NO SILVER BULLETS

Why ‘99% accurate systems’ could let you down 99% of the time

An extract from Perspective: The fraud and identity issue

Essential insights into the issues facing your industry today
Steven Hicklin, Head of Innovation and Research at Equifax, explains why ‘foolproof’ systems can prove both costly and inaccurate when it comes to fraud detection.

Let’s start with a statistics game. Please note: what you should know up-front about this game is that a lot of people have studied it and, even though it’s about statistics, people with a statistics background will get it wrong more often than someone who randomly guesses the result.

A man – let’s call him Bob – goes for an annual health check and his doctor carries out some blood tests. A week later, he’s called back and the doctor tells Bob that the blood test shows he has a very deadly and very rare tropical disease, which affects one in 10,000 people. The doctor assures Bob that the test is 99% accurate and that he should put his affairs in order.

What are Bob’s chances of actually dying of this disease?

a) 99%  
b) 90%  
c) 10%  
d) 1%

The answer is d) – he has 1% chance of actually being ill. That’s feels strange right? The test is SO accurate… 99%!

Think about it this way:

We select one million people at random off the street and we carry out this blood test on all of them. 100 of them (one 10,000th of the population) will have the deadly disease. Our test is 99% accurate, so it catches 99% of them – that’s 99 people.

If our test is 99% accurate, then it means it has a 1% error rate – so it declares 1% of the million people in the test as being ill – 10,000 people.

So – we have 10,000 positive tests – of which only 99 are accurate. This means that the chance of any specific one of those 10,000 people really being ill is only 99/10,000 – a fraction under 1%. An alternative way of reading this is, ‘there are way more false positives than real results’.

Being even more frank: ‘humans are poor at judging the meaning of odds accurately’. There has been a lot of research in this area and it’s a very commonly identified issue.

So – that was all very jolly, but what does it have to do with fraud?

The odds of fraud

Let’s change the question: I am a business with one million credit lines. I think about 1 in 10,000 of my accounts are fraudulent on any given day. I have a fraud detection system that scans all of my accounts every day and has a 99% accuracy rate….

Sound familiar?

In this example, the fraud detection system will detect 99 of the 100 frauds, but will fire 9,901 false positives into work queues to be manually checked. If each case takes one minute to check, then that will drive 9,901 minutes of work – 165 hours or 23 full-time equivalent (FTE) each day just to cover the false positives. Remember: that is 23 FTE per million credit lines – 10 million credit lines would drive 230 FTEs; although, more likely, it might drive 100 FTEs and people spending less time per case, or possibly even just hitting the delete key… a lot. It happens.

Let’s take this a little further. If those 23 FTEs were paid around £15 per hour to take into account a fully-loaded cost, and we assume that the work queues are processed every day of the year, then we get to an OpEx cost of around £579,000 per year – which may be more than the cost of the platform itself.

Scaling this up, 50 FTEs working fraud cases could be driving perhaps £1 million of cost per year.

This raises a few interesting questions – most of them revolving around the concept of ‘how can I spend less money’?

Reduce the cost, boost the results

Your first option is a ‘Defence in Depth’ approach. Let’s see what happens if we have three fraud detection systems. Let’s say that we use three systems that – critically – test for fraud in three totally different ways. We’ll say that each system is 99% accurate.

A different way of looking at this is saying: ‘We grab one million people off the street and test them for our rare disease with three completely different tests which are each 99% accurate. What are the chances that Bob is healthy, but all three tests tell him he is ill?’

This answer is significantly better – it’s 0.0278%. Out of our million people, we get only 278 false positives, instead of 9,901 false positives. Applying this to fraud, with our new approach, we still catch 99 of our 100 fraudsters. In cost terms, you no longer have 23 FTE working flat-out on false positives – you have one FTE working for only 4.5 hours per day. The result? You save 22 FTEs’ worth of cost.

There is a flip-side to this: you would need to change your processes a little. Only 86 out of the 100 active frauds would be caught by all three systems. 99% of them would be caught by two of the three, and almost 100% (99.99%) would be caught by one of the three.

Whether there is a value in using three separate high-quality systems is a trade-off. If you have a relatively small portfolio, you may not see the operational savings. With larger portfolios, they will be found. It will all boil down to a cost/benefit study, rating your portfolio size to the costs of false positives.
No silver bullets

Variations on a theme
And you’re not limited to a specific number of systems either – you could use one high-end 99% accurate system and implement alongside it a single low cost system giving a rate of, say, 95% – so a 5% error rate. This would still weed out the majority of the false positives while keeping your catch rate high. Costs still fall but not by quite as much. In this case, the million record portfolio would throw around approximately 500 false positives per day – which would lead to around 1.2 FTE worth of false positives per day.

The ‘best’ selection of numbers of systems and their qualities then boils down to a basic cost/benefit analysis. If nothing else, it makes for an easier conversation with your finance groups than perhaps more qualitative subjects like ‘friction’ and ‘customer experience’.

But...we have other options available to us to reduce cost and still keep catch rates high.

What else is out there?
One alternative is to use our existing systems to completely change the problem: let’s not use them to look for ‘fraud’ but to look for ‘honest people’ instead.

