

Executive Summary

Currently, LendSmart faces a profitability challenge as 28% of all approved loans end in default. This default rate is significantly above what management considers acceptable and well above the firm's risk tolerance. At this level of risk, every approval decision carries a material probability of loss, which erodes margins, reduces investor confidence, and pressures growth sustainability. The current approval process does not fully differentiate between genuinely high-risk applicants and those who are simply "unconventional but low-risk."

In this project, we analyzed LendSmart's historical applicant data and built a predictive model to more accurately identify high-risk applicants before making a lending decision. We specifically compared two established methods in credit-risk modelling: Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). Our goal was to determine which model provides the best balance between reducing costly defaults without unnecessarily rejecting creditworthy customers.

Key Findings

Our analysis using all available features confirmed that a core set of financial behavior indicators strongly drives default risk at LendSmart. Applicants with poor payment histories and unstable employment show higher default likelihood. In contrast, those with strong payment records and steady job histories are significantly less risky. High debt-to-income ratios and credit utilization emerged as clear red flags, whereas higher credit scores indicated safer borrowers. Secondary factors such as savings ratio and employment years also contributed modestly to risk differentiation.

In simple terms, the "danger profile" for LendSmart is: applicants who borrow aggressively relative to their income, already use most of their existing credit, and have inconsistent repayment behaviour. Importantly, these factors were far more predictive of default than demographic variables such as age, marital status, or education.

Model Performance & Recommendation

At the beginning of the analysis, both LDA and QDA models produced perfect results, reaching 100% accuracy. This indicated a case of data leakage, where some variables had

such high correlations with the target variable, that they revealed the target. After identifying this and rerunning the models without this leakage, they produced slightly more realistic results. Both models still performed strongly, achieving around 97% accuracy, but LDA slightly outperformed QDA in terms of precision, recall, and overall stability. LDA therefore can be seen as more reliable for distinguishing between defaulters and non-defaulters while minimizing false rejections of profitable applicants.

In business terms, this means that LDA allows LendSmart to better identify high-risk borrowers without unnecessarily turning away creditworthy clients. Therefore, LDA is recommended as the primary decision model for the company's loan approval process. Implementing this model will help reduce the current 28% default rate, improve portfolio quality, and protect profitability.