

# Loan Approval Prediction

## XGBoost Performance Results

### Final Features :

```
['lux_asset_ratio',  
 'residential_assets_value',  
 'com_asset_ratio',  
 'loan_income_ratio',  
 'loan_term',  
 'bank_asset_ratio',  
 'education',  
 'loan_amount',  
 'loan_lux_asset_ratio',  
 'res_asset_ratio']
```

### Classification Report :

Metric	Train Data	Test Data
ROC AUC	0.6744	0.5825
Gini	0.3487	0.1650
Recall	0.9868	0.9736
Precision	0.6303	0.6206
Accuracy	0.6316	0.6136

### Confusion Matrix for Train Data:

```
[[ 60 1230]  
 [ 28 2097]]
```

### Confusion Matrix for Test Data:

```
[[ 7 316]  
 [14 517]]
```

- **ROC AUC:** The XGBoost model achieved an ROC AUC of 0.5825 on the test data, indicating its moderate ability to distinguish between positive and negative cases.
- **Gini:** The Gini coefficient for the XGBoost model is 0.1650. Gini is derived from ROC AUC and is another measure of model discrimination. A higher Gini is generally desirable.
- **Recall:** The model's recall (sensitivity) is 0.9736 on the test data. This indicates that the model correctly identified a high percentage of actual loan approvals.
- **Precision:** The precision of the XGBoost model is 0.6206 on the test data. This means that out of the cases predicted as loan approvals, 62.06% were actually loan approvals.
- **Accuracy:** The model achieved an accuracy of 0.6136 on the test data, which measures overall correctness. However, accuracy can be influenced by class imbalances.

## Logistic Regression Performance Results

### Final Features :

```
['loan_term_optbinned',  
 'lux_asset_ratio_optbinned',  
 'loan_res_asset_ratio_optbinned',  
 'bank_asset_ratio_optbinned',  
 'com_asset_ratio_optbinned',  
 'luxury_assets_value_optbinned',  
 'loan_com_asset_ratio_optbinned',  
 'education_optbinned',  
 'loan_income_ratio_optbinned',  
 'no_of_dependents_optbinned',  
 'commercial_assets_value_optbinned',  
 'loan_bank_asset_ratio_optbinned',  
 'self_employed_optbinned',  
 'loan_amount_optbinned',  
 'income_annum_optbinned',  
 'residential_assets_value_optbinned',  
 'res_asset_ratio_optbinned',  
 'bank_asset_value_optbinned',  
 'loan_lux_asset_ratio_optbinned']
```

### Classification Report:

Metric	Train Data	Test Data
ROC AUC	0.6581	0.5867
Gini	0.3163	0.1734
Recall	0.8631	0.8493
Precision	0.6614	0.6434
Accuracy	0.6398	0.6136

### Confusion Matrix for Train Data:

```
[[ 351  939]  
 [ 291 1834]]
```

### Confusion Matrix for Test Data:

```
[[ 73 250]  
 [ 80 451]]
```

- **ROC AUC:** The Logistic Regression model achieved an ROC AUC of 0.5867 on the test data, slightly better than XGBoost.
- **Gini:** The Gini coefficient for the Logistic Regression model is 0.1734, also slightly better than XGBoost.
- **Recall:** The model's recall (sensitivity) is 0.8493 on the test data, indicating that it correctly identified a high percentage of actual loan approvals.

- **Precision:** The precision of the Logistic Regression model is 0.6434 on the test data, meaning that out of the cases predicted as loan approvals, 64.34% were actually loan approvals.
- **Accuracy:** The model achieved an accuracy of 0.6136 on the test data, similar to XGBoost.

Both XGBoost and Logistic Regression models seem to perform reasonably well. Since the number of features is less than logistic regression, I have chosen XGBOOST as the final model.