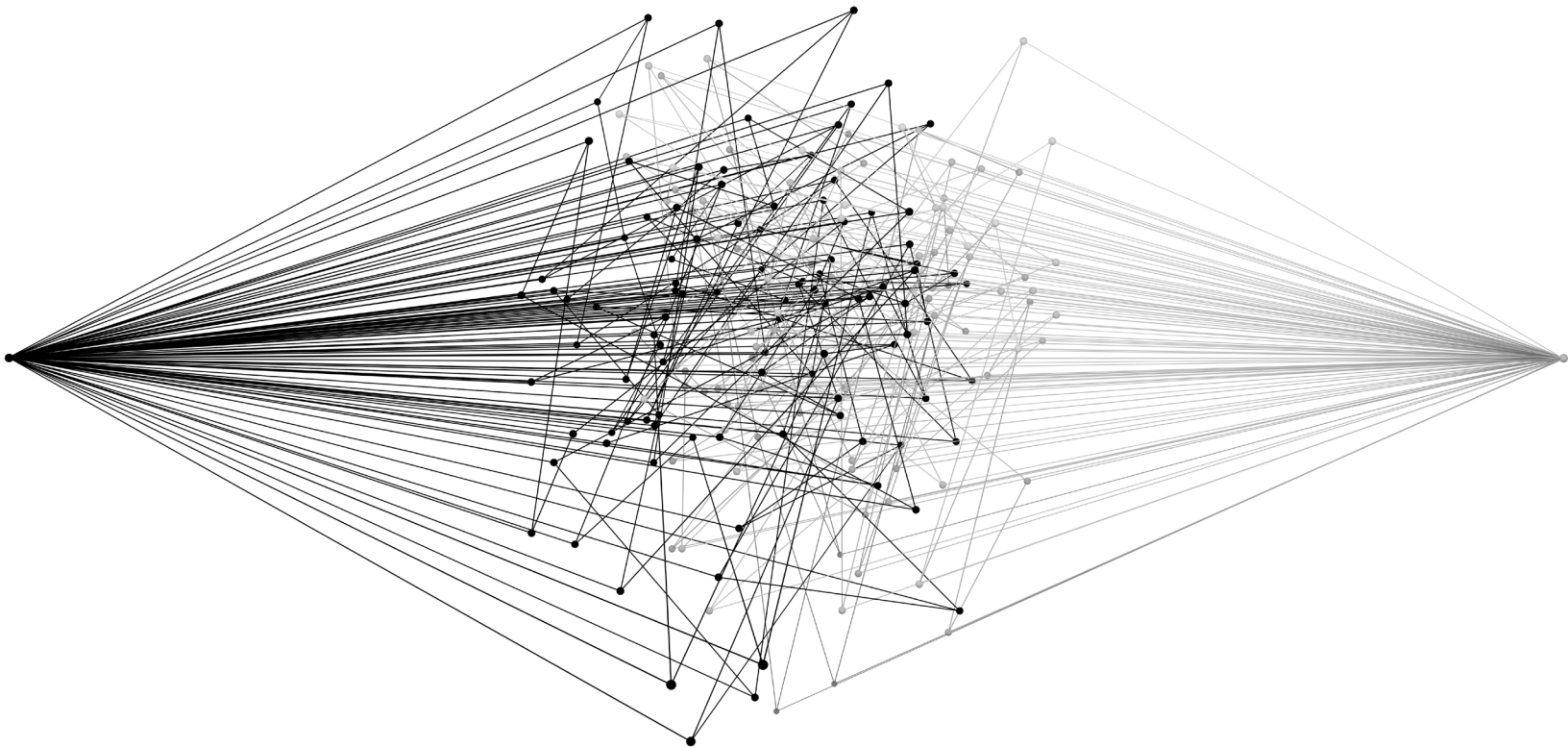


Risk Practice

# Flushing out the money launderers with better customer risk-rating models

Dramatically improve detection rates by simplifying model architecture, fixing underlying data, and using machine-learning algorithms to identify high-risk behavior.

*by Daniel Mikkelsen, Azra Pravdic, and Bryan Richardson*



**Money laundering is a serious problem for the global economy**, with the sums involved variously estimated at between 2 and 5 percent of global GDP.<sup>1</sup> Financial institutions are required by regulators to help combat money laundering and have invested billions of dollars to comply. Nevertheless, the penalties these institutions incur for compliance failure continue to rise: in 2017, fines were widely reported as having totaled \$321 billion since 2008 and \$42 billion in 2016 alone.<sup>2</sup> This suggests that regulators are determined to crack down but also that criminals are becoming increasingly sophisticated.

