



Creating a Recession-Proof Collections Model

How Experian Used Machine Learning to Challenge Traditional Collections Modeling



Executive Summary

By using the most accurate and up-to-date collection models, businesses can increase their ability to collect debts during a recession. Experian® recently created a collection model, benchmarking it against three machine learning methods: gradient-boosted trees, random forests and neural networks. This ensures the model provides the best performance regardless of the final modeling technique. It also allows businesses to focus their efforts on customers more likely to repay their loans.

Problem Statement

The current economy is strong, but experts see some signs of volatility. Inevitably, we will go through another recession, in either the near or distant future. One of the keys to a company surviving a recession is its ability to collect outstanding invoices and debts from customers. During a recession, many customers can't pay their bills because of the recession's impact on their own financial situation. Lenders can decrease the risk of nonpayment with first-party collection or third-party debt buyers can prioritize the collection efforts by using a collection model.

Background

Businesses using an accurate and intelligent collection model can significantly increase their ability to collect debts. Many customers will pay their debts in response to simple collection strategies, such as calls, emails and letters. However, businesses only have limited time and need to know which customers their collection team should focus on.

When a business spends significant time working with a customer who isn't going to repay a debt regardless of the offers or efforts, then the business loses out on revenue from customers who very likely will pay with simple outreach. Also, many customers will pay their debts given time, regardless of the collection efforts. By knowing which customers will self-cure, your team can focus on the customers who need focused efforts to pay their debts.

Businesses also should use collection optimization strategies, such as PowerCurve® Collections, to determine which collection action is most likely to be effective for each customer. During the optimization process, businesses can set up a parameter, such as the minimum that needs to be collected for each customer or the maximum number of calls to make each month. While contact optimization takes place outside the model, the score drives the collection activity dependent on the strategies the collection agency puts in place for their agents.

Many previous collection models were built using pre-2008 recession data, meaning they don't include more recently added data or repayment behavior changes after the recession. Now there is new information available that can increase the collection model's effectiveness, such as recent collection data and new data sources, including SBFE Data™. Additionally, the market now accepts new technology and platforms with improved capabilities. To prepare for a future recession, businesses need a collection model built with updated data and the most current methodologies. Using our data and extensive experience in the collection industry, we've built a new model to help businesses with their collection efforts.

Over the past few years, we've added new data sources, including our own consortium data, and we've become an SBFE Certified Vendor. As our core products have improved and expanded, we've built new scores and models based on our core data to create a more predictive model. While previous products are effective in terms of performance, using more recent data provides more predictive results because it more accurately reflects the current market.

Experian compared current machine learning technologies with more recent technologies to create predictive services, a platform that combines the latest technology and data. Because the new platform allows for the deployment of machine learning techniques, you can create a process, review inputs and outputs, and target particular behaviors.

Solution

Our new collection model prioritizes businesses that are more likely to pay a collection amount. There is no offer of credit with the new model, which means businesses don't need to use adverse actions on the back end of the process. Because of regulations, this makes for a much simpler model deployment and governance process.

Financial institutions (first party) and collection agencies (third party) will have their portfolios scored to get a ranking of the specific debts they're attempting to collect on. For input, the model uses the amount to be collected and business credit behavior provided by Experian. The model returns a score from 300 to 850, with the higher values indicating the business is more likely to pay on the collection.

An important component of model development is creating a relevant development data set. We have access to third-party collection data and leveraged this data to determine how likely a business is to pay on its collection. We developed sample data from 2016 to 2018 to create and test the model. We also tested an out-of-time period. The resulting model is a segmented logistic regression model, but it's benchmarked against different machine learning modeling techniques.

Comparing machine learning models against traditional approach

Experian started developing the model by using the development data with a traditional logistic approach with multiple segments. With this model, segments are created from the population and then compared. By evaluating the segments, the model detects when a segment or a portion of a segment acts differently from the rest of the population. This method typically is very hand-tailored, meaning that an analyst works directly with the algorithm to refine the model for the best results.

But we wanted to ensure the new model took advantage of the current technology, especially machine learning capabilities. In recent years, the market accepted the new methods, and businesses are using these models for predictive analytics. We built models using the three approaches outlined below and compared results with the traditional logistic approach to ensure the new methods were at minimum on par.

