Why focusing on reducing false positives is not the right strategy in fraud and anomaly detection

Fraud detection and false positives are inextricably linked. Whether it is rule-based or data-driven detection, there is always a large number of signals that prove not to be fraud; the false positives. The biggest problem with false positives is the amount of resources it requires to identify them, whole departments exist just to call customers, to analyze the raw data again, or by working through a protocol. Even the big names as LinkedIn and Facebook, sometimes wrongly detect or block a picture, post, or profile. There are many organizations that focus primarily on reducing false positives, is this wise?

When it comes to fraud, the answer is "no." The problem with fraud and anomaly detection mainly lies in the low prevalence. Consider how many transactions a large organization performs each day? Millions? Hundreds of thousands? And how many of them are fraudulent? In most cases, much less than 1 in 100 (1%).

Simple Bayesian statistics provides insight into the problem. Suppose the risk of fraud of a randomly selected customer or transaction is 1% in a large organization. We take a high quality detection system, assuming that if the customer commits fraud, it provides an alert in 95% of cases. The chance of a false positive with this detection is low, lets define 8%. What is the probability that a customer / order / transaction is fraudulent, given the alert from the detection system?

The answer is a disappointing: 10.71%

So even with a great detection system; a high probability of an alert in the event of fraud and few false positives, almost 90% of the alerts is false. If the prevalence increases from 1% to 5%, and the number of false positives in the detection system goes down from 8% to 5%, then 50% of the alerts is really fraudulent; half of the alerts are completely unnecessary. Please consider that it requires many times more energy, talent, and money to increase the accuracy of the classification from 80-90% to compared to the 70-80% step.

Of course, false positives are an ideal metric, they are easy to understand and quantify. Making it an attractive concept for non-technical staff. In addition, it can be a perfect excuse for a passionate and knowledgeable specialist to apply advanced cluster and classification algorithms (scientists love those). Also, it is likely that data-driven detection is better to reduce false positives.

But the conclusion is simple. In situations with rare events such as fraud, it makes little sense to base the focus of the organization on the quality of the classification.

Use this online tool to make your own calculations and play with the input.