

Feedback for Anomaly Detection

Nice write up! Thanks for spelling out the column headers and dataset dimensions, that level of clarity makes a big difference.

Great first steps; it will be really interesting to see whether the supervised or the unsupervised approach ends up working better for this data set.

One idea; have you thought about combining some of your columns? Specifically for both the destination and the origin account, there is a new and old amount. Can you simplify those by have an original amount, and the delta, or change? Even further, is the change always equal to the transaction amount? If it is, you shouldn't need to add the ending balance column. Just including the starting amount and the transaction should do the job.

You also have the origin and destination names. Have you run a quick analysis to see if any of these are repeats? A really quick “network” analysis could just be looking for repeat offenders on doing the fraud.