

PREDICTING LOAN DEFAULT RISK USING SAS

PREDICTIVE MODELING

1. Overview

Client:

A mid-sized U.S. fintech lender managing personal loans and small business credit lines

Objective:

To develop a logistic regression-based predictive model in SAS for identifying borrowers at high risk of default. The goal was to enhance credit decisioning and support automated risk scoring for new loan applications.

2. Background

The client had rising default rates among first-time borrowers. Their existing rule-based system failed to adjust to changing macroeconomic patterns. They needed a statistically sound, interpretable model built in SAS that could handle large datasets and integrate directly with their underwriting workflow.

3. Data Summary

Dataset:

Loan application and performance data (N = 28,500 records)

Variables Used:

Variable	Type	Description
Default_Status (0/1)	Binary	Target variable: loan default within 12 months
Income	Continuous	Declared monthly income (USD)
Loan_Amount	Continuous	Amount disbursed
Credit_Score	Continuous	FICO score at application
Employment_Length	Categorical	<1 year, 1–3 years, 3–5 years, 5+ years
Home_Ownership	Categorical	Rent / Own / Mortgage
Loan_Purpose	Categorical	Debt consolidation, personal use, etc.

DTI_Ratio	Continuous	Debt-to-income ratio (%)
-----------	------------	--------------------------

4. Methodology

Software Used:

SAS 9.4 (Base + Enterprise Guide)

SAS Procedures Employed:

1. Data Cleaning & Transformation:

- PROC IMPORT, DATA steps for formatting
- Dummy variables created using PROC GLMMOD
- Outlier handling via winsorization and boxplots

2. Descriptive Analysis:

- PROC MEANS, PROC FREQ, PROC UNIVARIATE
- Checked multicollinearity with PROC CORR and VIF analysis

3. Model Development (Logistic Regression):

- PROC LOGISTIC with stepwise selection
- Response variable: Default_Status
- Goodness-of-fit via Hosmer–Lemeshow test
- AUC and KS statistic calculated

4. Model Validation:

- Train/test split: 70% training, 30% validation
- PROC LOGISTIC + ROC analysis
- Scorecards generated using PROC SCORE for internal risk team

5. Key Results

Variable	Odds Ratio	p-value	Interpretation
Credit_Score	0.67	<0.001	Higher score reduces likelihood of default

DTI_Ratio	1.15	0.008	Higher debt burden increases default odds
Employment_Length (<1 yr)	1.42	0.012	Instability linked to elevated risk
Loan_Amount	1.06	0.022	Larger loans marginally increase default probability

Model Performance:

- AUC (Validation): 0.78
- KS Statistic: 0.46
- Hosmer–Lemeshow $p = 0.61$ (well-calibrated model)

6. Visual Outputs (from SAS)

- ROC curve with training vs. validation comparison
- Decile-wise lift chart
- Risk score distribution plot
- Influence and leverage plot for model robustness

7. Deliverables

- .sas code with macro-enabled scoring functionality
- Full report (18 pages) covering:
 - Model development workflow
 - Variable interpretations and business logic
 - Model performance and validation plots
 - Score thresholds and segmentation strategy
- Risk dashboard slides for credit team (6 pages):
 - Credit policy implications
 - Risk segmentation by income and credit score bands
 - Application in decision tree logic for loan approval

8. Client Application & Outcome

- Integrated into automated underwriting system within 4 weeks
- Reduced default rate on new loans by 17% within two quarters
- Informed the creation of a tiered interest rate system based on borrower risk profile

9. Strategic Value Delivered

- Delivered an **interpretable, production-ready risk model**
- Enabled **automated, data-driven credit decisions** with reduced human bias
- Improved **portfolio risk management** and borrower segmentation logic