Market Abuse Surveillance TechSprint (July 2024) video Transcript.

Team 6. Conatix

Delegate 1

I'm David, CEO of Conatix, and I'm proud to say it's our second FCA TechSprint. We're a cybersecurity startup and we work with universities. We've won awards. We were named one of the most innovative cyber startups in the UK by DCMS, the Government Cybersecurity Directorate. We're very active in the UK, but we have a global footprint and team around the world. And what we do is at we work at the juncture of AI for cybersecurity at Compliance Analytics.

And one of the main things we do is insider fraud detection on the other side of banks in the retail and commercial side, where you're looking at what employees and suppliers, vendors are doing on your network for suspicious activity behaviour. And that led us to think we might apply some of those approaches and concepts to the trading side. And, and that's what we've done.

Is there a time? Oh, OK, sorry.

So we applied this to problem 3, anomaly detection, which we do a lot of in other contexts on for market abuse surveillance. And so we used something that in cybersecurity is called a user entity behavioural analytics, looking at the behaviour of actors in the in the system, in the in your network and in the context of banking. You want to make of a retail bank, you want to make a model of what each individual is doing and compare it to what they've been doing historically.

Are they doing something today that's not normal based on what they've been doing for the past six months or a year and themselves. And maybe it's not, is it normal or is it extremely different from what their peer group is doing? And you can find a peer group organically by clustering people based on their historical behaviour. So in banking, someone in the marketing department, people in the marketing might have distinct behaviour on the network that's different from people and human resources, different from people in IT.

And if let's say during the lunch hour, someone in a call centre who deals with customer accounts suddenly starts behaving like someone in the IT department who accesses certain kinds of files and directories that a call centre person doesn't normally do.

If they jump a cluster, that's interesting and potentially an alert for anomalous and malicious behaviour. And we can apply the same thing to the markets. And So what we did was with this data, we clustered the actors by the asset class and, and the size. So the traders by their type of portfolio.

We extracted a measure of portfolio diversity and, and how big their trades are. So we and then so we came up with six clusters and we took out the large players and the large equities and we concentrated on the small players with small and medium portfolios and the medium sized players with medium and small portfolios to look for when are they doing something that will that where they're gaining a lot of value, making a lot of money in a single day versus their peers. So we're not using any rules or parameters.

Everything is relative to their peers or to their history to a baseline. And so when you do it for all the entities together, you get a lot of noise, you get a lot of potential anomalies. But when you segment people by cluster then, or entities and players by cluster then you cut out the noise. You get to see things through a different lens. This is based on actual data. So each of these is a massive value gain in a single day, more than Z score, more than 7 standard deviations from the norm of their peers.

So that's someone made a lot of money relative to people with similar portfolio construction, similar trading sizes in a single day. So each of these is potentially something that we want to look further into. And so all we did so far in the time that we had was to identify these anomalies.

The next step would be to look inside the spikes and the time series for The Who, what, when, why and how. And, and you can do that, you can automate that, you can do that, you know, it's a starting point for forensic analysis. And of course you can build feedback loops to, to be able to say, well, this kind of pattern is OK, this is normal, or we're willing to accept this pattern.

So in the context of a bank, someone you know, worked past 8:00 PM or looked at customer service centre, looked at 20 times more customer accounts than they normally do. Well, this was a busy day or a productive day. That's OK. They weren't trying to steal data about customers. So you can give it feedback and it knows what to ignore in the future in, in giving you alerts. And so we quickly, oh, and I wanted to say what I meant to say at the beginning.

We've done a lot of these. We've been around a little while. We've done a lot of accelerators, TechSprints. And the quality of the mentoring on this one for the mentors we spoke to was exceptional.

We want to thank Steve from Morgan Stanley and Shane from Bank of America, especially Devendra from Barclays, who inspired us to try to extend our anomaly detection even to dabble in problem set three. And to look at specifically the use case of spoofing where someone puts in a large order, buy or sell, and then cancels in order to move the price and then quickly executes an opposing order. And so we actually, and so again, we used our clusters of, of, of, of entity and asset trading types and, and styles and we used fake data this time. But the bubbles represent the order size. And if it's the same colour, then maybe they're working both, it's the same entity working both sides of spoofing on the buy side and the sell side. And you, you can do, you can create something like this with without great difficulty. And you could then put in philtres by security type, by time, by Z score.

So how extreme was the value gain? So you see somewhere up at the top, but each of these here we just look for one standard deviation and above. But you could set it however you like. And so here you're not using rules, you're, you're not setting parameters from the top down, you're deriving them from the bottom down. And so, and then, you know, the next steps would be, you know, forensic analysis, matching anomalies to actual events, of course, using real data for the simulated data and oh, and filtering everything so that you can look in more detail at, at, at what's actually happening. So sometimes even deconstructing machine learning into simpler methods can also be effective when you apply them carefully.

Thank you.