Paper 3637-2019

Improve your Anti-Money Laundering Monitoring While Reducing Costs with SAS® Viya®

Calvin Crase and Chris St. Jeor, Zencos Consulting

ABSTRACT

Businesses required to comply with Bank Secrecy Act and Anti-Money Laundering (BSA-AML) regulations are very familiar with the burden of false positive alert generation. Of the thousands of potential money-laundering alerts generated each month, only a small handful are positively identified as requiring additional investigation. According to industry statistics, over 95% of system-generated alerts are false positives, and nearly 98% never result in a suspicious activity report (SAR). These false alerts cost billions of dollars annually in wasted investigation time.

The conventional rules-based detection approach cannot keep pace with today's constantly evolving money laundering topologies. By leveraging the predictive power of SAS® Viya® Visual Data Mining and Machine Learning (VDMML), financial institutions can reduce risk by reassessing and enhancing their detection strategies with a data-driven, risk-based approach that is recommended by regulators. This paper will demonstrate how VDMML can allow users to quickly build Machine Learning models to predict which alerts will generate productive investigations and elevate your company's core AML detection process.

INTRODUCTION

Those familiar with the financial sector know the responsibility placed on financial institutions to participate in BSA-AML regulations. Current AML transaction monitoring consists of complicated business logic that tracks demographic and financial transactions to identify potential money laundering practices. Any combination of financial transactions that fall within the specified business logic is flagged for alerts and require further investigation.

The problem with this approach is that thousands of routine financial transactions are flagged each month that ultimately lead to unproductive alerts. These unproductive alerts result in billions of dollars spent each year with analysts spending increased amounts of time conducting investigations, and the problem is only getting worse. Research by McKinsey & Company shows that resources dedicated to AML compliance at major banks in the US have increased tenfold over the last 5 years (Stuart Breslow, 2017). While rule-based monitoring is the current industry best practice, it does not need to remain the only solution to AML monitoring.

This paper walks through how financial institutions can use the data created through the alert generation process and use predictive models to identify which alerts can be ignored and which should be investigated. We use mocked data built to replicate what the typical AML solution would have readily available and provide examples of potential predictive models that could be used to couple with the current rule-based approach.

WHAT CAN BE DONE ABOUT MONEY LAUNDERING?

Money laundering is "the concealment of the origins of illegally obtained money, typically by means of transfers involving foreign banks or legitimate businesses." (Oxford University Press, 2019)

These activities have devastating social consequences, the effects of which provide "the fuel for drug dealers, terrorists, arms dealers, and other criminals to operate and expand their criminal enterprises." (Financial Crimes Enforcement Network, 2019) For these reasons, financial institutions are charged with the responsibility to employ certain strategies to detect money laundering and uphold the integrity and stability of our economy.

DEFINING STRUCTURING SCENARIOS

As part of the AML process, financial institutions employ business logic rules (scenarios) that flag for certain behaviors we typically associate with potentially illicit financial activity. The process generally consists of looking at all the transactions at an entity level you are interested in monitoring, such as a bank customer. We take all the transactions associated with a customer for some predefined period and put them through a set of rules to determine if any of their behavior could be considered suspicious.

If it triggers on one of our scenarios, we say it has "alerted," and an analyst must determine whether this entity requires further investigation, or if it is a false positive. After an alert is investigated, if the analyst finds there is sufficient evidence that the transactions are suspicious then the alert has a SAR filed and is passed to the federal government for additional investigation. The specific scenario we consider in this paper is a structuring scenario.

The Federal Financial Institutions Examination Council defines structuring as a person or persons conducting or attempting to conduct one or more transactions in currency, in any amount, at one or more financial institutions, on one or more days, in any manner, for the purpose of evading the reporting requirements (FFIEC, 2004).

One such report is a currency transaction report (CTR), which must be filed to be evaluated by the FinCEN if a bank customer deposits $10,000 or greater. To avoid reporting, a customer may use "structuring" to make a series of different deposits over a short period that aggregate over $10,000 but are each individually less than $10,000. This behavior seeks to avoid detection from the mandatory CTR filing.

DEALING WITH FALSE POSITIVES

While the rule-based transactional monitoring is considered the industry best practice, the problem with this standalone solution is that roughly 95 percent of the time these alerts are false positives, which can be costly to a financial institution because they require valuable resources to investigate (PWC, 2010). Billions of dollars are spent each year investigating alerts that never needed to be investigated. While the goal is to create logic that captures suspicious activity, the inherent problem is that actual money launderers are very good at making their illicit activity appear similar to the billions of legal transactions happening daily.

To combat this problem, industry experts propose that financial institutions incorporate predictive models to complement the current rule-based process to filter out the noise and identify the actual illicit activity. To keep up with increased federal regulation and the responsibility of AML monitoring, McKinsey & Company have found that "Leading banks are trying to crack these problems by turning to new technologies. Machine learning, real-time data-aggregation platforms … offer a fundamentally new approach to managing

compliance." (Stuart Breslow, 2017) SAS Viya VDMML offers an end-to-end solution that banks can adopt to address the problems mentioned above and harness the predictive power of machine learning to complement their current AML processes.

