Credit Risk Assessment Using Data Mining Techniques

Approving and denying loan applications can be done more efficiently using the information our bank has at our disposal. By analyzing the data, we can discover patterns in our clientele. We can discover if there is a characteristic, or group of characteristics, that a client has that will predict if they will have a good or bad credit standing.

I was able to uncover several interesting patterns through analyzing information about 425 clients that our bank has done business with. I discovered that if a client is foreign and their credit history is categorized as “All paid” or “Bank Paid”, there is an 88.64% likelihood that they’d result in a bad credit standing. This pattern is shown in our previous data 44 times, with 39 of those clients with this pattern showing a bad credit standing. This probability rises to 94.44% if the client is also a woman.

Foreign & Credit History; All Paid or Bank Paid	44		
Foreign & Credit History; All Paid or Bank Paid => bad	39	0.0918	0.8864
Foreign & Credit History; All Paid or Bank Paid & Woman	18		
Foreign & Credit History; All Paid or Bank Paid => bad	17	0.0400	0.9444

I also discovered that if a client has no checking account and will be using the loan to finance either a used car, their education, furniture, or a small appliance, there is a 94.44%, 87.5%, 72%, and 86.67% likelihood they will have a good credit standing respectively. Considering each of these percentages only accounts for a small portion of the data (>=10%), grouping these purposes is beneficial as it allows us to account for over a fifth of clients (22.5%) while still retaining an estimate of 84.38% likelihood that they will have a good credit standing.

No Acct Checking & Purpose; Car Used, Education, Furniture,or Small Appliance	96		
No Acct Checking & Purpose; Car Used, Education, Furniture,or Small Appliance=> good	81	0.1906	0.8438

One more pattern I found was that clients who were between ages 18-24 and either had a low or low-medium amount of savings were 74% and 75% likely to have a bad credit standing respectively. By grouping these 2 characteristics, 15% of the overall client data is accounted for while still retaining an estimate of 74.24% likelihood that they'd result in bad credit standing. The probability for this pattern rises to 85.19% if the client also rents their housing.

Age 18-24 & Low or MedLow	66		
Age 18-24 & Low or MedLow => bad	49	0.1153	0.7424
Age 18-24 & Low or MedLow & Rent	27		
Age 18-24 & Low or MedLow & Rent => bad	23	0.0541	0.8519

Finding how characteristics affect the likelihood of a client’s credit standing can help our bank make less risky investments. We’ll be able to reduce the number of bad investments and boost the number of good investments. Continuing the search for patterns in our data and implementing the ones we have found will ensure profit growth and cost minimization.

Appendix

The data mining technique used is referred to as "Association Rule Analysis". This technique consists of evaluating series of if-then logical statements to find patterns. Using these patterns, predictions can be made when encountering the "if" portion of a logical statement.