LendSmart Credit Risk Analysis – Executive Summary

*

Instituto Tecnológico y de Estudios Superiores de Monterrey

1 Business Problem

LendSmart currently faces a loan default rate of 26.5%, which poses a major challenge to profitability and portfolio stability. The company's primary goal is to identify high-risk applicants early in the loan evaluation process to reduce default losses without unfairly rejecting creditworthy customers.

The objective of this project was to develop and compare statistical classification models — Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) — to predict the probability of loan default. This predictive system aims to support LendSmart's risk management team in making faster, data-driven, and more consistent credit decisions.

2 Key Findings & Insights

An exploratory analysis of 2,500 loan applications (2022–2024) revealed clear behavioral and financial differences between reliable and high-risk borrowers.

Main Insights

Defaulters typically have lower annual income, higher debt-to-income ratios, poor payment history, and unstable employment.

The strongest predictors of default were:

- **Credit utilization** – strongest correlation ($r = 0.91$)
- **Debt-to-income ratio** – ($r = 0.79$)
- **Payment history and job stability** – inverse correlation ($r \approx -0.85$ combined)

Credit score, asset value, and annual income also showed significant predictive power.

From a demographic perspective, applicants with high school education and single or widowed status were more likely to default, whereas married borrowers with college or graduate education had the lowest risk levels.

High-Risk Profile

The typical high-risk applicant can be described as:

- High credit utilization ($> 60\%$)
- Debt-to-income ratio > 0.6
- Credit score below 650
- Irregular payment history and unstable employment
- Low savings and limited asset base

These insights show that financial discipline and repayment stability are the most decisive factors in predicting loan default.

3 Model Performance & Selection

Both LDA and QDA models were trained and evaluated on an 80/20 train–test split after feature standardization. Results demonstrated perfect separation between good and bad borrowers, yielding the following metrics:

Table 1: Model Performance Metrics

Metric	LDA	QDA
Accuracy	1.00	1.00
Precision	1.00	1.00
Recall	1.00	1.00
AUC	1.00	1.00

In business terms, this means the model correctly identifies 100% of true defaulters (Recall) and does not mistakenly reject creditworthy customers (Precision = 1.00).

Since both models performed identically, LDA was selected for deployment due to its simplicity, interpretability, and robustness against overfitting. The top predictive features according to LDA coefficients were credit utilization, debt-to-income ratio, payment history score, and job stability.

4 Recommendation

LendSmart should deploy the LDA model as a decision-support tool in its loan approval process.

Rationale

- The model significantly reduces exposure to default risk by identifying high-risk borrowers early.
- Its linear structure allows for clear regulatory explanations and easy integration into current approval workflows.
- The selected variables are transparent and business-relevant, promoting trust and accountability.

Trade-off

While a small portion of low-risk applicants ($\approx 10\%$) may be misclassified as high-risk in real-world scenarios, this trade-off is acceptable given the expected reduction in default-related losses.

Next Steps

- Integrate the model into LendSmart's loan screening system and retrain semi-annually.
- Expand the dataset with 2025–2026 applications to validate long-term performance.
- Develop a dashboard for explainable AI to monitor model behavior and fairness.

5 Conclusion

This analysis demonstrates that LendSmart can predict loan default risk with near-perfect accuracy using interpretable statistical models. Deploying the LDA model will allow the company to minimize default losses, improve decision consistency, and strengthen customer confidence through data-driven risk management.

6 Link del video

Link del video `https://youtu.be/vYC9FNnuB6s`