

# Machine learning for customer risk ratings

June 2017



## **Analytical contacts**

**Manish Kumar, PhD**  
Director, Risk & Analytics  
Manish.kumar@crisil.com

**Alan L Paris, CAMS**  
Director, Risk & Analytics  
alan.paris@crisil.com

## Financial crime and compliance analytics

Financial institutions are mandated to conduct anti-money laundering (AML) risk assessments to determine the overall risk rating of their customers. The assessment and subsequent rating are derived using a risk rating methodology based on industry standards and rules defined by regulatory bodies. Customer risk rating is a typical multi-class problem because financial institutions have two or more categories of ratings. Most firms highlight the importance of judgment/qualitative factors in determining the risk rating of customers. However, in practice, a mathematical model using quantitative variables/attributes plays an important role in determining customer risk ratings because the process is convenient and consumes less time. This article discusses the application of machine learning for customer risk ratings.

### Objective

This document proposes a framework/methodology based on machine learning approach to establish the risk rating score (normally a low, medium or high score) of customers using various drivers/attributes as input variables. The methodology is used for AML risk assessment and rating during the on-boarding phase, and the procedure will be performed on an ongoing basis throughout the period of the customer's association with the financial institution.

### Data and variables

For the analysis, granular customer-level data to determine the risk ratings is required. Also needed are independent variables (regressor) for a relatively longer period, covering an entire economic cycle. Necessary discussions with line of business and senior stakeholders to shortlist the initial set of potential drivers and independent variables also need to be considered for the model building exercise.

To identify the variables that contribute the most to the assigned customer risk ratings of the sample, the procedure is to follow correlations, pre-filtering, and forward/backward, and stepwise, variable selection approaches.

Some key variables considered for model development are:

- Customer's name and country assessment
- Type of customer and customer's industry
- Customer's source of assets/wealth
- Customer's reputation
- Anticipated account activities and purpose of the account
- Account's beneficial owners
- Known associates and network
- Customer behaviour and transaction history

## **Methodology: Champion/challenger approach**

Customer risk rating is a typical multi-class problem because financial institutions have two or more categories of ratings. Most firms highlight the importance of judgment/qualitative factors in determining the risk rating of customers. However, in practice, a mathematical model using quantitative variables/attributes plays an important role because the process is convenient and consumes less time.

Several statistical and machine learning techniques have been applied for such rating predictions. Among them, artificial neural networks (ANN) have been widely used because of their broad applicability to many business problems such as fraud detection, bond ratings, firm ratings, application and behavioural scorecard, default prediction, and financial time-series prediction. Of late, the Random Forest methodology has also become a popular choice for addressing classification and prediction problems because of its robustness and high accuracy.

Typically, classical ordered logit and probit models are used as a challenger/benchmark to evaluate the performance of the champion model built using methodologies such as ANN and Random Forest.

Hence, we propose machine learning models, namely ANN and Random Forest, to handle multi-class classification of customer risk ratings. Further, the results of these models will be benchmarked against the ordered logit/probit model.

### **Performance metrics**

To systematically evaluate the robustness of the proposed model, use the cross-validation technique in addition to in-sample and out-of-sample testing. The performance of the machine learning-based approach and classical ordered logit/probit model will be evaluated using Hit Ratio, Confusion Matrix, Mean Square Error, and Population Stability Index.

### **Broader impact of the proposed methodology**

The evaluation of customer risk ratings of banks and other financial organizations is very challenging due to the opaqueness of the banking sector and high variability in the sector's customer risk. Risk evaluation is, however, extremely important, given the growing number of regulations, policies, and penalties. The magnitude of losses in terms of goodwill, reputation, and fines, and its cascading effects on the financial system and economy, have enormous consequences.

The proposed customer risk rating approach can be applied to develop an accurate and objective customer risk rating model. The model would enable financial institutions to rate their customers on a proactive basis at a desired frequency and respond to the needs of developing internal customer risk rating systems.

The proposed model's role as an early warning system is extremely critical in view of

- I. the financial costs incurred by banks in case of policy and regulatory violations
- II. continued failure of banks in identifying and deterring money laundering activities over the last 30 years
- III. the risk of spill-over across the financial system and economy (from terror financing activities, drug syndicates, and child trafficking).

## **About CRISIL Limited**

CRISIL is a global, agile and innovative analytics company driven by its mission of making markets function better. We are India's foremost provider of ratings, data, research, analytics and solutions. A strong track record of growth, culture of innovation and global footprint sets us apart. We have delivered independent opinions, actionable insights, and efficient solutions to over 100,000 customers.

We are majority owned by S&P Global Inc, a leading provider of transparent and independent ratings, benchmarks, analytics and data to the capital and commodity markets worldwide.

## **About CRISIL Global Research & Analytics (GR&A)**

CRISIL Global Research & Analytics (GR&A) is the world's largest and top-ranked provider of high-end research, risk and analytics services. We are the world's largest provider of equity and fixed-income research support to banks and buy-side firms. We are also the foremost provider of end-to-end risk and analytics services that include quantitative support, front and middle office support, and regulatory and business process change management support to trading, risk management, regulatory and CFO functions at world's leading financial institutions. We also provide extensive support to banks in financial crime and compliance analytics. We are leaders in research support, and risk and analytics support, providing it to more than 75 global banks, 50 buy-side firms covering hedge funds, private equity, and asset management firms. Our research support enables coverage of over 3,300 stocks and 3,400 corporates and financial institutions globally. We support more than 15 bank holding companies in their regulatory requirements and submissions. We operate from 7 research centers in Argentina, China, India, and Poland, and across several time zones and languages.

## **CRISIL privacy notice**

CRISIL respects your privacy. We use your contact information, such as your name, address, and email id, to fulfil your request and service your account and to provide you with additional information from CRISIL and other parts of S&P Global Inc and its subsidiaries (collectively, the 'company') you may find of interest.

For further information, or to let us know your preferences with respect to receiving marketing materials, please visit [www.crisil.com/privacy](http://www.crisil.com/privacy). You can view the company's customer privacy policy at <https://www.spglobal.com/privacy>

**Last updated: April 2016**