Mortgage Risk Modeling: Chase Case Study

Taniya Adhikari

# Overview

This case study describes a mortgage risk analysis for a Chase Bank a nation’s one of the largest banks. In 1996 after the mergers and acquisition of the company, Chase bank also acquired new risk of mortgage. Their numbers for mortgage holders were larger after the merging, and each mortgage had some micro-risk associated with it. The Chase Bank has multi-million dollars’ worth of mortgages. These micro-risks are small but can have a snowball effect if not paid attention to it. Usually, there are two types of risks associated with credit risks, when a mortgage holder does not pay back the loan and when a holder pays it early. These risks can be minimized with the use of predictive analytics. Decisions for mortgage are made based on predictive analytics results.

# Business Understanding

After the 1996 merging, Chase bank was facing a new risk with the large number of mortgages. Essentially, mortgage risk consists of 2 types of micro risks:

A: Customer fails to pay the mortgage payments.

B: Customer pays the whole mortgage early.

The risk A is a complete loss of the mortgage amount, whereas risk B displays a different type of risk. Although, risk B is paid early and the full amount is recovered, the total interest over the loan term is reduced thus reduces the income of the bank. For this case study we will focus on risk B. Chase’s mortgage wants to find which mortgage holders will prepay the mortgage within 90 days? In other words which customer is highly likely to prepay in next 90days? What are the factors about mortgage that tells this?

# Data Understanding

Data used for this study is Chase bank mortgage customer data and each mortgage consist of hundreds of variables some of the variables are mortgage amount, property value, property type, interest rate, borrower’s annual income, credit score, late payments, customer’s age, marital status, education, and employment history, etc.

# Data Preparation

For this case study data was carefully analyzed and arranged for decision tree model. The data was split into training and testing set. Because the decision tree algorithm used in the analysis could handle missing values, it is assumed that no missing values were imputed. The model itself picks the variables for prediction.

# Modeling and Modeling Interpretation

The data was sampled from large Chase mortgage portfolio and had training set of about 22,000 cases and 5,486 cases for test set. The model used was simple decision tree model to categorize cases by low and high-risk group as it keeps growing. The learning model had 39 segments (*Refer to Fig 1 for 10 segment decision tree model diagram*). For model evaluation, a single metric lift was used to measure the performance of the model. It determines how many more target customers model can identify compared to without model. Results indicate that the 20% of the cases are most high risk including 60% of all the estimated defectors. 60 percent multiples by 20 percent gives us 300 percent as many as without the model, thus lift at 39 segments is 3.0. Results from the testing set also indicate similar results (*Refer to Table 1 for results*).

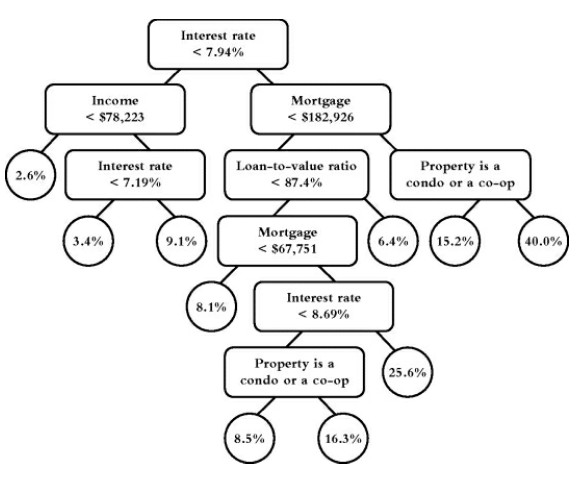


Fig 1: Decision Tree Model with 10 Segments

|  |  |  |
| --- | --- | --- |
| **Decision Tree** | **Lift of Training Set** | **Lift of Testing Set** |
| 4 Segments | 2.5 | 2.5 |
| 10 Segments | 2.8 | 2.8 |
| 39 Segments | 3.0 | 3.0 |

Table 1: Evaluation Metric Results

# Deployment

Chase needed multiple models for different categories. Chase wanted to use the results to estimate the expected future value of the mortgages to decide if it is a good idea to sell them to other banks. CART trees were developed for fixed rates versus variable-rate, for different mortgage terms and different stages during the period. After categorizing mortgages in different groups, each had decision tree. Each model employed its own variables in different ways. These models further generated millions of profits in the initial first year after the deployment.

# Summary and Conclusions

To predict whether the customer will likely pre-pay the mortgage early or not. This was a classification problem; thus, decision tree model was applied. Simple rules of yes and no were applied to variables to determine high risk or low risk groups among the cases. The model also generated quite favorable results. It identified 20% of the cases belong to the high-risk group. It predicted 74% of mortgage pre-payments beforehand.