Meeting with Conrad 1/31/2022

Purpose: Gain application knowledge of outlier detection methods; Learn about various methods of outlier detection methods.

Q: What is the main purpose of your models? Is it to just flag the fraud or is it designed to block the transactions?

An outlier doesn’t mean fraud: just because it is anomalous doesn’t mean it is fraudulent.

They don’t tell what the risk value is

Typically fraud looks like everything else because people may buy things at various velocities. It is normal to buy something a couple times a week.

Domain wise: outlier and anomaly are very similar

There are different types of fraud: payment, identity, password

They track velocity rather than total number of users. They take the derivative of the users count

Q: What type of models do you use?

They use normal models, not a typical outlier detection model. Cyber security uses outlier detection methods to help them catch things.

Q: in context of your domain, what do you look for when trying to find outliers.

They track trends, user behavior, location, and spikes.

Q: how would you handle novelty detection? Do you build the model on the current data first, then put that model in your data pipeline?

They host the model

Q: I’ve learned about various models already, but one thing that I haven’t quite found an answer to is how to categorize them. Ex: which ones are considered global outlier detection, local outlier detection, or can they be both but it comes down to the way you tune them.

Q: What are some good things/tips/tricks to keep in mind while trying to remove outliers from the dataset?

Context of the outlier. What are you willing to sacrifice?

Q: Do you build your own custom models?

Q: Is it normal to have an ensemble of methods in a pipeline?

They use decision trees, normally just one model.

Neural nets are great for problems that involve spatial data

Don’t be afraid to not use a neural net.

The organization of the data science team is just as important as the tech.

Do you expect the Data Science team to put things into production. ML engineers are your CS people who

**Hosting a model**

If you export a model in sci-kit you cannot use it in tensor flow. Spark is what Kount uses. The models are not cross compatible. This is why they host the models. Don’t think you can make a model and someone else uses it. You create an API around the model.