Customer risk-rating models are one of three primary tools used by financial institutions to detect money laundering. The models deployed by most institutions today are based on an assessment of risk factors such as the customer's occupation, salary, and the banking products used. The information is collected when an account is opened, but it is infrequently updated. These inputs, along with the weighting each is given, are used to calculate a risk-rating score. But the scores are notoriously inaccurate, not only failing to detect some high-risk customers, but often misclassifying thousands of low-risk customers as high risk. This forces institutions to review vast numbers of cases unnecessarily, which in turn drives up their costs, annoys many low-risk customers because of the extra scrutiny, and dilutes the effectiveness of anti-money laundering (AML) efforts as resources are concentrated in the wrong place.

In the past, financial institutions have hesitated to do things differently, uncertain how regulators might respond. Yet regulators around the world are now encouraging innovative approaches to combat money laundering and leading banks are responding by testing prototype versions of new processes and practices.<sup>3</sup> Some of those leaders have adopted the approach to customer risk rating described in

this article, which integrates aspects of two other important AML tools: transaction monitoring and customer screening. The approach identifies high-risk customers far more effectively than the method used by most financial institutions today, in some cases reducing the number of incorrectly labeled high-risk customers by between 25 and 50 percent. It also uses AML resources far more efficiently.

## **Best practice in customer risk rating**

To adopt the new generation of customer risk-rating models, financial institutions are applying five best practices: they simplify the architecture of their models, improve the quality of their data, introduce statistical analysis to complement expert judgment, continuously update customer profiles while also considering customer behavior, and deploy machine learning and network science tools.

### **1. Simplify the model architecture**

Most AML models are overly complex. The factors used to measure customer risk have evolved and multiplied in response to regulatory requirements and perceptions of customer risk but still are not comprehensive. Models often contain risk factors that fail to distinguish between high- and low-risk countries, for example. In addition, methodologies for assessing risk vary by line of business and model. Different risk factors might be used for different customer segments, and even when the same factor is used it is often in name only. Different lines of business might use different occupational risk-rating scales, for instance. All this impairs the accuracy of risk scores and raises the cost of maintaining the models. Furthermore, a web of legacy and overlapping factors can make it difficult to ensure that important rules are effectively implemented. A person exposed to political risk might slip through screening processes if different business units use different checklists, for example.

---

<sup>1</sup> "Money-laundering and globalization," United Nations Office on Drugs and Crime, [unodc.org](http://unodc.org).

<sup>2</sup> Gavin Finch, "World's biggest banks fined \$321 billion since financial crisis," *Bloomberg*, March 2, 2017, [bloomberg.com](http://bloomberg.com).

<sup>3</sup> The US Treasury and banking agencies have together encouraged innovative anti-money laundering (AML) practices; see "Agencies issue a joint statement on innovative industry approaches," US Office of the Comptroller of the Currency, December 3, 2018, [occ.gov](http://occ.gov). In China, the Hong Kong Monetary Authority has backed the wider use of regulatory technology, and in the United Kingdom, the financial regulator has established a fintech sandbox to test AML innovations.

Under the new approach, leading institutions examine their AML programs holistically, first aligning all models to a consistent set of risk factors, then determining the specific inputs that are relevant for each line of business (Exhibit 1). The approach not only identifies risk more effectively but does so more efficiently, as different businesses can share the investments needed to develop tools, approaches, standards, and data pipelines.

