



PREDICTING LOAN
APPROVAL RISK USING
APPLICANTS DATA.

PROJECT OVERVIEW

This project seeks to analyse and determine factors that increase and decrease loan approval likelihood by building machine learning models to predict loan approval outcomes. Logistic regression and decision trees classifiers were implemented in this study.

PROBLEM STATEMENT

- ▶ Financial Institutions face significant risk when it comes to loan approvals for clients who may default. Approving a high risk loan may result in financial losses. This project aims to build a classification model that minimizes the risk of approving loans to applicants who may default.

OBJECTIVES

- ▶ 1. Predict loan approval likelihood based on applicant information.
- ▶ 2. Build a model that minimizes risk.
- ▶ 3. Provide lenders with insights that can help them make more informed decisions.

METHODOLOGY

WHAT DATA DID WE USE?

1. Applicant information (age, income, employment)
2. Loan details (amount, interest rate, purpose)
3. Credit history indicators
4. Past loan default history

The target variable was loan status: Approved(1) , Reject (0)

MODELS USED

We tested multiple predictive approaches to determine which one best identifies risky loans. These were:

1. Logistic Regression Model
2. Decision Trees Classifiers

Key Challenges in the Data

1. Class imbalance: Not all loans were approved equally.

- ▶ In our dataset, there were many more rejected loans than approved loans. If left unaddressed the model would learn to reject loans more often. To handle this, we made sure that both approved and rejected loans were fairly represented when training the model. We also adjusted the model so that it paid extra attention to the less common outcome.
- ▶ This approach helped the model learn patterns from both types of loans, rather than being biased toward the majority group.

2. Correlated features:

- ▶ Age is highly correlated with person employment experience and person credit history length.
- ▶ person employment experience is highly correlated with person credit history length.
- ▶ Employment experience was dropped.

LOGISTIC VS DECISION TREES

- ▶ Both models are good at identifying loans that should be approved.
- ▶ The logistic regression model had a higher recall 0.92 compared to decision trees 0.83. This means that it is more likely to detect risky loans compared to the Decision trees model. The trade-off is slightly lower precision, meaning some loans that could be approved might be incorrectly flagged as risky. This trade-off is acceptable when minimizing losses is the priority.
- ▶ We would prefer a system that minimizes risk. Ideally, the system should be more likely to flag a loan that would otherwise be approved as risky, rather than approving a loan that should be considered high-risk.
- ▶ Overall logistic regression is recommended for predictive performance.

FINDINGS

► The following improve loan approval likelihood.

1. Higher income
2. Higher interest rates: Applicants who accept loans with higher interest rates.
3. Loans that take up a larger share of income; applicants who commit a higher percent of their income to repaying loans
4. Applicants who rent their homes

These factors suggest that the applicant has sufficient cash flow to meet payment obligations.

FINDINGS

► The following decrease loan approval likelihood.

1. Larger loan amounts
2. Lower credit scores
3. Home ownership
4. Education loans
5. Medical loans
6. Personal loans
7. Venture-related loans
8. Previous loan defaults

These factors suggest that the applicant has a lower likelihood of having a loan approved

CONCLUSIONS

1. Many factors influence loan approval: Higher income, accepting higher interest rates, loans that take up a larger share of income and applicants who rent their homes increase the likelihood of loan approval **while** Larger loan amounts, lower credit scores, home ownership, education loans, medical loans, personal loans, venture-related loans, and previous loan defaults reduce the likelihood of loan approval.
2. Applicants who have defaulted on loans before pose the highest risk while applicants who commit a higher percent of their income to repaying loans pose the lowest risk.
3. Certain types of loans like venture, education, personal, and medical loans are less likely to be approved.

RECOMMENDATIONS

1. Use past default history as a primary risk filter and apply higher requirements for repeat defaulters.
2. Look at how much of an applicant's income would go toward repaying the loan, not just their total income, when deciding whether to approve it.
3. Treat different loan purposes differently, and apply stricter review for higher-risk type loans.
4. Factors that may appear minor at first glance, like home ownership status, still have an impact on loan approval outcomes and deserve consideration.