USING THE POWER OF SAS VDMML

SAS VDMML is a web-based analytics package that offers everything from exploratory analysis to robust predictive analytic models. SAS VDMML offers an end-to-end solution for complicated business problems, allowing users to explore data and build models without having to write a single line of code. Work can easily be shared and collaborated on across business units and brings new unparalleled levels of efficiency to the business process. Some of the primary benefits of SAS VDMML include:

- Web-based GUI platform
- Provides end-to-end solution for complicated business problems
- In-memory distributed process for speedy results
- Collaborate across business units
- No coding needed
- Wide variety of predictive models
- Compare models quickly and export SAS score code to operationalize a champion model

TO INVESTIGATE OR NOT TO INVESTIGATE: IT'S A BINARY QUESTION

The problem we are trying to solve is to predict whether an alert generated by our AML structuring scenario described previously results in a productive alert. While the problem on the surface may seem complicated with several contributing factors, the question itself is quite simple. Will this alert lead to a productive alert? Yes or No?

This analysis is called binary supervised learning. For these types of binary problems, we can use a variety of modeling approaches, each ranging in predictability and interpretability. The goal for this type of analysis is to identify known variables or features that we can use to predict an unknown outcome, answering, "Will this alert be productive?"

UNDERSTANDING THE DATA

For our purposes, we created sample data that replicate what we would use for a common structuring AML scenario. In practice, however, you must track each alert and record whether an investigation was initiated (a binary indicator of Yes or No, 1 or 0, etc.) and then tie the results back to the actual scenario data. To build a predictive model, you must use data for which you know the outcome, then use that model to predict data for which you don't know the outcome.

Our models use mocked variables that would be fed into our structuring scenario. It is crucial to use predictor variables in your model that can easily be tracked and are readily available without having to know what the outcome is that you are trying to predict. It is difficult to operationalize a model that requires data that is not readily available.

The models we built use common structuring variables as well as other entity level data found in the SAS AML Solution. The variables we are using are defined as follows:

- **Productive_Alert (Target):** Binary indicator for whether an investigation was initiated.
- **Party_ID**: Party number to track historical alerts and transactions. We do not use this in the actual model but simply as a party key.
- **Number_of_Alerts_Generated:** Numeric variable that aggregates the total number of alerts generated for a given party key.
- **Time_Between_Transactions:** Numeric variable that aggregates the amount of time between transactions.
- **Past_CTR:** Binary indicator for whether the person has generated a past Currency Transaction Report.
- **Number_of_Transactions:** Numeric variable that aggregates the total number of cash transactions that were included with the alert.
- **Currency_Amount:** Numeric variable that aggregates the total amount transferred for the alert.
- **Cash_Intensive_Business:** Binary indicator for whether the initiator of the transaction is a cash-intensive entity.

CHOOSING THE CORRECT MODEL

The first step in building a model is to know what you want to get out of your model. It's not enough to simply make a prediction – what you will do with the prediction is critical in choosing the type of model to use.

PREDICTABILITY VS INTERPRETABILITY

When selecting the type of model to use you need to start with a fundamental question. What is more important? Predictability (the accuracy with which we make our predictions) or interpretability (the ability to explain why we made the prediction we did)? In the world of analytics, we often hear much discussion about the differences in predictive power between different types of models. Far too many projects die because the creators forget to account for the differences in interpretability across models. While it would be ideal to choose a model that has high predictability and interpretability, we, unfortunately, cannot have our cake and eat it too. So, we must decide what is more important for the end use of our model – the accuracy of the prediction or the interpretability?

A simple way to view this relationship between model predictability and interpretability is on the analytic spectrum chart in figure 1. While it would be great to have a model that fits in the top right quadrant of the chart, in reality, models usually fit somewhere in the top left or

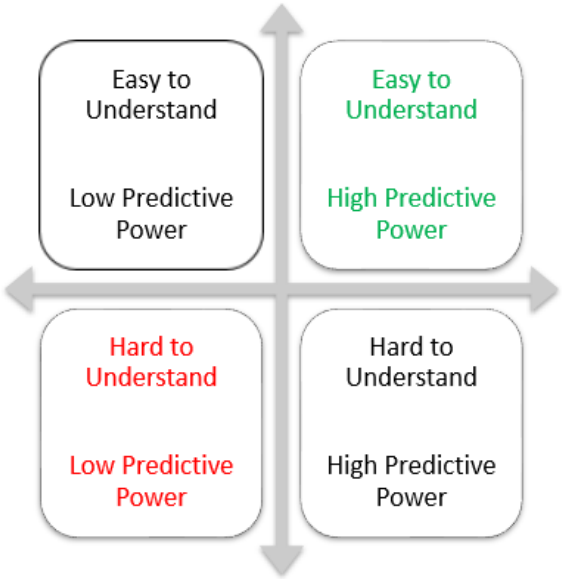

Figure 1 Analytic Spectrum

bottom right quadrants of the graph. Models like decision tree and logistic regression existing in the top left, and gradient boosting and neural networks models in the bottom right of the spectrum.

