Over het blindstaren op false positives in fraude en anomalie detectie

Fraude detectie en false positives zijn onlosmakelijk met elkaar verbonden. Of het nu rule-based of data-driven detectie betreft, er is altijd een groot aantal signalen die achteraf geen fraude blijken te zijn; de false positives. Het grootste probleem met false positives is dat het geld kost, hele afdelingen worden opgetuigd om klanten na te bellen, de ruwe data nog eens te bekijken, of een protocol door te werken. Zelfs bij de grote namen als LinkedIn en Facebook wordt er wel eens een foto, post, of profiel onterecht gesignaleerd of geblokkeerd. Er zijn dus veel organisaties die vooral focussen op het reduceren van false positives, is dit verstandig?

Als het gaat om fraude is het antwoord “nee”. Het probleem bij fraude en anomaliedetectie zit hem voornamelijk in de lage prevalentie. Beoordeel zelf: hoeveel transacties / handelingen doet een grote organisatie per dag? Miljoenen? Honderden Duizenden? En hoeveel daarvan zijn er frauduleus? In de meeste gevallen veel minder dan 1 op 100 (1%).

Simpele Bayesiaanse statistiek geeft inzicht in het probleem. Stel dat bij een grote organisatie de kans op fraude bij een random gekozen klant / handeling 1% is. We nemen een extreem goed detectie systeem: we gaan ervan uit dat, als de klant fraude pleegt, in 95% van de gevallen het detectiesysteem een alert geeft. De kans op een false positive met dit detectiesysteem is laag, we schatten 8%. Wat is nu de kans dat een klant / aanvraag / handeling frauduleus is, gegeven de alert van het detectiesysteem?

Het antwoord is somber: 10.71%

Dus zelfs met een geweldig detectiesysteem; een hoge kans op een alert in het geval van fraude en weinig false positives, zit bijna 90% van de alerts er gewoonweg naast. Gaat de prevalentie omhoog van 1% naar 5% en het aantal false positives in het detectiesysteem omlaag van 8% naar 5%, dan zijn 50% van de alerts echt frauduleus; de helft van de alerts volslagen onnodig. Neem hierbij in overweging dat het vele malen meer energie, talent, en geld kost om de stap te maken van 80-90% classificatie nauwkeurigheid dan de stap van 70-80%.

Natuurlijk zijn false positives ideaal als focuspunt, ze zijn makkelijk te begrijpen en te kwantificeren. Wat het voor niet-technische medewerkers een aantrekkelijk aanknopingspunt maakt. Daarnaast kan het voor gepassioneerde specialisten een ideaal excuus zijn voor het toepassen van geavanceerde cluster en classificatie algoritmen (*scientists* zijn daar dol op). Het is aannemelijk dat data-driven detectie beter is in het reduceren van false positives.

Maar de conclusie is simpel. In situaties waar het gaat om rare events zoals fraude heeft het weinig zin om de focus van een organisatie te baseren op de kwaliteit van de classificatie.

Gebruik deze online tool om zelf de berekening te maken en de spelen met de input.