Aaron Drexler

2/27/22

DSC 630

**Chase Mortgage Case Study**

For the case study, I decided to investigate Chase bank’s Mortgage structure. Soon after a megamerger in 1996 that rendered Chase Bank the nation’s largest, the bank’s home finance team recognized a new degree of risk. Their pool of mortgage holders had grown significantly. It was now composed of what had originally been six banks’ worth of mortgages: millions of them. Each one represented a bit of risk—a micro-risk. Chase’s mortgage portfolio faced risk factors amounting to hundreds of millions of dollars. Every day is like a day at the beach—each grain of sand is one of a million micro-risks. Once a mortgage application is stamped “low-risk” and approved, the risk management process has only just begun. The bank’s portfolio of open mortgages must be tended to like cows on a dairy farm. The reason? Risk lurks. Millions of mortgages await decisions as to which to sell to other banks, which to attempt to keep alive, and which to allow refinancing for at a lower interest rate.

**Data Understanding: Defining the Data**

To learn from data: the process isn’t nearly as complex as you might think. Start with a modest question: What’s the simplest way to begin distinguishing between high- and low-risk mortgages? What single factor about a mortgage is the most telling? If a mortgage’s interest rate is under 7.94 percent, then the risk of prepayment is 3.8 percent; otherwise, the risk is 19.2 percent. Drawn as a picture:

Diagram

Description automatically generated

What a difference! Based only on interest rate, we divide the pool of mortgages into two groups, one five times riskier than the other, with respect to the chances of prepayment (a customer making an unforeseen payoff of the entire debt, thereby denying the bank future earnings from interest payments). This discovery is valuable, even if not entirely surprising. Homeowners paying a higher interest rate are more inclined to refinance or sell than those paying a lower rate. If this was already suspected, it’s now confirmed empirically, and the effect is precisely quantified.

**Data Preparation**

This learning method, called decision trees, isn’t the only way to create a predictive model, but it’s consistently voted as the most or second most popular by practitioners, due to its balance of relative simplicity with effectiveness. It doesn’t always deliver the most precise predictive models, but since the models are easier on the eyes than impenetrable mathematical formulas, it’s a great place to start, not only for learning about PA, but at the outset of almost any project that’s applying PA.

**Diagram

Description automatically generated**

You can see the tree is growing downward. As any computer scientist will tell you, trees are upside down and the root is on the top (but if you prefer, you may turn this book upside down). As shown, the mortgage holder’s income is very telling of risk. The lower-left leaf (end point of the tree) labeled “Segment 1” corresponds with a subgroup of mortgage holders for whom the interest rate is under 7.94 percent and income is under $78,223. So far, this is the lowest-risk group identified, with only a 2.6 percent chance of prepayment.

Before modeling, data must be properly arranged in order to access its predictive potential. Like preparing crude oil, it takes a concerted effort to prepare this digital resource as learning data (aka training data). This involves organizing the data so two-time frames are juxtaposed: (1) stuff we knew in the past, and (2) the outcome we’d like to predict, which we came to find out later. It’s all in the past—history from which to learn—but pairing and relating these two distinct points in time is an essential mechanical step, a prerequisite that makes learning to predict possible. This data preparation phase can be quite tedious, an involved hands-on technical process often more cumbersome than anticipated, but it’s a small price to pay.

**Modeling**

For the model, I used decision trees. To use a decision tree to predict for an individual, you start at the top (the root) and answer yes/no questions to arrive at a leaf. The leaf indicates the model’s predictive output for that individual. For example, beginning at the top, if your interest rate is not less than 7.94 percent, proceed to the right. Then, if your mortgage is under $182,926, take a left. You end up in a leaf that says, based on these two factors, the risk that you will prepay is 13.9 percent. Here’s the tree on Chase mortgage data after several more learning steps (this depiction has less annotation—per convention, go left for “yes” and right for “no”):

Diagram

Description automatically generated

**Revisit Business Understanding**

Learning has now discovered 10 distinct segments (tree leaves), with risk levels ranging from 2.6 percent all the way up to 40 percent. This wide variety means something is working. The process has successfully found groups that differ greatly from one another in the likelihood the thing being predicted, prepayment, will happen. Thus, it has learned how to rank by future probabilities. For example, Sally Smithers, the example mortgage customer from earlier in this chapter, starts at the top (tree root) and answers yes/no questions:

Q: Interest rate , 7.94 percent? A: No, go right.  
Q: Mortgage , $182,926?  
A: Yes, go left.

Q: Loan-to-value ratio , 87.4 percent?  
A: Yes, go left (the loan is less than 87.4 percent of the property value).

Q: Mortgage , $67,751?  
A: No, go right.  
Q: Interest rate , 8.69 percent?  
A: No, go right.

Thus, Sally comes to a landing in the segment with a 25.6 percent propensity. The average risk overall is 9.4 percent, so this tells us there is a relatively high chance she will prepay her mortgage. **Business rules** are found along every path from root to leaf. For example, following the path Sally took, we derive a rule that applies to Sally as well as many other homeowners like her:

Graphical user interface, text, application

Description automatically generated

**Deployment**

But Chase’s plans put a new twist on the value of predicting churn. The bank intended to use the predictive scores to estimate the expected future value of individual mortgages in order to decide whether it would be a good move to sell them to other banks. Banks buy and sell mortgages at will. At any time, a mortgage could be sold based on its current market price, given the profile of the mortgage. But the market at large didn’t have access to these predictive models, so Chase held a strong advantage. It could estimate the future value of a mortgage based on the predicted chance of prepayment. In a true manifestation of prediction’s power, Chase could calculate whether selling a mortgage was likely to earn more than holding on to it. Each decision could be driven with prediction.

With data covering millions of mortgages, the amount available for analysis far surpassed the training set of about 22,000 cases employed to build the example decision trees depicted in this chapter. Plus, for each mortgage, there were in fact hundreds of predictor variables detailing its ins and outs, including summaries of the entire history of payments, home neighborhood data, and other information about the individual consumer.

The Chase project required numerous models, each specialized for a different cate- gory of mortgage. CART trees were grown separately for fixed-rate versus variable-rate mortgages, for mortgages of varying terms, and at different stages of tenure. After grouping the mortgages accordingly, a separate decision tree was generated for each group. Since each tree addressed a different type of situation, the trees varied considerably, employing their own group of variables in divergent ways.

**Summary and Conclusions**

The undertaking was an acclaimed success. People close to the project at Chase reported that the predictive models generated millions of dollars of additional profit during the first year of deployment. The models correctly identified 74 percent of mortgage prepayments before they took place and drove the management of mortgage portfolios successfully. Soon after the project launch, in 2000, Chase achieved yet another mammoth milestone to expand. It managed to buy JPMorgan, thus becoming JPMorgan Chase, now the largest U.S. bank by assets.

**Citations**

Siegel, Eric. (2016). *Predictive analytics: the power to predict who will click, buy, lie, or die*. New Jersey: John Wiley & Sons.