There are some good mathematical reasons for this. In a large book of business in the likes of a bank or a telco, fraud is relatively rare – perhaps 1 in 10,000 accounts in a single day or so. If you have a lot of accounts you still get a lot of fraud – there are 365 days in a year and, even at that low occurrence rate, a medium sized portfolio could still rack up thousands of frauds in a year – but in the grand scheme of things, you’re still looking for the occasional black swan amongst a huge numbers of white swans.

In statistical terms, we have a ‘heavily skewed’ dataset – lots and lots of honest people vs a few fraudulent people – and, to make it worse, we have a myriad of fraudsters carrying out different types and patterns of fraud. We therefore enter an area where regular statistics, like regressions, are not at their peak of accuracy.

So – what about if we don’t try and detect fraud? What if we try and detect ‘not fraud’ instead? Put a different way – what if we didn’t try and use complex methods to detect a needle in a haystack? What if we created some really simple tests to detect hay instead? There’s a huge amount of hay in the haystack, and so we can create some very good models to detect it. Then, wherever we feel confident that we have hay, we can put it to one side.

The magic of this approach is that while we have to use our modelling methods at the limits of their abilities to detect very rare needles, and so getting lots of false positives, there is so much hay that all of our favourite methods work much better. The result is that we can use simpler algorithms and get better results.

If we took this approach, how would it work? What we’d be left with is one large pile of hay we are extremely confident contains NO needles at all. We would have a second, much smaller pile which contains some hay and some needles. But now, rather than the needles being 1 in 10,000 occurrences, they are now going to be somewhere in the 1 in 100 occurrences. That’s brilliant – because the ratio of ‘Fraud: Not Fraud’ is so much closer, meaning we can use our existing fraud systems to get higher catch rates with far less false positive rates – and so incur less cost.

In effect, by changing what we model, we can keep all of the tightly integrated systems (which need large IT projects to significantly change) in place and hugely increase their effectiveness – which means a much simpler business case with the Finance teams.

In reality if we do this, we also see a few other benefits as well.

When we have our ‘big stack’ of honest people and our ‘little stack’ of potential fraudsters, an analysis of the ‘little stack’ will usually show that a lot of the ‘little stack’ consists of data errors. It’s often very hard to determine the difference between data errors and some types of fraud. Is Bob using a false address, or has his real address simply been keyed incorrectly? Is Alice trying to use a fake name, or has someone entered it incorrectly?

READ MORE...
What this may mean for business

While it may not get you the next conference slot at one of the Big Data conferences, data quality projects are an excellent way of driving value into, and cost out of, an organisation. False positive rates fall dramatically in fraud: credit checks return fewer thin files; debt collection is simpler; financial modelling is more accurate; and security measures are more precise.

It’s rare that a single activity positively affects so many arms of an organisation, but data quality is one of these. With the ‘little stack’ you have set aside, this is a prime location for your data quality or data science groups to build their patterns and models of any data issues to allow for a much simpler fix process with fewer sweeps through the data.

But how can we model ‘honest’? Well, we can use regression or kMeans (clustering models) as normal. Moving into the machine-learning world, recommender systems offer a great route into this approach because, instead of them being used to say ‘people who bought toothbrushes also bought toothpaste’, you can use them to say ‘Bob is honest – find more people like Bob’.

Systems like classifiers and clustering systems work very well too. Wherever possible, Equifax suggests using two-sided classifier algorithms, such as Support Vector Machines, for this work. The initial set-up is more complex and it’s a steeper learning curve to climb, but this family of algorithms has been shown to be especially well adapted to noisy and complex datasets – which is usually what is going to be processed. While more complex to build and run, when you are seeking to eke out the last drop of accuracy, these algorithms can be worth the effort.

Critically, they don’t need to be bleeding edge systems and models created by teams of PhDs if you choose not to go down this route. Building a few simplistic algorithms and using all of them helps us in the same way as using the multiple fraud detections systems above: you can run your client set through multiple algorithms and put anything that all three simple algorithms flag as ‘honest’ into the ‘big pile. You get very accurate results without having to create extremely accurate algorithms.

Safety in numbers

A number of simple algorithms will usually beat one very complex algorithm. If you want the absolutely most accurate method, then multiple complex algorithms will trump both of these.

For anyone who has worked in fraud detection for a while, the concept of ‘model normal’ feels very alien: our clients constantly tell us this. The way to think about it though is that, at some point, statistics had never been used for modelling fraud at all: to the first groups that were told to “go and use statistics to capture fraud….”, this would also have felt alien. As an industry, we are used to the idea of modelling fraud but, if it’s approached from first principles, it’s hard to find a good solid mathematical foundation for doing it that way.

Working out the ‘best’ approach for you and your organisation is also not always obvious. One route would be to look at it from a metrics-based approach, such as ‘which route leads to the best optimisation of cost/revenue and client satisfaction?’ Another would be to trial all the different methods in a champion/challenger approach and see which blend of them your organisation feels most comfortable with.

There is no silver bullet for defeating fraud. And, because it doesn’t happen very often, our existing methods can play tricks on us and drive in a lot more cost than we expect. If we want to keep using our existing methodologies, we can use a ‘Defence in Depth’ approach to drive cost out of our companies while still keeping catch rates high.

There are also alternative methods we can use which change the problem (instead of looking for new solutions) and make our lives easier as well as produce good results, but this requires a change of mind-set – and the hardest part of any business transformation – which this would be – is changing mind-sets.

To find out more about how we can help, you can contact Steve by emailing steven.hicklin@equifax.com or calling 07827 805710.