LendSmart Credit Risk Analysis: Technical Report

Sophia Gabriela Martínez Albarrán, Sibyla Vera Avila, Regina Pérez Vázquez

Instituto de Estudios Superiores Tecnológico de Monterrey CCM

November 2025

1. Introduction

This technical report summarizes the analytical approach used to evaluate credit risk for LendSmart, a consumer lending company experiencing a default rate of approximately 26.6%. The objective was to develop classification models capable of predicting high-risk applicants before loan approval. Two multivariate statistical methods were applied: Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

2. Dataset Overview

The dataset consists of 2,500 loan applications with 18 variables, including demographic, financial and behavioral indicators. The target variable, `loan_status`, classifies borrowers as either good (0) or defaulting (1).

Key predictive variables include credit score, payment history score, credit utilization, job stability, savings ratio, and debt-to-income ratio. The dataset is complete (0% missing data) and displays strong separation between good and defaulting borrowers.

3. Methodology

The analysis followed a standard multivariate workflow:

- Data preprocessing (scaling, exploration, distribution checks)
- Train–test split for model evaluation
- Implementation of LDA and QDA
- Performance assessment using accuracy, precision, recall, and AUC

Model Assumptions

LDA assumes equal covariance matrices across groups, while QDA allows class-specific covariance structures. Both models assume approximate multivariate normality, which was reasonably satisfied after preprocessing.

4. Results

Both models achieved perfect classification results:

- **Accuracy:** 1.000
- **Precision:** 1.00
- **Recall:** 1.00
- **AUC:** 1.000

Although performance was identical, QDA offers greater flexibility because it does not assume equal covariance matrices. This makes it more robust to changes in borrower distributions.

5. Key Insights

- Payment history is the strongest predictor of default.
- High credit utilization and low job stability are strong risk indicators.
- Borrowers with higher savings ratios and stable employment show lower probability of default.

6. Business Recommendations

- Deploy QDA as the primary risk classification model.
- Integrate behavioral risk factors (payment history, utilization) into automated decision rules.
- Monitor model performance and covariance structure quarterly to detect drift.

7. Conclusion

The discriminant analysis performed in this project provides a reliable and interpretable framework for identifying high-risk borrowers. With perfect predictive performance and strong theoretical foundations, QDA is recommended as the operational model for LendSmart's credit evaluation processes.