



Enhanced Vision

With predictive analytics, mortgage professionals can see further into the cycle of a loan to determine the best course of action to take.

By Jonathan B. Glowacki and Makho Mashoba

Financial professionals in the digital age are able to do something that was once impossible: see into the future. Using predictive analytics, a blend of statistical analysis and computer science, today's financial professionals solve problems, improve processes, and understand trends in consumer behavior. For example, they can process credit card behavior to identify potential fraud before approving a charge, reduce churn by targeting unhappy customers, and identify customers for new products and services based on what they have purchased in the past. Predictive analytics is not complicated; it just requires data and the right tools.

The mortgage industry has tremendous amounts of data. There is housing data, mortgage origination data, and servicing data, to name a few. The conditions are ripe for using predictive analytics.

To illustrate the benefits of predictive analytics, we'll examine one case example from publicly available Government National Mortgage Association (GNMA or Ginnie Mae) data in the servicing space.

Once a borrower is 90 days past due on his or her mortgage payment, the approved Ginnie Mae issuer and loan servicer has the option to buy out the nonperforming loan at par—100 percent of the unpaid principal balance (UPB)—from a GNMA

pool. Then, according to the GNMA mortgage backed securities (MBS) issuer handbook, the issuer can employ loss mitigation tools on those loans to cure the mortgage back to performing status. If the loans re-perform, the issuer can re-securitize the loans into a new issue pool. If the price is greater than par, or 100 percent of UPB, the issuer can get an immediate gain on the sale of the loans, earning revenue.

In many cases, the issuer has to buy out the loan to pursue loss mitigation strategies that change the terms of the mortgage (such as term extension or interest rate reduction). However, the handbook does not require a partial claim to alter the terms of the mortgage. In a rising interest rate environment, partial claims are likely to become a more prevalent loss mitigation strategy, leaving the issuer with more choices on whether or not to buy out nonperforming loans for a security.

There is an opportunity to buy out loans in the early stages of delinquency if the issuer expects them to re-perform. The opportunity can be profitable if the issuer can effectively identify which loans to buy out and which loans to move through the foreclosure process. In the latter cases, issuers need to evaluate the cost of interest advances and property maintenance expenses. Also, GNMA issuers must be mindful of the delinquency ratios on pools that

are monitored by GNMA. The delinquency ratio is the fraction of the loans in the issuer's GNMA portfolio that are either in foreclosure or 90 or more days delinquent. Buyouts can be an effective way to manage delinquency ratios. These metrics in some cases incentivize the issuer to pursue early buyouts to reduce those ratios.

To Buy Out or Not to Buy Out

So how can predictive analytics help GNMA servicers make a decision on whether or not to buy out a given mortgage? There is ample data on GNMA loan performance, and the data is available for download from GNMA's website. All that remains is to predict which loans will turn around from nonperforming to performing. There are a variety of techniques that can be used for this analysis, including a relatively new algorithm called XGBoost.

XGBoost has been drawing a lot of interest in the predictive analytics community, winning several international data competitions. The XGBoost algorithm works by systematically iterating through possible predictive models while reducing unexplained error at each step. At the end, the result is an estimated value with low error and outstanding predictive ability.

To perform the analysis, the GNMA data was downloaded and modified to extract a list of all

loans that were 90 days delinquent and received a loan modification. The outcome of the loan modification (i.e., re-perform or re-default) was added to the data as a binary indicator (0 or 1). Several economic variables were also included. The XGBoost algorithm was then applied to the data to create a model that predicted which loans would re-default following modification. If an issuer can effectively evaluate which loans will re-perform, then the issuer can target those loans to buy out from the pool and perform loss mitigation. Then, if the loans cure, the issuer can re-securitize them into new issue pools.

After running XGBoost on the performance data, the tool revealed insights that weren't accessible with more traditional techniques such as logistic regression. The graph in Figure 1 depicts the importance of various features for the GNMA algorithm. Figure 1 highlights the most predictive features on the top, with the less predictive features on the bottom. (The features here are shown in isolation, but the algorithm does take into consideration correlations between variables.)

Home price appreciation from loan origination to first delinquency was the most important feature for predicting loan performance after modification. In addition, changes in the local housing affordability index and unemployment rate had modest effects, while variables like number of





borrowers and unpaid balance of the loan weren't as powerful for forecasting future defaults among these loans. In other words, by analyzing the data, we are able to deduce that economic variables are far more important than borrower characteristics in evaluating the performance of mortgages after modification.

