LoanAI

Model Performance Report

Generated on February 21, 2026

Model Overview

This report presents the performance evaluation of the **XGBClassifier (Pipeline)** model trained for loan approval prediction. The model uses the **XGBoost** algorithm with **157** estimators, a maximum depth of **9**, and a learning rate of **0.1712**. The model was trained on **4,269** samples with **15** features.

Parameter	Value
Algorithm	XGBoost
Pipeline Name	XGBClassifier (Pipeline)
Number of Estimators	157
Max Depth	9
Learning Rate	0.1712
Subsample Ratio	0.768
Total Training Samples	4,269
Number of Features	15

Table 1: Model Hyperparameters & Configuration

Performance Metrics

The following metrics evaluate the model's classification performance on the full dataset. All values are rounded to four decimal places.

0.9995 Accuracy	**0.9996** Precision	**0.9996** Recall	**0.9996** F1 Score	**1.0000** ROC AUC
0.9994 Specificity	**0.9990** MCC	**0.9990** Cohen's κ	**0.0026** Log Loss	**1.0000** Avg Precision

Figure 1: Key Classification Metrics Summary

Confusion Matrix

The confusion matrix below shows the model's prediction accuracy. Out of **4,269** total samples, only **2** were misclassified — **1** false positive(s) and **1** false negative(s).

	Predicted Rejected	Predicted Approved
Actual Rejected	**1612**	**1**
Actual Approved	**1**	**2655**

Figure 2: Confusion Matrix (TN, FP, FN, TP)

Cross-Validation Results

5-fold stratified cross-validation was performed to assess generalization. The table below shows the scores for each fold along with the mean and standard deviation.

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
Accuracy	0.9988	0.9953	0.9941	0.9977	0.9988	**0.9970**	0.0019
F1	0.9991	0.9962	0.9953	0.9981	0.9991	**0.9976**	0.0015
Precision	1.0000	0.9925	0.9944	0.9981	1.0000	**0.9970**	0.0030
Recall	0.9981	1.0000	0.9962	0.9981	0.9981	**0.9981**	0.0012

Table 2: 5-Fold Stratified Cross-Validation Scores

Dataset Class Distribution

The dataset contains **4,269** samples in total, with **2,656** approved (62.2%) and **1,613** rejected (37.8%). The moderate class imbalance was addressed during model training.

Class	Count	Percentage	
Approved (1)	2,656	62.2%	
Rejected (0)	1,613	37.8%	

Figure 3: Approved vs Rejected Class Distribution

ROC & Precision-Recall Summary

The model achieves an **ROC AUC of 1.0000** and an **Average Precision of 1.0000**, indicating near-perfect discrimination between approved and rejected loan applications.

Curve	Key Metric	Value	Interpretation
ROC Curve	AUC	**1.0000**	Perfect separation between classes
PR Curve	Avg Precision	**1.0000**	Excellent precision across all thresholds

Table 3: ROC and Precision-Recall Curve Metrics

Feature Importance

The chart below shows the relative importance of each feature in the model's decision-making process. CIBIL Score dominates with 73% importance, followed by Loan-to-Income ratio (10.1%) and Loan Term (9.8%).

Rank	Feature	Importance	Visual
1	**cibil_score**	73.01%	
2	**loan_to_income**	10.12%	
3	**loan_term**	9.79%	
4	loan_to_asset	3.57%	
5	asset_to_income	1.18%	
6	no_of_dependents	0.39%	
7	income_annum	0.34%	
8	residential_assets_value	0.27%	
9	loan_amount	0.26%	
10	total_assets_value	0.25%	
11	luxury_assets_value	0.24%	
12	bank_asset_value	0.21%	
13	self_employed_Yes	0.15%	
14	education_Not Graduate	0.12%	
15	commercial_assets_value	0.10%	

Figure 4: Feature Importance Ranking (XGBoost)

Key Insights

• ■ **CIBIL Score** is the dominant feature (73.01%), confirming that credit history is the strongest predictor of loan approval.

• ■ **Loan-to-Income Ratio** (10.12%) is the second most important feature, reflecting the applicant's debt burden relative to income.

• ■ **Loan Term** (9.79%) significantly impacts predictions — longer terms may indicate higher risk profiles.

• ■ **Financial Ratios** (Loan-to-Asset: 3.57%, Asset-to-Income: 1.18%) together contribute ~4.75% importance.

• ■ **Raw Asset Values** (residential, commercial, luxury, bank) each contribute less than 0.3%, as the engineered ratios capture most information.

CIBIL Score Distribution

The CIBIL score distribution reveals a clear decision boundary around the **550 score mark**. Almost all rejections occur below CIBIL 550, while approvals are concentrated in the 550–900 range.

CIBIL Range	Approved	Rejected	Total	Approval Rate
300-350	41	312	353	12%
350-400	31	320	351	9%
400-450	43	303	346	12%
450-500	33	317	350	9%
500-550	37	348	385	10%
550-600	346	2	348	99%
600-650	335	0	335	100%
650-700	367	4	371	99%
700-750	373	1	374	100%
750-800	358	3	361	99%
800-850	354	1	355	100%
850-900	338	2	340	99%

Table 4: CIBIL Score Distribution by Loan Decision

Conclusion

The **XGBClassifier (Pipeline)** model demonstrates **exceptional performance** across all evaluation metrics:

■ **Near-perfect accuracy** of 99.95% with only 2 misclassifications out of 4,269 samples.

■ **Excellent generalization** confirmed by 5-fold cross-validation (mean accuracy: 99.70%, $\sigma = 0.19\%$).

■ **CIBIL Score** is the most influential feature (73%), followed by loan-to-income ratio (10.1%) and loan term (9.8%).

■ **Clear decision boundary** at CIBIL ≈ 550, with high approval rates above and high rejection rates below this threshold.

■ **Perfect AUC score** of 1.0000 indicates optimal trade-off between true positive and false positive rates.