Tecnológico de Monterrey

Campus Ciudad de México

November 9, 2025

Aplicación de métodos multivariados en ciencia de datos

Group 602

LendSmart Credit Risk Analysis

Names and IDs

Helena Eridani Escandón López - A01659511

Ofelia Gabriela Góngora Méndez - A01666131

Armando Atanasio Navarrete Yépez - A01658529

Professors

Juliho David Castillo Colmenares

Jonathan Montalvo Urquizo

1. Business Problem

LendSmart, a FinTech company specializing in personal and small business loans, is facing a high loan default rate and needs to reduce their current 28%. This percentage threatens its profitability and long-term sustainability since each time a high-risk applicant is mistakenly approved, the institution incurs significant financial losses.

Our team was asked to develop a statistical classification model capable of identifying high-risk loan applicants before a loan is approved. The goal is to support LendSmart's credit department in reducing defaults by flagging potentially risky applicants for additional review or rejection.

2. Key Findings & Insights:

Our analysis showed that the dataset provided by LendSmart has perfect separability between clients who pay their loans correctly and clients who default, meaning that it can clearly distinguish between good clients and risky clients, with no mistakes. The data contains strong predictive indicators of credit behavior; from a business perspective, a high-risk client is one that closely matches the profiles of the default clients present on the current dataset.

The profile of a high-risk applicant was mainly determined by the credit score (strongest predictor), the debt-to-income ratio, the employment stability, education level, the asset ownership and savings. This means that an applicant with a low credit score, whose total debt represents a high percentage of their income, with irregular employment histories, a lower education, with few declared assets and minimal savings represent a greater risk to the company and therefore shouldn't be considered as possible clients.

3. Model Performance & Selection:

We compared two models (LDA and QDA) and both of them had a Recall score of 1, which translates to identifying 100% of all true defaulters and with this, no good applicants are incorrectly rejected. This implies that both models have a perfect performance in this context and thanks to the perfect AUC of 1.0, both would be suitable for distinguishing between high-risk and low-risk applicants. Even though choosing any of the models would be adequate, we consider that using LDA in this case is preferable since it is simpler, more interpretable, and less prone to overfitting, which is ideal to explain in business terms.

4. Final Recommendation:

We recommend LendSmart to adopt the LDA as a default risk-assessment tool but a validation on new data should be made to confirm that it keeps its perfect performance or at least a higher performance than QDA.

In this test, there were no false negatives or false positives, meaning the model perfectly classified all applicants. Using this model would represent a substantial reduction in losses due to default clients, and even a slightly lower accuracy rate (from 98% to 95%) would still be beneficial to LendSmart's profits. While real-world data might reduce performance slightly, even a 95% accuracy rate represents a strong business advantage.

As mentioned, if the performance remains consistent, LendSmart can drastically reduce loan defaults, saving substantial losses tied to high-risk approvals. The model can also optimize credit approvals for low-risk clients, increasing operational efficiency.

Future work should focus on validating the model on a new and unclean dataset with real data to review its performance and decide whether to continue with LDA or choose QDA for better results. After this, the LDA model or the chosen model should be integrated into the system to identify high-risk clients and stop their application.