



Mitigating Risk With External Data, A Guide For CROs



Table of contents

Reducing risk without reinventing the wheel	3
The rising cost of fraud	6
Leveraging machine learning for risk reduction	9
Use cases: is your data really enough?	13
Final thoughts: can you handle the pace of change?	18
About Explorium	21



Reducing risk without reinventing the wheel

Risk is becoming harder to identify as the world becomes more volatile and fraud continues to outpace our ability to identify it. It's the very last thing any CRO wants to hear — but unfortunately, it's true.

This makes assessing risk an incredibly slippery task. In volatile times, risky behavior and fraud attempts increase, while indicators that something is seriously amiss are shifting all the time. Looking back at how you worked these things out a year or even a month ago will seem less relevant and less useful. This makes preparing for future risk and fraud threats even more challenging.

It doesn't help that fraudsters are getting more and more sophisticated. Loan stacking and identity theft are rampant. Cybercriminals are so good at leaving no traces that the average breach isn't detected for many months after the fact. Accurately pinpointing incidents of fraud


and money laundering without interrupting perfectly normal business operations is a constant headache. Turning down unreliable applicants for credit while granting loans to underserved customers is a delicate balancing act.

To protect your business without losing your competitive edge, you need to broaden your horizons.

You need to look beyond the methods you relied on in the past to build better, stronger risk profiles. You need to design risk prediction tools that will work today, and tomorrow.

You need to consider whether the data you have can tell you enough to cope with emerging risks and pressures — and if not, what external data sources will give you the context and insight you need.

You need, in short, to find ways to reduce risk from your models without reinventing the wheel.



You need to consider whether the data you have can tell you enough to cope with emerging risks and pressures — and if not, what external data sources will give you the context and insight you need.



The rising costs of fraud

Fraud is big business, all over the globe. In the US, the FBI calculates that the cost of insurance fraud, excluding health insurance, comes to \$40 billion per year. The burden of these costs is typically passed on to customers, pushing up premiums for the average family by \$400-\$700 every year. In the UK, when you include all types of fraud, it totals a whopping \$224 billion per year, according to Experian. That's the equivalent of around \$7000 lost *every second*.

... And given our rapidly changing economic landscape, it's likely to get a lot worse.

It's not just the direct cost of the money stolen that makes fraud so expensive. Many organizations invest significant resources in the process of identifying and pursuing fraudulent activity after the fact. To make this more demoralizing, if your customers (or their details) are

the target of fraudulent activity, the chances are you will have lost their trust and loyalty forever — even after you've spent time and money investigating, and have refunded any money they've lost.

What's more, many financial institutions and other organizations tackling fraud are still using technology, strategies, and even data that's woefully inadequate for the challenge.

Some rely on clunky, outdated, rules-based legacy systems, which are inefficient and ineffective, often involving tedious, error-prone processes that push up costs. Because they work by testing claims or other documentation against set rules, they are only equipped to pick up on problems that are well-known and have occurred plenty of times in the past. That means they have no way of recognizing suspicious behavior that is very new, more nuanced, or otherwise falls outside these parameters.

Others recognize the need to be led by data, but they aren't using the right data in the right way to meet the challenge head-on. They're focused on analysis, extracting insights from historical data to figure out what happened in the past and use this to infer risks in the present.

Interrogating in-house data like this is certainly better than nothing. You're in a much better position to ensure the same problems don't slip through the net in the future, potentially reducing costs over time. If you plan to significantly reduce risk, though — and with it, the costs associated with fraud — this approach simply isn't up to scratch. As we'll see in a moment, you need to look outside the business for the data that will really make a difference.



Leveraging machine learning for risk reduction

Another issue with systems designed to prevent fraud and minimize risk is what happens when the pendulum swings too far the other way.


Sure, on paper it might be an effective anti-fraud strategy to simply treat more and more types of activity as suspicious, blocking transactions, delaying deposits or getting people to jump through a ton of extra security hoops at the slightest sign of trouble. But the costs of this are also high: annoyed, frustrated customers may soon be a competitor's customers.

At the very least, if the process is too arduous, both your customers and your organization's employees may be tempted to look for workarounds that actually make things riskier in the long term. This is where AI starts to come into its own. Machine learning programs sift through and aggregate information from masses of historical

transaction data, including ones that turned out to be fraudulent, looking for patterns in their characteristics. The resulting algorithm is applied to new transactions, credit applications, insurance claims or so on, in order to flag up anything that fits the pattern and thus indicates a potential case of fraud.

