Case Study: Prediction of Mortgage Risk

Pushkar Chougule

Overview

The mortgage industry, measured in the trillions of dollars, serves as the financial cornerstone of homeownership. Mortgage lending forms a major portion of a Bank's lending business, especially with average loan application size to the tune of $250,000 and some of them touching $500,000 or even more. Once pool of mortgage holders grows significantly, there is always the risk that some of its customers may prepay the mortgage due to refinancing with competing bank or by selling the house. Prepayment is a loss to the bank as the principal is returned to the bank prematurely and thus the bank misses out on the mortgage's planned future interest income. Especially in the case of falling interest rates scenario, the prepayment risks are higher. And if the bank has many such mortgages closing due to prepayments, then banks are losing out their regular income in the form of loan interest. So the banks are constantly on lookout for identifying such "Prepayment risk" mortgages and selling them off or transferring them to other banks to gain profits and reduce the income risk. One such case is with the Chase Bank, which we will be looking at as a part of this case study.

Chase Bank is one of the Top 10 mortgage lenders in the country. This case study describes a predictive analytics project to evaluate the mortgages as to whether the risk of prepayment exists or not, based on variety of factors. They decided to seek help from Dan Steinberg in achieving this goal for risk assessment of the mortgage portfolio and decide upon the risk levels of the mortgages within their portfolio. By performing this risk assessment, the could either retain such mortgages with lowest risk of prepayment and sell the mortgages with higher prepayment risks to the other banks.

Business Understanding

Chase's mortgage portfolio faced risk factors amounting to hundreds of millions of dollars. For a bank, one single mortgage customer may not be a big risk i.e. it is a micro risk. However, multiple micro risks may add up and will convert into huge financial risk. Hence, bank tries to mitigate some of the micro risks by selling some of the existing mortgages to other banks or credit institutions, or to maintain with the bank, or allow refinancing for at a lower interest rate. The decision on selling the existing mortgages to other banks is to be made preferably on the mortgages which pose risk of prepayment i.e. the ones which may not be profitable in the longer run because of early payments of principal amounts which in turn results in loss of interest income from the customers.

As a part of the case study to solve this problem, Chase bank decided to go with Predictive Analytics technique to help identify the loans with potential prepayment risks and make decisions on either selling those loans to other banks based on the current market values and whether selling such loans will be more profitable or whether retaining them and making attractive offers would be more beneficial from the bank's perspective.

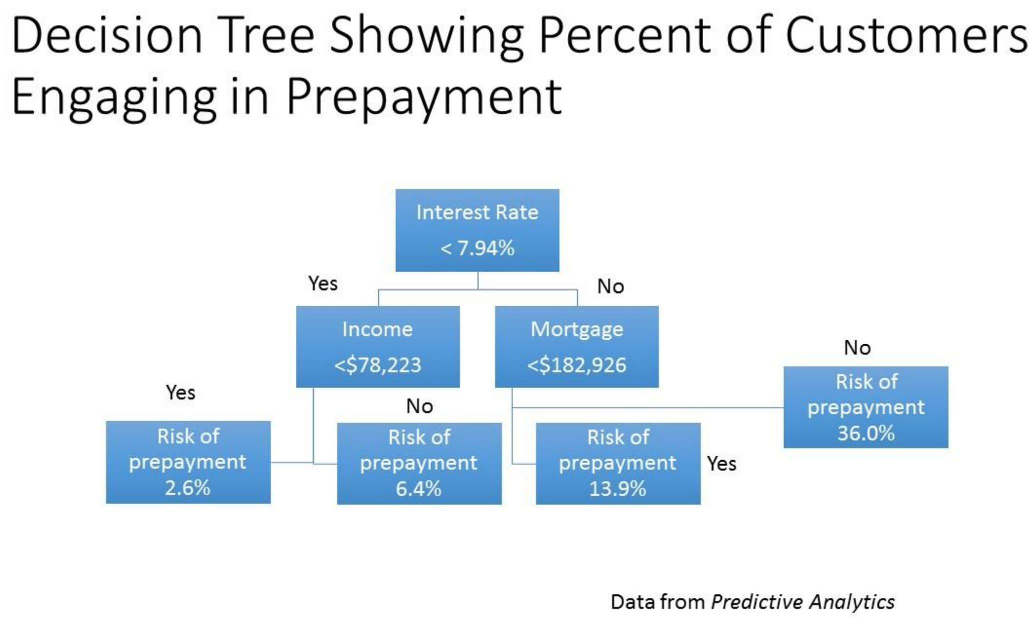
The objective was to build a model that helps identify the mortgages with higher prepayment risks in the next 90 days and sell such loans to other banks based on current market value and strike profitable deal so that selling the mortgage would earn more profit rather than retaining which would be foreclosed and result in the loss of interest income. This will help bank to add to its profitability and income.

