

Market Abuse Surveillance TechSprint (July 2024) video Transcript.

Team 3. Pytilia

Delegate 1

So good afternoon, everybody. I'm Neil Sinclair and here with Tim Silversais today presenting for Petalia and we'll be joined by Ian Dalton for the Q&A. Just wanted to kick off by y looking at the the jigsaw piece puzzles that we brought together over the course of the tech Sprint.

So we're kind of thinking a little bit about Pytilia. We've got a strong financial services domain expertise, particularly wealth management and have implemented a rules based trade surveillance solution in that space. We've recently one of the winners of Fintech Scotland innovation call, which was looking at adding AI power into our surveillance solution. And on that theme, we've got a number of research partnerships with universities back in Northern Ireland looking, looking at various AI solutions.

On that academic thing, we are grateful to, to Barry, also Queen's University and Nick for their input and to the various academic papers and models that we're able to, to bring to play. Also bringing the day-to-day experience.

Thanks also to Dev Pandit just to get them the, the real world feedback there as to what is the areas to focus on and from that equity, equity derivative being their priority area. Also interesting to, to kind of find out what their challenges are. And really goes back to to what we were saying earlier about just the number of false positives. And then finally echoing what you were saying earlier, Jessica, just getting access to the high quality data is just really goldmine words can't describe the benefit of that. Just handing over to Tim then to talk us through the methodology.

Delegate 2

Sure. Thanks Neil. So yeah, it just won't take a couple of minutes to talk through problem statement, our understanding of that and our approach to a solution. So as Neil mentioned, our data ingest combined and cleaned all three data sources provided in the FCA sandbox.

It's the order book, the transaction data and the provided news feed samples. And what are we trying to do with these? Before we jump into the detail, I guess probably three things I would call out. One is we're focusing initially on deflection of cross product abuse using equities and equity derivatives as a sample, acknowledging that there are other forms of cross product abuse that are interesting and that our model can be extended to address those as

required. Secondly, that we're aiming to target both transaction driven and quote driven manipulation techniques and thirdly to provide an outcome which is both explainable and robust in the face of market volatility. So at the other end, a large proportion of our effort was spent on feature engineering. In other words, identifying which features we should derive from the raw data supplied by the FCA and use as inputs to our chosen models.

So our data was initially aggregated into time based buckets per product family classified using a moving time window to avoid blind spots. And I will list every feature that we've listed a few up in the slides, but perhaps a few interesting ones to call out include order book imbalance, a bid ask spread V pin and V1. And we can go deeper into those as required in Q&A or offline later. But broadly speaking, we're aiming to catch a blend of time and volume based metrics. And it's probably worth saying that our feature selection was informed by recent academic research including a couple of papers that were alluded to on the previous slide.

And so having taken the initial pass through that feature engineering activity, I guess the next step for us was to choose an appropriate model for anomaly detection given this feature set we derived. And again, our decisions here reflected sense of background reading and model selection plus hands on experimentation with some likely candidates.

And our conclusion was that non solver model made most sense where we supplied the same feature set to two separate models, isolation forest and auto encoder, and combined the models like on outputs using a composite scoring methodology.

The risk of getting into the weeds rationale for this was really that each model had strengths and weaknesses which complemented each other and combining both allowed us to reflect both short term activity and longer term pathings patterns with a view to building in that robustness to which we called out as a goal.

The output from this process then is a set of anomalies, each with contextualised output referencing relevant order book sequences and transactions.

Post processing analysis which is an ongoing activity and provides additional human readable context on the type of suspected mark manipulation detected and the reasons for identifying each anomaly, each cluster of orders and transactions as an anomaly in the 1st place. And just very briefly take a look at how the components of the solution can be represented in the architectural block diagram as that hopefully provides a useful input into the road map discussion, which may also update us three.

Delegate 1

Thanks Tim. So where are we and what's next?

So as we as we touched on earlier, we we've got our initial rules based deployment and thanks to modular architecture there, we're able to pull in quite a number of Booker dealer feats. It was from 24 Booker dealers and up to 72 bespoke file formats there.

Great to get sort of user feedback on that, complementing the the quality of the data and the ease of use we mentioned then adding the AI part aspect in Fintech Scotland. I was looking at buy side and actually random forest model in that case. Tim's covered what we've done during the market abuse tech Sprint. And no surprise there that we've kind of got more ideas than we had time to kind of fit into the available window.

So what are we, what are we wanting to do next?

Again, that idea of the, the feedback loop, arguably the most important part of the process that the compliance analyst and kind of learning from, from their experience and incorporating that also kind of adding more into the, the data ingest. We, talked about false positives earlier, but that's actually an interesting source of input.

So why not get a best of both from the existing tooling, feed that into this processing and see what the verdict is hoping to do some sentiment analysis on, on the news as well. And as well as a a time clock looking at a a volume clock idea as well.

We, focused on equity and equity derivative as per the the mentor guidance, but we, we've architected for correlation matrix to let us explore further areas.

So just to to close in Italia's kind of got innovation R&D at its heart. We've got that kind of solution, surveillance solution that that's kind of out there got strong user feedback and we've demoed more capability during the tech Sprint and we've got lots of ideas about where we want to go next.

During Fintech Scotland, one of the ideas that came out was compliance, enabling innovation, regulation enabling innovation. That's something that we want to keep going with. And I guess our asks then is kind of for the, I suppose, ongoing kind of type of input we've had from the mentors around compliance expertise and experience.

And maybe in summary, then we're interested in your problems bring, bring it on and we'll, we'll solve them together.

So it just remains for as we move into Q&A, just for me to say big thank you to everyone.

Big thank you for everyone who who's helped us during or to working together.