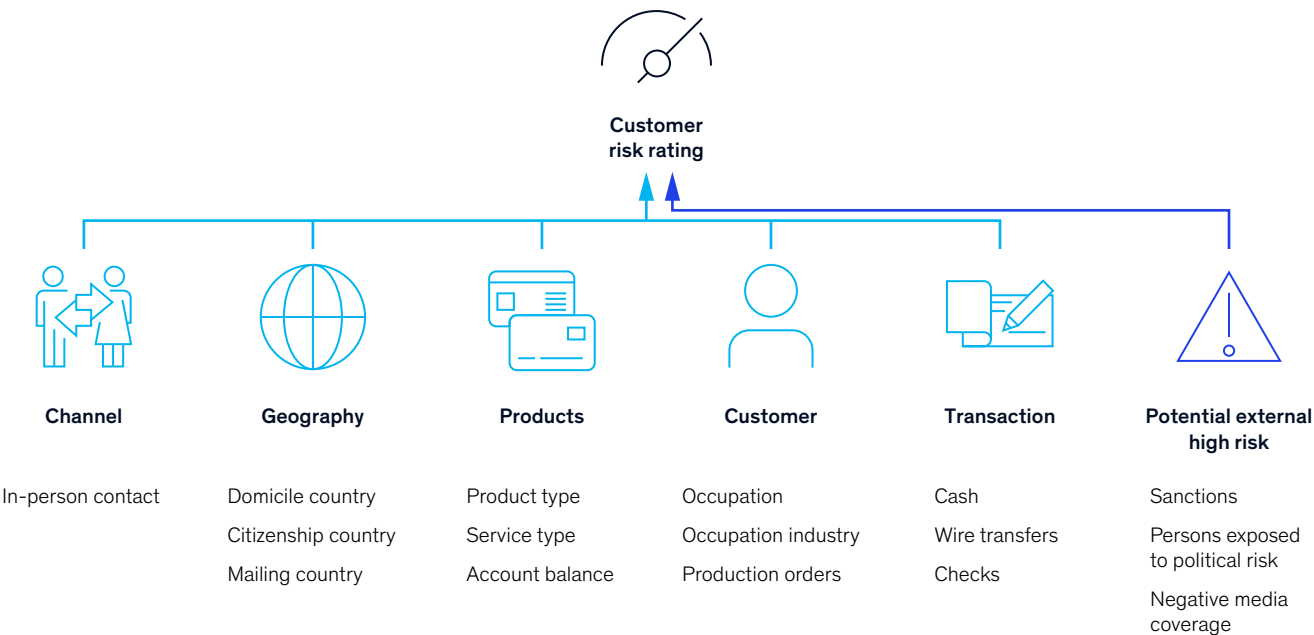
**2. Improve data quality**

Poor data quality is the single biggest contributor to the poor performance of customer risk-rating models. Incorrect know-your-customer (KYC) information, missing information on company suppliers, and erroneous business descriptions impair the effectiveness of screening tools and needlessly raise the workload of investigation teams. In many institutions, over half the cases reviewed have been labeled high risk simply due to poor data quality.

The problem can be a hard one to solve as the source of poor data is often unclear. Any one of the systems that data passes through, including the process for collecting data, could account for identifying occupations incorrectly, for example. However, machine-learning algorithms can search exhaustively through subsegments of the data to identify where quality issues are concentrated, helping investigators identify and resolve them. Sometimes, natural-language processing (NLP) can help. One bank discovered that a great many cases were flagged as high risk and had to be reviewed because customers described themselves as a doctor or MD, when the system only recognized “physician” as an occupation. NLP algorithms were used to conduct semantic analysis and quickly fix the problem, helping to reduce the enhanced due-diligence backlog by more than 10 percent. In the longer term, however, better-quality data is the solution.

Exhibit 1

**Effective, efficient risk-rating models use a consistent set of risk factors, though inputs will vary by business line.**



### **3. Complement expert judgment with statistical analysis**

Financial institutions have traditionally relied on experts, as well as regulatory guidance, to identify the inputs used in risk-rating-score models and decide how to weight them. But different inputs from different experts contribute to unnecessary complexity and many bespoke rules. Moreover, because risk scores depend in large measure on the experts' professional experience, checking their relevance or accuracy can be difficult. Statistically calibrated models tend to be simpler. And, importantly, they are more accurate, generating significantly fewer false-positive high-risk cases.

Building a statistically calibrated model might seem a difficult task given the limited amount of data available concerning actual money-laundering cases. In the United States, suspicious cases are passed to government authorities that will not confirm whether the customer has laundered money. But high-risk cases can be used to train a model instead. A file review by investigators can help label an appropriate number of cases—perhaps 1,000—as high or low risk based on their own risk assessment. This data set can then be used to calibrate the parameters in a model by using statistical techniques such as regression. It is critical that the sample reviewed by investigators contains enough high-risk cases and that the rating is peer-reviewed to mitigate any bias.

Experts still play an important role in model development, therefore. They are best qualified to identify the risk factors that a model requires as a starting point. And they can spot spurious inputs that might result from statistical analysis alone. However, statistical algorithms specify optimal weightings for each risk factor, provide a fact base for removing inputs that are not informative,

and simplify the model by, for example, removing correlated model inputs.