It is possible to benchmark these results against performance on holdout test data. After splitting the data 80-20, an analyst can train an XGBoost model on 80 percent of the data. The analyst can then compare the actual default rates against the XGBoost predictions for the remaining "unseen" portion of the data. Using a standard receiver operating characteristic (ROC) curve, it is possible to evaluate and compare model performance.

An ROC curve works by plotting the true positive rate (how many predictions the model got correct) against the true negative rate (how many predictions the model estimated to occur that did not). The technique is often used for optimal model selection. The closer the curve is to a 45-degree line, the worse it performs. Here we see, in Figure 2, that the specificity (or the true positive rate), as well as the sensitivity (the true negative rate), are nearly ideal. The XGBoost model is nearly optimal.

Compare these results with the ROC curve of a traditional scoring algorithm such as logistic regression, shown in Figure 3.

The predictions from the traditional algorithm aren't as accurate as those of the XGBoost model. The

"Considering the growing amounts of data available, the mortgage industry should pay attention to predictive analytics tools."

ROC curve from the traditional algorithm is much closer to the diagonal, meaning that the true positive and true negative rates aren't performing as well. XGBoost has an advantage over the traditional algorithm in this holdout data set.

The Future Depends on Predictive Analytics

The XGBoost model, like similar algorithms, is easy to implement. Once the mechanics of the technique are understood, and the parameters are tuned correctly, the model can be turned on a data set to produce accompanying predictions. The model can be updated continuously each month based on new data feeds. Pointing an XGBoost program toward a new data set and running it again is virtually all that is needed to refresh the results. It is also possible to retune the parameters for the update to further enhance the effects.

A use case of this type of model would be to pursue early buyouts for mortgages that have a high probability of re-performing and potentially not pursue early buyouts for mortgages that have a low probability of re-performing, as long as this policy is consistent with GNMA servicing guidelines.

This same technique can be used on a variety of data for alternative purposes. Predictive analytics can capture predictive power from internal data, whether that involves established and go-to data sets or whether that involves bringing together data from across an organization to make predictions. Predictive analytics can also help a firm leverage industry data and other outside sources to forecast trends or improve decisions. This case is a concrete example of how using the tool should result in higher return on investment on GNMA early buyouts.

Considering the growing amounts of data available, the

mortgage industry should pay attention to predictive analytics tools. Investing in the technology has proven to generate significant returns. GNMA issuers is just one group to which predictive analytics can be applied. Predictive analytics can be applied to many other techniques and tools to increase efficiencies within the mortgage industry. The future depends on it. **M**



JONATHAN B. GLOWACKI, FSA, CERA, MAAA is a principal and consulting actuary with Milliman, and a fellow and chartered enterprise risk analyst through the Society of Actuaries.

Glowacki's areas of expertise include mortgage scoring, structured finance, model validation, and loan-loss reserving. He has provided quantitative and strategic consulting services on topics including residential mortgage collateral, credit insurance, debt protection products, reverse mortgages, mortgage insurance, financial guaranty insurance, deposit insurance, commercial mortgage-backed securities, and financial institution credit risk.



MAKHO MASHOBA is a financial analyst with Milliman. He joined the firm in 2016, specializing in econometrics and the statistical analysis of consumer behavior.

Mashoba has worked on projects involving model validation as well as borrower risk evaluation and mortgage industry trend analysis.

Figure 1: XGBoost ROC Curve

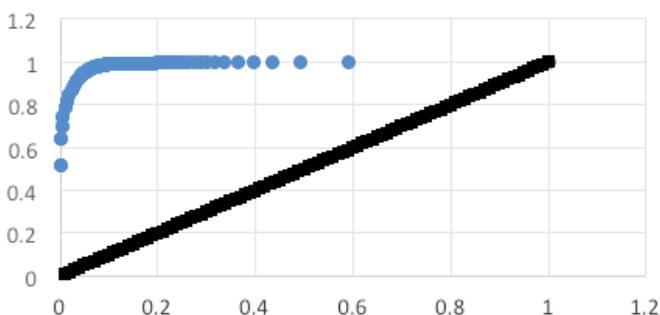


Figure 2: Logistic ROC Curve

