Summer Intern Research Project: The Use of Machine Learning in AML

Grace Stapkowski

August 2024

Contents

1 Introduction

AML (Anti Money Laundering) is an international group of regulations, laws, and procedures to uncover money that has been disguised as legitimate income. This plays a pivotal role in a variety of services provided by A&M as well as any financial services institution.

The main two parts of AML are to (1) know your customers and their risk profiles, and (2) monitor and report suspicious behavior to overall identify, evaluate, and mitigate money laundering risks. These components can detect money laundering at any stage at a variety of levels. With the rise of AI/ML, more criminals are able to use advanced algorithms at any stage of money laundering. This means that financial institutions must leverage the same technologies to be able to detect this crime most effectively. In machine learning, there is client risk profiling and suspicious behavior flagging.

This document offers a condensed, high level overview of the machine learning techniques commonly used for client risk profiling and suspicious behavior flagging for those who yearn for LaTeX write-ups.

2 Client Risk Profiling

The main goal of client risk profiling is to assign each client a risk score, looking at their accounts and transactions broadly as one unit.

Let $x_c \in \mathbb{R}^d$ be a vector of d features specific to client c. Features can include any fact about the client, such as volume of transactions made in sanctioned countries, average transaction amount, number of accounts, etc.

Let $P = \{L, M, H\}$ be a generic set that measures risk (L for low risk, etc.) Note the set P can contain other risk levels/indicators and can be larger or smaller depending on the use case.

To obtain the money laundering risk for a client c, we apply the function

$$p : \mathbb{R}^d \to P \tag{1}$$

to the feature vector x_c by performing $p(x_c)$). This function is depicted in a simplified form in Figure 1 below.

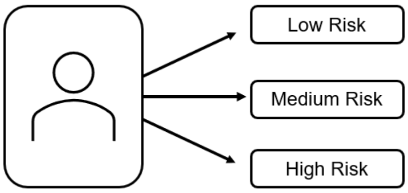


Figure 1: Client Risk Profiling

3 Suspicious Behavior Flagging

Suspicious behavior flagging works at a lower level than client risk profiling, typically focusing on the transaction level (although it can be used at higher levels as well).

Assume a client c has $a = 1, ..., A_{(c)}$ accounts where each account (c, a) has $t = 1, ..., T_{(c,a)}$ transactions.

Thus we have a vector $x_{(c,a,t)} \in \mathbb{R}^d$ specific to a transaction (c, a, t). In this case, instead of mapping to a generic set P, we map to the binary set $\{0, 1\}$ (i.e.{no flag, flag}).

Thus our AML suspicious behavior flagging uses the function $s : \mathbb{R}^d \to \{0, 1\}$ where $s(x_{(c,a,t)}) = 1$ indicates that a flag was raised on the transaction (c, a, t).

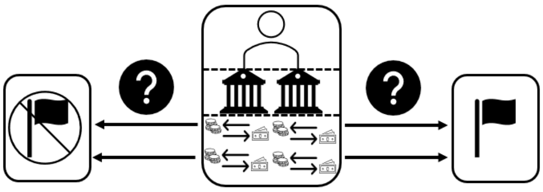


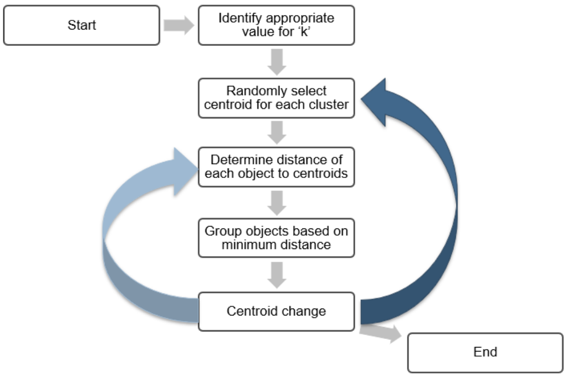
Figure 2: Suspicious Behavior Flagging

4 Unsupervised Algorithm: K-Means Clustering

K-means clustering is a commonly used algorithm for clustering data points. In AML, it can be used to cluster clients or transactions to assess risk. This algorithm is unsupervised, which means it does not use training data.

4.1 General Algorithm

Figure 3 below shows the general steps of the k-means clustering algorithm as shown in the presentation. The darker blue arrow represents the outer process that repeats only when the lighter blue inner process runs enough times for the clusters to converge.



Start
Identify appropriate value for 'k'
Randomly select centroid for each cluster
Determine distance of each object to centroids
Group objects based on minimum distance
Centroid change
End

Figure 3: K-Means Clustering Algorithm Steps

Step 1: To begin, the algorithm must know the value of 'k'. One possible way to determine 'k is to increase the value until the within cluster variance (WCV) starts to plateau. Typically, the WCV will decrease as 'k increases, but there is always a point where the decrease in variance becomes less severe. (See: elbow method).

