FINANCIAL CONDUCT AUTHORITY

Market Abuse Surveillance TechSprint (July 2024) video Transcript.

Team 2. eflow

Delegate 1

So thank you all for joining me, Jonathan Dixon and E Flow for our market abuse surveillance tech Sprint today. So what are we trying to achieve and what is the impact from the work we have done? Firstly, we're aiming to use machine learning to identify market moving events, client actions and trading strategies that can precede cases of market abuse.

Secondly, we will use these events and client actions against historical activity to identify where trading activities within normalised levels, both reducing false positives and enhancing detection capabilities. Next, I want to discuss our methodology which comprises of four key areas.

Firstly, data preparation. This involves the scaling and normalisation of data together with the creation of time series windows. Secondly, we use the supervised machine learning model trained with rules based surveillance processes to accurately assess distinct threads within a wider data set.

Thirdly, the machine learning model is trained within the current neural network.

And lastly, we use proprietary risk scoring together with correlation assessments between traditional rules based events and machine learning output for A2 way feedback loop.

This feedback loop using the analyst risk scoring of individual alerts will be used as the final piece in the puzzle to both validate and tune machine learning risk scored output. And it is this feedback loop that allows to create client centric surveillance and be innovative in the detection of risk events.

Next, the innovation. Here I will examine how what we do makes a difference and how it does so in ways that are not currently seen within the marketplace.

Firstly, thanks to the FCA, we could train machine models, machine learning models through large real world data sets. This combined with traditional parameter based events model learning has allowed to leverage the strength of our traditional and future states events.

Secondly, the feedback loop in our case management system for the identification of false and true positives will help drive the Third Point, the machine learning driven risk scoring of alerts and events. This in turn will lead to the last and perhaps most pertinent point. The machine learning can identify events that pinpoint the initiation of market abuse.

The confusion matrix in this slide allows us to see the result of machine learning training on alert output, with the bottom two quadrants showing where a traditional parameter based surveillance solution for spoofing identify potential malfeasance to alert generation and the right hand two side quadrants, the route on the right hand side show where machine learning models how to identify potential market abuse.

So what's this mean in practise?

Firstly, at the bottom right, we can see 281 examples of where both traditional and machine learning models created alerts. In essence, these the high value alerts those you want your analyst limited time being focused on at the beginning of the day.

Next, at the top right, we can see 31 examples of where the machine learning model identified potential market abuse that was not seen by traditional surveillance systems. This is the value add the additional alerts not seen by a traditional solution.

Next, we can see 147 examples of where the traditional solution identified potential market abuse but did not have it validated by the machine learning model. These are alerts that could generate a lower risk score, allowing analysts to focus in their initial efforts on those higher value 281 alerts first referenced.

And lastly, we have over 20,000 time series window events that did not generate alerts true negatives. This further is further extrapolated to risk levels on this slide, which will be available for reference after the presentation.

Next, we can see the solution in action. The top left shows the results of an insider dealing alert and the ensuing risk scores based on the degree of deviation from the trade surveillance parameters established when it comes to the parameter based solution, together with the AISS risk score looking at the impact and risk of news events as well as the associated E communication channels.

Scrolling down, which I'm sure will happen in just a minute. Oh yeah, scrolling down we can see the trades in question as well as beneath that the relevant E communication points. These can include voice communications, e-mail, WhatsApp and all relevant non structured channels. Beneath that we can see the timeline in graph format and the case management tool at the very bottom left, allowing analysts to select risk rankings and reinforce the feedback loop.

So what's next?

Firstly, training machine learning models across multiple surveillance typologies.

Secondly, integrating risk scoring to identify high risk events that should be the focus of surveillance analysts limited time.

Thirdly, feedback loops to reinforce machine learning processes and further refine the alert output and risk scoring.

And lastly fourthly, the utilisation of non structured data such as ecoms and news to help identify intent behind trading activity and further refine machine learning models.

So in conclusion, machine learning output has produced output. Machine learning has produced output that does four key things.

Firstly, it found its traditional surveillance activity confirming high value alerts.

Secondly, it provides additional hitherto unknown alerts for additional investigation. The value added proposition.

Thirdly it allows for a risk scoring of unmatched traditional alerts that were not validated through a machine learning process.

And lastly feedback loops via human risk scoring of alerts to reinforce the machine learning model and allow for a client centric dynamic parameterisation. Machine learning is here to add value to existing processes until it can be fully risk accepted as a replacement.

Similar to what Rob was saying, we're in an area of transition as we move towards the integration of machine learning methodologies away from traditional solutions. We aim to prove the proof of concept and advance the technology. Measuring twice and cut once, we hasten slowly towards a new beginning.

I'd like to thank you all for your time and I'd like to extend A-Team wide.

Thank you to the FCA, without whose support and data this work would not have been possible.

I'd also like to thank Marsha and Ben for me for giving me the opportunity to be up here and perhaps most importantly, our engineers who are sitting with us over there without whose hard work and diligent work we would not be able to generate the results that we have.