Anti-Money Laundering False Positive Solution

Automatically Classifying False Positive Alerts Using Machine Learning

Ashrith Barthur

H_2O.ai

October 16, 2019

Overview

What is the Anti-Money Laundering Solution?

1. The Anti-Money Laundering Solution is an end-to-end Machine Learning-Based False Positive Money Laundering alert detection system.
2. It is deployed in a financial institution to reduce False Positive Money Laundering alerts generated by Rule-Based alerting systems.

Anti Money Laundering

What is money laundering?

"the concealment of the origins of illegally obtained money, typically by means of transfers involving foreign banks or legitimate businesses."

Three Stages of Money Laundering

Anti Money Laundering

How are we solving Anti Money Laundering up until now?

1. We use Rule-Based systems that detect instances of money laundering.

What are these rule-based systems?

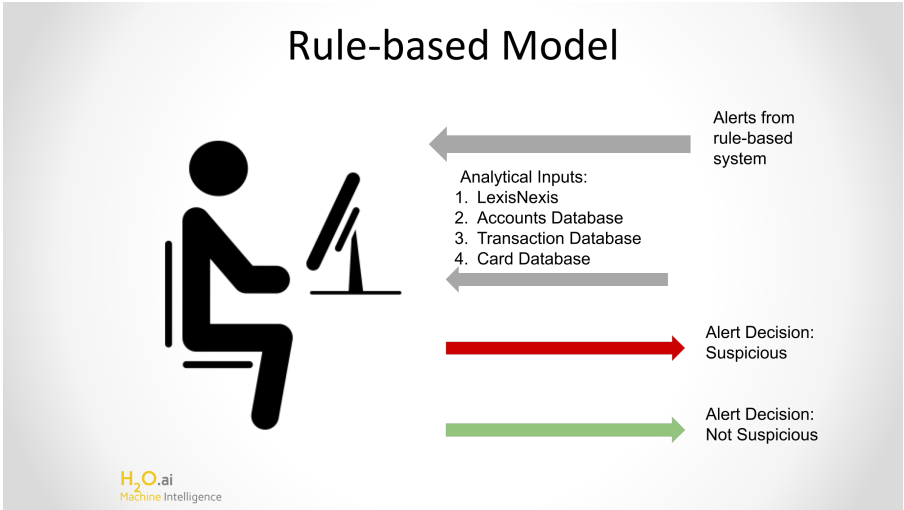
1. FICO
2. fiserv
3. SAS AML
4. Actimize

Anti Money Laundering

If rule-based systems are solving Money Laundering then what remains to be a problem?

1. Process is manual
2. Human intervention required at many steps
3. Alerts have a high false positive rate - 75% to 99%
4. The Rule-Based systems are slow to evolve.
5. Rule gaps, and complexities exist.
6. Rules are stratified, evolution of new rules requires *only* AML systems to design these rules.

Current Work Flow



Current Work Flow

The steps in the current rule-based workflow are:

1. An alert is generated by the alerting system (mostly false positive - 75% to 99%)
2. The investigator reviews it.
3. The alert is approved as True Positive, or classified as False Positive

What Does the Anti-Money Laundering Solution Do?

1. Fundamentally, it reduces false positives.
2. The solution is strategically placed between the AML system and the investigator.
3. It classifies alerts as False Positive, or True Positive, using an out-of-loop ML approach.
4. A curated set of alerts are given to the investigator.

What are the advantage of the AML Solution?

1. Speed and Speed - Usual investigation time is about 45 - 90 days. The solution cuts it down to seconds.
2. It reduce human inaccuracies.
3. It reduce required person-hours.
4. It fills rule-gaps by innovative features.

What is the *basic* modeling requirement to successfully use the AML Solution?

1. The AML Solution requires the problem to be defined as a *Supervised Machine Learning problem.*
2. Therefore for training a model, the Solution requires historic alerts to be pre-marked as False Positives, or as True Positives, by the investigators.

What are the data sources required by the AML Solution?

1. Pre-marked AML Alert Data
2. Banking Transactional Data
3. Banking KYC (customer) data

Demo

How is the AML Solution deployed?

1. The AML Solution does not modify any ***existing*** AML alert generating system.
2. FIs invest considerable resources around Rule-Based AML alert generators, and replacing them with a complete ML-Based solution is slow. Therefore the AML Solution works around these Rule-Based alert generators.
3. The AML Solution ingests Alerts from the database table (or a copy of,) that the AML alert generating system writes to (e.g. SAS).
4. It then applies the AML specific feature transformations and creates a ML-Based model using DriverlessAI.
5. Using the ML-Based model, alerts generated by the Rule-Based AML alert generating systems are classified as True Positve, or False Positive.
6. This classification is written back to the database for further investigation or to create a Suspicious Activity Report.

Things to Remember

1. The ability of the model is dependent on the risk appetite, which is dependent on the False Negative percentage.
2. High risk appetite can give better False Positive and True Positive classification, but will also increase the False Negatives.
3. The ability of the model will also depend on the quality and size of the data.

Thank you & Questions