Step 2: After the value of 'k' has been selected, the algorithm will randomly select 'k' "centroids" (i.e. "centers") to make 'k' clusters.

Step 3: Once the centroids have been selected, the Euclidean distance between each data point and centroid is calculated (i.e. point one to centroid one is calculated, then point one to centroid two, until point one to centroid 'k' is measured. Then point two to centroids 1,...,k and so on).

Step 4: From there, each data point is assigned to the centroid that it is closest to.

Step 5: Once the assignment is completed, the centroid is then recalculated based on the cluster assignments to be the true mean of the cluster. Steps 3-5 are repeated until the centroid no longer changes. When this point is reached, the algorithm returns to step 2 to repeat steps 3-5 again with new randomly selected centroids.

The user controls how many times steps 2-5 repeat (i.e., how many different clusterings the algorithm will attempt). The choice for this number of iterations depends on a variety of factors such as data size and computing power.

After all outer iterations have run, the best clustering is determined by evaluating the within cluster variances. The within cluster variance (WCV) can be found through solving the following function.

$$WCV = \sum_{k=1}^{K} \sum_{p_k} ||x_c - \mu_k||^2 \tag{2}$$

where μ_k is the center of cluster 'k'. That is, the algorithm is looking for the clustering with the minimum squared Euclidean distance between every point and its respective centroid.

4.2 Application of K-Means in AML

4.2.1 Client Risk Profiling

Client risk profiling can be done with k-means combined with decision trees. An example I came across was a k-means clustering with $k = 7$, 2 of the 7 clusters being risky. From there, decision trees can be used to find and spell out the classification rules that can recreate the clusters. (This is more to make the analysis more digestible for bank tellers).*

4.2.2 Suspicious Behavior Flagging

For suspicious behavior flagging, k-means can be combined with principal component analysis (PCA). At a very high level, principal component analysis takes data points/transactions with many features and compresses them into a lower dimensional data set, which then can be clustered using k-means to identify variance between the data. (Here is a great video from statquest on PCA).*

*Note: there are many possible combinations of algorithms that can produce strong results. The ones noted here are the ones that I found easiest to understand.

5 Supervised Algorithm: Logistic Regression

Logistic regression is a commonly used algorithm for binary classifications (e.g., true/false, yes/no, and in our case flagged/not flagged). The theory for this algorithm is heavily based on probability;

specifically conditional probability. This algorithm is supervised, which means that it requires historical training data.

5.1 General Algorithm

The algorithm models the equation

$$p(X) = Pr(Y = 1|X) \tag{3}$$

where

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \tag{4}$$

In this general case, the algorithm seeks the conditional probability that the output is 1 given the variable X and the logistic regression line fit on historical data. For AML, it aims to predict the probability that a given transaction or client should be flagged given its feature vector and historically flagged or not flagged training data.

Diving deeper into the equation for $p(X)$, β_0 and β_1 are coefficients that can be estimated via maximum likelihood using numerical methods. In the logistic regression equation, β_0 can be thought of as the intercept and β_1 as the slope. To see this more clearly, we can rearrange the function for $p(X)$.

$$log(\frac{p(X)}{1 - p(X)}) = \beta_0 + \beta_1 X \tag{5}$$

Again, once the model is trained with historical data, it can predict the probability that a transaction or client should be flagged. The probability is a number between 0 and 1. For AML, a financial firm might want to flag any client or transaction with a probability > 0.75.

5.2 Application of Logistic Regression in AML

5.2.1 Client Risk Profiling

For client risk profiling, logistic regression can be combined with social network analysis. The social network analysis is done first, building graphs of client information and relationships. The logistic regression is then done on the resulting graph(s).*

5.2.2 Suspicious Behavior Flagging

For suspicious behavior flagging, logistic regression can be combined with gradient boosted trees. Logistic regression is performed first with the purpose of filtering out clearly non-illicit activity. From there, gradient boosted trees are used to flag transactions.*

*Note: there are many possible combinations of algorithms that can produce strong results. The ones noted here were the ones that I found easiest to understand.

6 Sources

Fighting Money Laundering with Statistics and Machine Learning by Rasmus Ingemann Tuffvenson Jensen and Alexandros Iosifidis

Machine Learning in Fraud Detection and AML — The Sumsuber

Machine learning for Anti-money laundering - KPMG Belgium

AML risk-rating models — McKinsey

Statistical Classification - Wikipedia

Statistical Learning: 4.1 Introduction to Classification Problems by Stanford Online

Statistical Learning: 12.3 k means Clustering bY Standford Online

KMeans — scikit-learn 1.5.1 documentation

K-Means Clustering in Python: A Practical Guide – Real Python

LogisticRegression — scikit-learn 1.5.1 documentation

Logistic Regression in Python – Real Python

Gradient Boosting - Wikipedia

Social Network Analysis - Wikipedia

Principal Component Analysis - Wikipedia

Decision Tree - Wikipedia