Data Understanding/Data Preparation

The dataset consisted of total of 27,302 cases / records. We will be using 21,816 cases (80%) in the training dataset and remaining 5,486 cases (20%) of the cases were kept aside for testing purpose. Some of the important features to consider in the data understanding and preparation are the interest rates, mortgage amount, loan-to-value ratio, customer annual income, type of property are some of the important factors to consider for the identification of loans with higher prepayment risk. Interest rate is the first important factor and it can help drive the next stages / branches in the modeling phase and when the interest rates are higher, customers choose to refinance the mortgages at lower rate. Customer annual income is another factor of importance – customers with lower income ranges tend to continue with the mortgage payments longer as compared to customers with higher annual income. Other factors like mortgage amounts and loan-to-value ratio are important as well since higher value mortgages or the mortgages with higher loan-to-value ratio tend to continue for longer rather than getting foreclosed with prepayment. Considering all these factors, we need to ensure that required features are maintained within the dataset. We will look through the dataset and ensure missing values and the outliers are cleaned from the dataset appropriately.

Modeling / Evaluation

The data will first be split into training and testing data set. The training set will be used to train / build the decision tree model. Decision Tree model is used when we are required to break down complex data into more manageable sections so that ultimately it can aid in the decision making. We will use 80% of the dataset as training set and the remainder 20% set for testing purposes so as to ensure the selected metrics, precision and accuracy are well within expected ranges. We can evaluate how many of the predicted outcomes result in accurate classification as customer prepaying the loan vs. continuing with the mortgage for longer duration. We can calculate the positive and negative prediction results ratios for correct (true) and incorrect (false) result counts and ensure the accuracy and precision level of the model is above 70%, since this is the first such predictive analytics endeavor of this kind for the company. We can use the k-fold (e.g. 10-fold, 11-fold etc.) cross validation technique to try to fit the model. Once the model fitting is performed, we can evaluate the model against the test data set and validate whether the model predicted the prepayment risks correctly in terms of predicted outcomes vs. actual cases. If the results are within the expected accuracy levels, then we will continue to use the model or train / evaluate / fine-tune the model iteratively until the predicted results meet the expectations of accuracy over 70% levels. The criteria for efficient model building is with help of evaluation metrics like precision, accuracy, lift, recall etc. If criteria were met on these various parameters, model can be deployed to production environment and monitored for performance and continued improvement opportunities. We will look to branch the tree out to optimal number of segments to achieve lift of around 3.



Deployment

Prediction and management of the mortgages based prepayment risks does come with its own challenges. Identifying which customers will leave and which ones will continue to stay on correctly is a big challenge in itself and depends on variety of considerations like how big are the mortgage amounts, and respective loan-to-value ratio? What are the interest rates and do we expect them to go down or up? What is the customer income for the corresponding mortgages and, the type of property etc.? This is the reason why carefully deploying the Decision Trees model is so important in the mortgage value estimation process. It helps identify which mortgage holders will prepay within the next 90 days and then the mortgages are valued accordingly in order to decide whether to sell them to other banks.

Conclusion

The decision tree model correctly identifies 74 percentage of mortgage prepayments before they took place, and drove the management of mortgage portfolios successfully. It helped tremendously to either retain the mortgages or to price the mortgages based on its current market price at point of time. Thus the mortgages could be sold at any point of time based on the profile of the mortgage. Chase could estimate the future value of a mortgage based on the predicted chance of prepayment. Hence calculations whether selling a mortgage was likely to earn more than holding on to it could drive the profitable strategy. Chase has had an advantage of the predictive model power as compared with the peer / competitor banks. The use of the predictive models generated millions of dollars of additional profit just during the initial year of deployment and delivered the great results with the faith put in the predictive modeling strategy by the bank.

References

1) Abbott, D. (2014). *Applied Predictive Analytics – Principles and Techniques for the Professional Data Analyst.* John Wiley & Sons.
2) XIAO, Z (2021). Interest rate risk, prepayment risk and bank's securitization of mortgages. Retrieved on Oct 5, 2021 from https://blogs.worldbank.org/allaboutfinance/interest-rate-risk-prepayment-risk-and-banks-securitization-mortgages
3) Siegel, E. (2016). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.* John Wiley & Sons.
4) Thakur, M. (2021). *Prepayment Risk.* Retrieved on Oct 4, 2021 from https://www.wallstreetmojo.com/prepayment-risks/.
5) Vickers, M. (2016). A REVIEW OF ERIC SIEGEL'S PREDICTIVE ANALYTICS. Retrieved on Oct 6, 2021 from https://foresightr.com/2016/08/09/a-review-of-eric-siegels-predictive-analytics-part-one/