We reviewed the expected dollar amount collected and the Kolmogorov-Smirnov test for the machine learning methods compared with the traditional model. We created a traditional model that leverages the transparency of traditional techniques and updated data, which performs comparably to newer machine learning techniques.

We benchmarked against the following types of machine learning collection models:

- **Gradient-boosted trees** — This model starts with one decision tree and identifies incorrectly predicted records. Next, the model creates a new decision tree that predicts all the records by focusing on those the model got wrong and attempting to learn from its mistakes. This continues until a stopping criterion occurs, such as all records being classified correctly. During scoring time, the trees all vote on the unseen record, with earlier trees having more weight. While the model is running, it's continually cleaning up and improving for the next tree. Gradient-boosted trees work similarly to a child learning colors. The child's teacher may tell them they got five out of 10 wrong, and they're directed to focus on the colors they got wrong to refine their definition of a color.
- **Random forest** — A random forest model seeks to create multiple decision trees where each has a slightly different view of the data. The model learns multiple trees by resampling from the development data and choosing a large sample of features. The goal is to have a forest of trees that vote on the best output by sampling from disparate voters. As a result, the trees become experts on the topic and then vote on different outputs, with the majority vote winning. The random forest model works similarly to a group of people trying to decide an outcome. Instead of relying on a single person's vote, they use a consensus to provide a more accurate result.
- **Neural network** — To understand a neural network, think about stimulating the brain and neurons to create a feedback loop in which a neuron's output becomes the input to another neuron. This model takes input — which in this case is the credit behavior inputs and the placement amounts. It then weights the importance of each piece. By comparing these inputs, the model predicts the outcome. If the result isn't correct, then the model goes back and readjusts the weighting of the components. For example, you tell a child a color is red, and then they begin identifying the color as red. If they incorrectly identify the color, their brain takes the feedback and adjusts the connections between the neurons.

Overcoming challenges of the machine learning model

Machine learning methodologies are slowly being accepted in the market and gaining momentum. One of the reasons for the slower adoption is that the approach is still overcoming some challenges:

- **Lack of transparency** — With the logistical approach, businesses and regulators can better understand the features and their relationship to the predicted outcome. Machine learning models often are more complex because they include large numbers of trees and contain the most features with various weights. The results often aren't intuitive to humans and can be challenging for people to accept. However, once businesses see the results are accurate, many begin to trust the results from machine learning models.
- **Stability** — Because machine learning is very focused on current data and trends, there is some concern about the model's longevity. Additionally, the fact the model is less transparent adds to the concern about its stability. However, the models we created mimic the data they see and take market changes into account to increase the model's stability and longevity.
- **Overfitting** — When you create a model that's very specific to the data used to train and develop, there is a risk of overfitting. This means the model may not accurately predict collections in the future. By using a wide range of data during development, a validation set and an out-of-time sample, Experian reduced the risk of overfitting when developing the new collection model.

Challenges are a natural and expected part of using new technology and processes. As the technology matures, many of these issues will level out or be resolved. Businesses can reduce the impact of these challenges by working with a trusted and experienced solutions provider.

Conclusion

In summary, collections play a very important role in a company's future, especially during challenging financial times, such as a recession. Companies should evaluate their collection methods and models before a recession hits so they are well-prepared with the right processes and technology. Once a recession occurs, it can be challenging to correct the collections issue midstream, especially with other business pressures. Now is the time to get ready for the next recession.

However, your first step is knowing which customers to focus your limited time and financial resources on for collections. By using Experian's new collection model, you can reduce the time spent with the wrong customers, so employees can focus on working with the right customers to efficiently collect payment. Since Experian benchmarked the collection score against state-of-the-art machine learning methods, users can be confident the collection score will provide top performance. If you update your collection processes now with the most accurate and modern models, you'll have the tools in place to increase the odds of being recession-proof.

About Experian's Business Information Services

Experian's Business Information Services is a leader in providing data and predictive insights to organizations, helping them mitigate risk and improve profitability. The company's business database provides comprehensive, third-party verified information on 99.9 percent of all U.S. companies, as well as millions of companies worldwide. We provide market leading tools that assist clients of all sizes in making real-time decisions, processing new applications, managing customer relationships and collecting on delinquent accounts.

For a free consultation, contact your local Experian sales representative, call 1 800 520 1221 or visit www.experian.com/biq.