The big difference between this and more conventional rule-based or analytics-based approaches is that machine learning models are dynamic and continually evolving. The system “learns” from mistakes and new information, refining the model over time to predict instances of fraud with greater accuracy.

Plus, it’s applied in real time, meaning that situations it flags up as suspicious are blocked before money changes hands, rather than investigated after payment or other attempted actions are completed. This has the potential to save your organization millions of dollars, while helping to ensure that you don’t get overzealous, turning away potential customers or denying transactions that posed no threat.



Your machine learning program can only base its decision-making process on the training data it received, so if this is incomplete or in some way unrepresentative of real world scenarios, you could find that it contains inherent biases that undermine its value.

Here's the kicker though: you're still relying on whatever data you already have available within the business. Your machine learning program can only base its decision-making process on the training data it received, so if this is incomplete or in some way unrepresentative of real world scenarios, you could find that it contains inherent biases that undermine its value. To avoid this, you really need to ensure you access the best quality, most representative and extensive datasets you can. Oftentimes, these will need to come from external sources.



Is your data really enough?


As we've seen, you need plenty of data to do your job right. But you also need plenty of the *right* data.

The overwhelming likelihood is that you don't have access to all the data you could possibly find useful within your organization, especially in these uncertain times. You need to look outside your own data depositories to put this data in context and understand where the real risks lie.

Let's take a look at three examples of how that works in practice:

→ **Credit risk scoring for underserved customers**

Deciding what data to trust is also an important consideration when you're deciding who deserves to be given the green light for credit and loans. To manage risk, many organizations tend to rely exclusively on scoring systems from big credit rating agencies, but



The overwhelming likelihood is that you don't have access to all the data you could possibly find useful within your organization, especially in these uncertain times. You need to look outside your own data depositories to put this data in context and understand where the real risks lie.

this is a flawed system that can result in many perfectly reliable people getting turned away.

This is especially true if the applicant has a somewhat non-traditional work setup — they're self-employed, a freelancer, or work in the gig economy, for example — or have not attempted to access credit before (which, in our rapidly changing world, covers an ever-expanding segment of the population!).

For this reason, many lenders are looking to alternative data sources to get a more accurate picture of the applicant's credibility. This can mean a broad range of external datasets, from domain information and business registration to government filings to social media data.

Incorporating this kind of information and nuance into your models is an incredible way to improve decision-making and reduce risk. Just bear in mind that, to make it feasible, you will need a powerful platform in place to streamline or automate the process of acquiring, matching and cleaning data so that it's ready for use.

→ **Anomaly detection in anti-money-laundering (AML) efforts**

Many financial organizations delve into time-series data to try and spot oddities in account behavior that are consistent with money laundering activities.

This is a good approach, but it tends to lead to a high number of false positives. You can greatly improve accuracy by bringing in more contextual information that helps you understand the transactions you're looking at and whether they are in keeping with the type of customer or business. That means incorporating external data on the company and the individual into the model and combining this with your AML efforts to get a complete picture.

→ **Uncovering insurance fraud in the medical sector**

In the last chapter, we saw how machine learning algorithms play a vital role in accurately detecting fraudulent insurance claims, based on historical data.

You can enhance and speed up your results, though, by adding carefully chosen external data to the mix that gives you a deeper understanding of the customer's context. That may include scanning documents or bringing in other types of data on the applicant's online history and behavior.






Final thoughts: can you handle the pace of change?

When the situation takes a hard swerve, you need to adapt and recalibrate fast.

Whatever rules, models, or strategies you've been using to calculate and monitor risk and fraud in your organization inevitably need updating. At the same time, you don't have time to wait around for your data to catch up.

The great news is: you don't have to. With the right technology, you can tap into the most valuable, up-to-the-minute, external datasets, giving you excellent visibility over risk in your business. A powerful platform will automatically clean and harmonize this so it's ready for use, streamlining the process of combining this with your own data.



Whatever rules, models, or strategies you've been using to calculate and monitor risk and fraud in your organization inevitably need updating. At the same time, you don't have time to wait around for your data to catch up.

Yes, there are more dangers out there than before. But solutions exist to help you keep on top of these threats and to adapt to them at lightning speed. It's up to you to make sure your organization embraces them.



About Explorium

Our automated data discovery and feature generation platform automatically connects a company's data to thousands of relevant premium, partner, and open data sources to extract an optimal feature set based on model impact. We're creating a new category as the first company to empower and service business leaders and data scientists with end-to-end automation of data discovery and feature generation —

fueling superior decision-making and driving real business impact.

