# Bank Loan Default Analysis

## (Executive Summary Prepared for Bank Of America)

The banking industry is primarily funded through the interests earned on loans. As the number of defaulted loans increase, a bank may lose profits, lose money, and potentially go out of business. Because of this risk, it is critical that banks continually analyze the factors involved in defaulted loans and adjust their policies to minimize that risk. This analysis indicates which variables are most important for predicting loan default risk.

While this analysis attempted to use all variables, some were removed due to a linear relation between two predictor variables. When this occurs, the modeling process is unable to discern the impact of either variable. The bankruptcies variable conflicted with the Max Open Credit. Years of Credit History conflicted with Number of Open Accounts.

Bankruptcies was removed since very few loan applicants had bankruptcies. This means that it would have had negligible impact on the final model. Years of Credit History was removed since it details past credit health vs current credit health. While both are likely impactful, historic trends may not predict current or future financial health. It was felt that the current credit health would be more impactful to the risk of the loan.

The Loan Status variable, with binary values of “Fully Paid” and “Charged Off”, was used as the response variable to train the models. The remaining variables centered around financial health, level of equity, loan specifics, and additional risk factors.

|  |  |  |  |
| --- | --- | --- | --- |
| **Financial Health** | **Level of Equity** | **Loan Specifics** | **Additional Factors** |
| Credit Score | Home Ownership | Current Loan Amt | Tax Liens |
| Annual Income |  | Term | Years in Cur. Job |
| Monthly Debt |  | Purpose | No. Credit Problems |
| No. Open Accounts |  | Mo since delinquent |  |
| Cur. Credit Balance |  |  |  |
| Max. Open Credit |  |  |  |

The loan amounts varied between about $10,000 to and $790,000, and were relatively evenly divided between short and long term loans. Annual income varied between about $77,000 to $36,000,000, which most likely represents a business and not a personal account. The lowest credit score was 585. For most applicants, tax liens, bankruptcies, and credit problems were not issues. Overall, most numerical measures were evenly split between both categories of successful and unsuccessful loans. This illustrated that loan risk is not exclusive to any specific financial segment of the population.

In addition to the numerical values, categorical values involved the purpose of the loan, years of employment, home ownership, and the loan term of either short or long. While the purposes indicated by loan applicants varied, attempts to categorize them together did allow one of the models to improve its predictive ability. Those categories were grouped as:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Largest Expense** | **Major Expense** | **Medium Expense** | **Smaller Expense** | **Other Expense** |
| Purchase House | Small Business | Major Purch. | Moving | Other |
| Educ. Exp. | Business Loan | Med. Bills | Debt Cons. |  |
|  | Home Imp. | Buy a Car |  |  |
|  |  | Take a Trip |  |  |
|  |  | Vacation |  |  |
|  |  | Wedding |  |  |

The two types of models employed were Linear Regression and Neural Network Analysis. Both methods output the probability of a result, which can be translated into a binary classification. Both models can be adjusted over time to better fit new data if needed. In this analysis, the Neural Network model had the highest level of successfully predicting both successful and unsuccessful loans.

The Neural Network Model indicated that current loan amount was the major, positive indicator for successful loans. This suggests that applicants who tend to require larger loans are more likely to pay the loan off successfully. Other positive indicators were Annual Income, and Open Accounts. It is possible that people who are financially active may have the financial intelligence to repay a loan successfully. Some of the negative influences were requiring a loan for a major expense, the level of monthly debt a customer has, the length of their current employment, and their credit score.

Overall, the model indicated these five items were the most impactful.