### **4. Continuously update customer profiles while also considering behavior**

Most customer risk-rating models today take a static view of a customer's profile—his or her current residence or occupation, for example. However, the information in a profile can become quickly outdated: most banks rely on customers to update their own information, which they do infrequently at best. A more effective risk-rating model updates customer information continuously, flagging a change of address to a high-risk country, for example. A further issue with profiles in general is that they are of limited value unless institutions are considering a person's behavior as well. We have found that simply knowing a customer's occupation or the banking products they use, for example, does not necessarily add predictive value to a model. More telling is whether the customer's transaction behavior is in line with what would be expected given a stated occupation, or how the customer uses a product.

Take checking accounts. These are regarded as a risk factor, as they are used for cash deposits. But most banking customers have a checking account. So, while product risk is an important factor to consider, so too are behavioral variables. Evidence shows that customers with deeper banking relationships tend to be lower risk, which means customers with a checking account as well as other products are less likely to be high risk. The number of in-person visits to a bank might also help determine more accurately whether a customer with a checking account posed a high risk, as would his or her transaction behavior—the number and value of cash transactions and any cross-border activity. Connecting the insights from transaction-

**While statistically calibrated risk-rating models perform better than manually calibrated ones, machine learning and network science can further improve performance.**

monitoring models with customer risk-rating models can significantly improve the effectiveness of the latter.

**5. Deploy machine learning and network science tools**

While statistically calibrated risk-rating models perform better than manually calibrated ones, machine learning and network science can further improve performance.

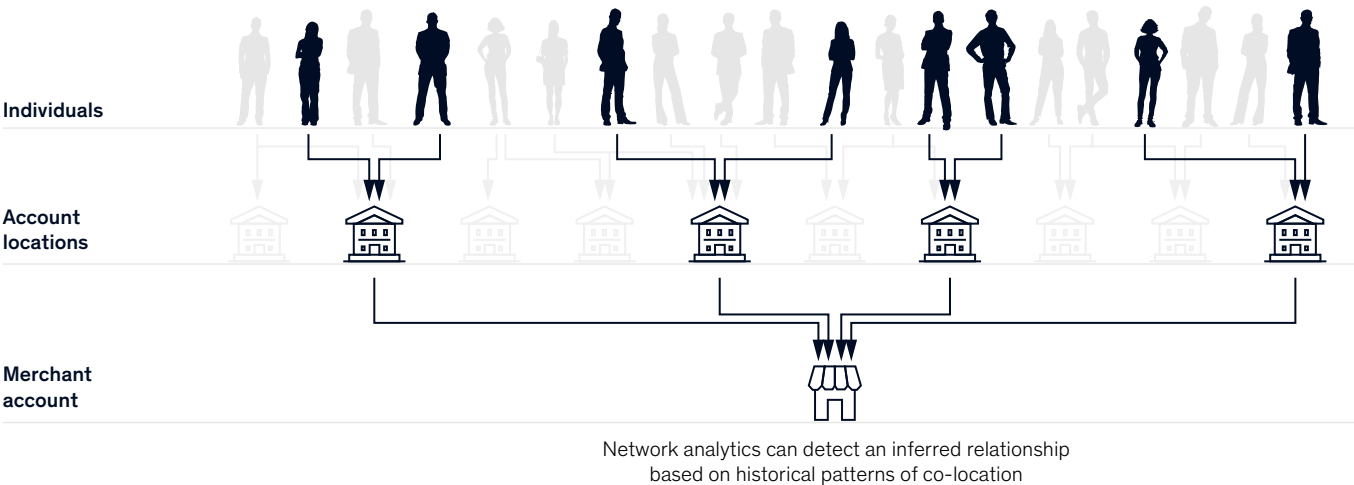
The list of possible model inputs is long, and many on the list are highly correlated and correspond to risk in varying degrees. Machine-learning tools can analyze all this. Feature-selection algorithms that are assumption-free can review thousands of potential model inputs to help identify the most relevant features, while variable clustering can remove redundant model inputs. Predictive algorithms (decision trees and adaptive boosting, for example) can help reveal the most predictive risk factors and combined indicators of high-risk customers—perhaps those with just one product, who do not pay bills but who transfer round-figure dollar sums internationally. In addition, machine-learning approaches can build competitive

benchmark models to test model accuracy, and, as mentioned above, they can help fix data-quality issues.

