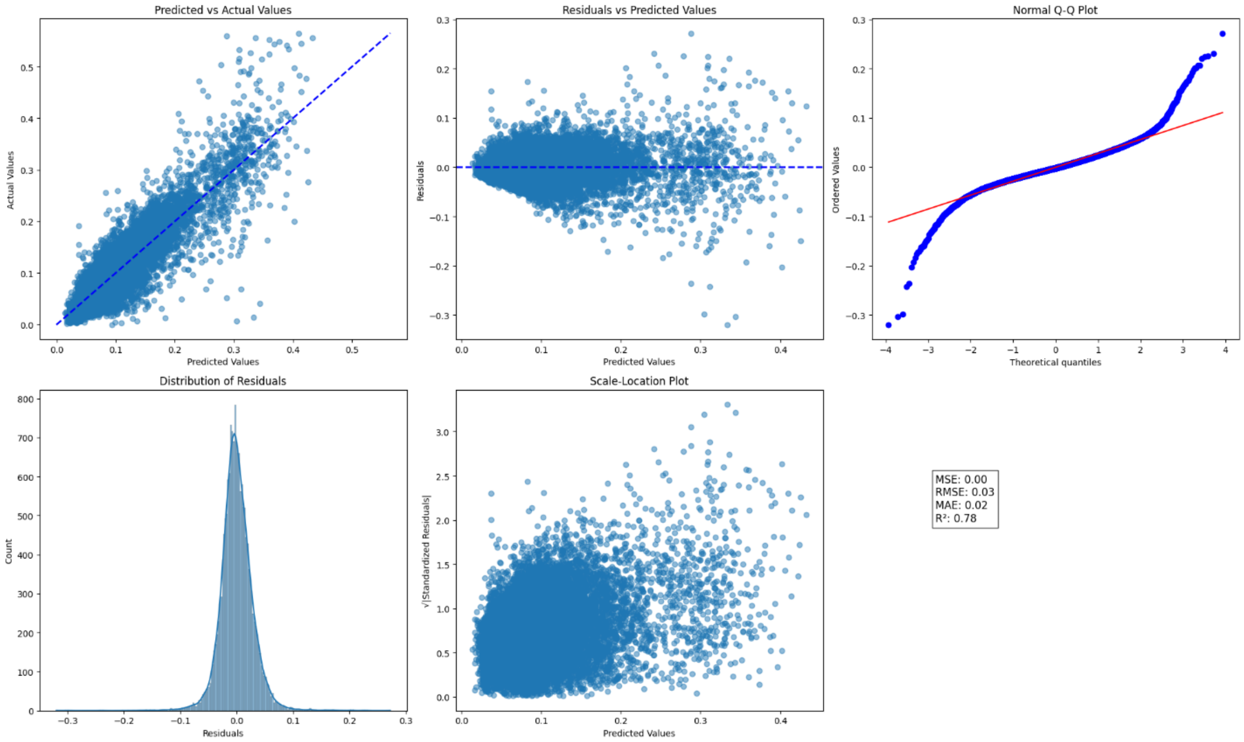
```
Linear Regression Model Evaluation:
MSE: 0.00118, MAE: 0.02460, RMSE: 0.03428, R2: 0.70372

Random Forest Model Evaluation:
MSE: 0.00086, MAE: 0.02080, RMSE: 0.02924, R2: 0.78439

XGBoost Model Evaluation:
MSE: 0.00094, MAE: 0.02162, RMSE: 0.03058, R2: 0.76420
```

Results:


Residuals Plot
Actual Values
Residuals
Residuals (y_test - predictions)
Actual Values (y_test)
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.0
0.1
0.2
0.3
0.4
0.5


Predicted vs Actual Values
Actual Values
Predicted Values
Residuals vs Predicted Values
Residuals
Predicted Values
Normal Q-Q Plot
Ordered Values
Theoretical quantiles
Distribution of Residuals
Count
Residuals
Scale-Location Plot
√|Standardized Residuals|
Predicted Values
MSE: 0.00
RMSE: 0.03
MAE: 0.02
R²: 0.78