LendSmart Credit Risk Analysis

Course: Application of multivariate methods in data science

Professor: Juliho David Castillo Colmenares

Team members:

Alan Ulises Luna Hernández *A01424523*

Axel Palacios Granados *A01666972*

Cristian Cruz Orozco *A01665590*

Octubre, 2025

0.1. Executive Summary

0.1.1. Business Problem

LendSmart is a *FinTech* company that provides personal and small business loans. Currently, about **28 %** of its loans end in default, which leads to direct losses and a level of risk that management considers too high.

The goal of this project was to analyze the history of **2,500 applicants** in order to build a model that estimates the probability that a new applicant will default on their loan (`loan_status`). With this model, LendSmart can support its approval/rejection decisions and adjust conditions (amount, interest rate, etc.) to reduce losses.

Two related statistical approaches (LDA and QDA) were evaluated to select the most suitable one to integrate into the credit risk process.

0.1.2. Key Findings and Insights

The analysis of historical information, together with the distribution plots and the correlation matrix, reveals a very clear profile of borrowers who default versus those who pay on time.

2.1 Low-Risk Customer Profile

Customers who pay their loans on time tend to have:

- **High credit scores**, typically concentrated between 680 and 780 points (median close to 720).

- **Higher annual incomes**, with median around 75,000–80,000 USD.

- **Low debt-to-income ratios**, typically between 0.15 and 0.40.

- Moderate use of their available credit lines (loans are not "at the limit").

High-Risk Customer Profile

Borrowers who default are characterized by:

- **Lower credit scores**, concentrated between 500 and 650 points.

- **Lower annual incomes**, with a median clearly below the non-default group.

- **High debt-to-income ratios**, often in the 0.60–0.80 range.

- **High credit utilization**, i.e., usage close to the limit of their credit lines.

The LDA model confirms these patterns: the factors with the strongest influence on default risk are **payment history**, **job stability**, **credit utilization**, **debt-to-income ratio**, and **overall credit score**. A strong payment history, stable employment, and a healthy level of debt act as **protective factors**; being near the limit of credit lines and having a large share of income already committed to payments are **clear warning signs**.

Sociodemographic variables (education and marital status) show some groups with higher default rates (e.g., *High School* level or *Single/Widowed* status), but their impact is smaller compared to financial indicators.

0.1.3. Model Performance and Selection

Two models were built and compared using the same information:

- **Model A:** LDA.

- **Model B:** QDA.

On the test set (customers not used to train the model), both approaches delivered **exactly the same result**:

- 367 customers who in reality **did not default** were classified as **low risk**.

- 133 customers who **did default** were identified as **high risk**.

- No false positives and no false negatives were observed in this set.

In other words, on these data the model separates good and bad payers perfectly. The ROC curves for both models overlap with an AUC of 1.0, confirming this separation.

Since LDA and QDA show the same performance, the decision is driven by ease of use. The LDA model is simpler to interpret and explain, so it is recommended as the **chosen technical model**.

0.1.4. Final Recommendation

Proposed Decision

The recommendation is **"Go"**:

> LendSmart should adopt the LDA model as a decision-support tool to evaluate credit applications.

This model helps minimize the most costly scenario (approving loans that are very likely to default) while keeping under control the risk of rejecting good customers, thanks to the strong separation observed in the historical data.

Suggested Next Steps

Although historical results are very favorable, in real-world operations there will always be some errors. Therefore, the following is proposed:

1. Deploy the model initially in a **pilot phase**, comparing its recommendations with the current decisions of the risk team.

2. Establish **monthly monitoring** of the actual default rate and of the good customers the model may be rejecting.

3. Adjust the **decision threshold** according to LendSmart's risk appetite.

In conclusion, the analysis indicates that LendSmart can rely on a **solid, data-driven credit risk model** that can support strategic decisions to reduce the default rate and improve the profitability of its loan portfolio.