Network science is also emerging as a powerful tool. Here, internal and external data are combined to reveal networks that, when aligned to known high-risk typologies, can be used as model inputs. For example, a bank’s usual AML-monitoring process would not pick up connections between four or five accounts steadily accruing small, irregular deposits that are then wired to a merchant account for the purchase of an asset—a boat perhaps. The individual activity does not raise alarm bells. Different customers could simply be purchasing boats from the same merchant. Add in more data however—GPS coordinates of commonly used ATMs for instance—and the transactions start to look suspicious because of the connections between the accounts (Exhibit 2). This type of analysis could discover new, important inputs for risk-rating models. In this instance, it might be a network risk score that measures the risk of transaction structuring—that is, the regular transfer of small amounts intended to avoid transaction-monitoring thresholds.

Exhibit 2

**Network science can reveal suspicious connections between apparently discrete accounts.**



## The journey toward sophisticated risk-rating models

### *Getting started: How to move from horizon one to two*

Assemble a team of experts from compliance, business, data science, and technology and data.

Establish a common hierarchy of risk factors informed by regulatory guidance, experts, and risks identified in the past.

Start in bite-size chunks: pick an important model to recalibrate that the team can use to develop a repeatable process.

Assemble a file-review team to label a sample of cases as high or low risk based on their own risk assessment. Bias the sample to ensure that high-risk cases are present in sufficient numbers to train a model.

Use a fast-paced and iterative approach to cycle through model inputs quickly and identify those that align best with the overarching risk factors. Be sure there are several inputs for each factor.

Engage model risk-management and technology teams early and set up checkpoints to avoid any surprises.

### *Becoming an industry leader: How to move from horizon two to three*

Begin to build capabilities in machine learning, network science, and natural-language processing by hiring new experts or identifying potential internal transfers.

Construct a network view of all customers, initially building links based

on internal data and then creating inferred links. This will become a core data asset.

Set up a working group to identify technology changes that can be deployed on existing technology (classical machine learning may be easier to deploy than deep learning, for example) and those that will require longer-term planning.

Design and implement customer journeys in a way that facilitates quick updates to customer data. An in-person visit to a branch should always prompt a profile update, for example. Set up an innovation team to continuously monitor model performance and identify emerging high-risk typologies to incorporate into model calibration.

Although such approaches can be powerful, it is important that models remain transparent. Investigators need to understand the reasoning behind a model's decisions and ensure it is not biased against certain groups of customers. Many institutions are experimenting with machine-based approaches combined with transparency techniques such as LIME or Shapley values that explain why the model classifies customers as high risk.

### **Moving ahead**

Some banks have already introduced many of the five best practices. Others have further to go. We see three horizons in the maturity of customer risk-

rating models and, hence, their effectiveness and efficiency (Exhibit 3).

Most banks are currently on horizon one, using models that are manually calibrated and give a periodic snapshot of the customer's profile. On horizon two, statistical models use customer information that is regularly updated to rate customer risk more accurately. Horizon three is more sophisticated still. To complement information from customers' profiles, institutions use network analytics to construct a behavioral view of how money moves around their customers' accounts. Customer risk scores are computed via machine-learning approaches utilizing transparency techniques to explain the scores and accelerate

## Moving along three horizons, the model becomes more sophisticated and thus greater in its effectiveness and efficiency.



investigations. And customer data are updated continuously while external data, such as property records, are used to flag potential data-quality issues and prioritize remediation.

Financial institutions can take practical steps to start their journey toward horizon three, a process that may take anywhere from 12 to 36 months to complete (see sidebar, “The journey toward sophisticated risk-rating models”).

As the modus operandi for money launderers becomes more sophisticated and their crimes more costly, financial institutions must fight back with innovative countermeasures. Among the

most effective weapons available are advanced risk-rating models. These more accurately flag suspicious actors and activities, applying machine learning and statistical analysis to better-quality data and dynamic profiles of customers and their behavior. Such models can dramatically reduce false positives and enable the concentration of resources where they will have the greatest AML effect. Financial institutions undertaking to develop these models to maturity will need to devote the time and resources needed for an effort of one to three years, depending on each institution's starting point. However, this is a journey that most institutions and their employees will be keen to embark upon, given that it will make it harder for criminals to launder money.

**Daniel Mikkelsen** is a senior partner in McKinsey's London office, **Azra Pravdic** is an associate partner in the Brussels office, and **Bryan Richardson** is a senior expert in the Vancouver office.

Designed by Global Editorial Services  
Copyright © 2019 McKinsey & Company. All rights reserved.