To help create some context around this idea let's play a little game we like to call story time.

Story One: You Want Predictability

Let's say you are heading to Vegas for the weekend and you want to lay some money down on a Dallas Cowboys game. You are a huge Cowboys fan, but let's be honest, this is your money on the line, and you want to make sure you are betting on the correct outcome. All you care about is whether the Cowboys are going to win. It doesn't matter if Dak throws six picks or if Zeke runs for 3 touchdowns. All that matters is that your money comes back to you with interest. In this case, you would want to use a highly predictive model, like a gradient boosting model, and ignore the interpretability of the model entirely.

Story Two: You Want Interpretability

For this story let's pretend you are the manager of the data science team at a large hospital and you are trying to build a model to identify patients that are at risk of developing high blood pressure, which helps you better care for your patients. Just telling a patient that they have a 75% chance of developing high blood pressure is not enough. You need to be able to tell them why they are at risk and what they can do to avoid it. For this type of problem, you want a model that can be easily interpreted and used to create an actionable plan to help your patients avoid the problem altogether. Logistic regression would be a great candidate for this type of scenario.

In summary, when selecting the model to use, you first need to decide what is most important: the ability to understand the prediction or the accuracy of the prediction? Answering this question up front helps set the project up for success because a model is only as good as the end user's ability to use it.

JUDGING MODEL PERFORMANCE STATISTICS

We will walk through some models available in SAS VDMML and how to use them later in this paper. Let's first consider how to evaluate the performance of your models. For most modeling approaches, the user rarely knows ahead of time all the variables they want to use or even the type of model they want to use. To accurately assess the performance of your modeling solution and chose a champion model from your list of candidate models, you need to decide on a standardized approach that allows you to compare apples to apples.

There are a host of statistics (R^2, AIC, BIC, and so on) to compare statistical models and their predictive performance. For simplicity's sake, we are going to use the misclassification rate, which essentially tells us how often our predictions are incorrect. The lower the number, the better the model is at making predictions. This is a performance statistic available across all binary predictor models and is very straightforward. Bear in mind that the misclassification rate works well when you have an even split in the target that you are trying to predict. If you have a heavily disproportional split in your target and you want to use misclassification rate, you may want to oversample your data before building your model.

PREDICTING PRODUCTIVE ALERTS

SAS VDMML has a variety of binary predictive models you can choose. This section of the paper discusses some basic models available in SAS VDMML and the pros and cons of each model.

The models demonstrated in this section include logistic regression, decision trees, and gradient boosting.

USING LOGISTIC REGRESSION

Logistic regression is one of the most widely used predictor/classification models. While the underlying math can seem a little complicated, in its purest form logistic regression attempts to find the relationship that each of the predictor variables has with the variable you are trying to predict. It then uses those relationships to calculate the overall probability of an event happening for a given observation.

The feature that makes logistic regression so attractive when compared to other models is the calculated relationship each predictor variable has with the target. Each variable is assigned a coefficient value which represents the log odds for a one-unit change in the value of the variable. So, say for example you are predicting whether the Cowboys are going to win on Sunday. If quarterback passer rating is one of your predictor variables and it has a coefficient of 1.34, this would mean that for each additional percentage increase in quarterback passer rating, the Cowboys are 1.34 times more likely to win the game. This analysis allows the end user not only to understand the end prediction of a win or loss but also to understand the specific effect each predictor variable has on the target variable and the final prediction.

SAS VDMML allows you to build logistic regression within the drag-and-drop interface. Using the scenario data previously discussed, we can quickly build a logistic regression model to predict if an alert was productive. The logistic regression model and the model's misclassification rate of 11% can be viewed in figure 2.

Figure 2 Logistic Regression Output from SAS VDMML

You can quickly see which variables are significant in your model as well as the cumulative lift and residual plots for your model. Using the output above, you can quickly assess the significance of each variable in the model. Coupling this output with the user's industry expertise can help expedite the exploration of other potential variables with previously unknown correlations. The ability to quickly adjust and rerun models without having to

manage a single line of code dramatically decreases the amount of time spent on other modeling processes.

USING DECISION TREES

Decision trees are one of the most straightforward statistical models you can use for binary predictions or classifications. A decision tree uses a host of predictor variables (both continuous and categorical) and identifies the best splits of those variables that create the "purest" splits of the data. This is done through an iterative process until the specified conditions of the model are met. With each split, the target variable is forced down one of two branches. The goal of each split is to get the greatest separation of the target as possible. Following the path of the final bins created through the splits of the predictor variables provides a logical path that provides valuable business logic and insight into why the prediction was classified the way it was.

Using our sample data, we can quickly create a decision tree in SAS VDMML. Once you specify which variable to use as the target and select the predictor variables you want to use, SAS VDMML automatically creates the model for you. The output can be seen in figure 3 with a misclassification rate of 11%:

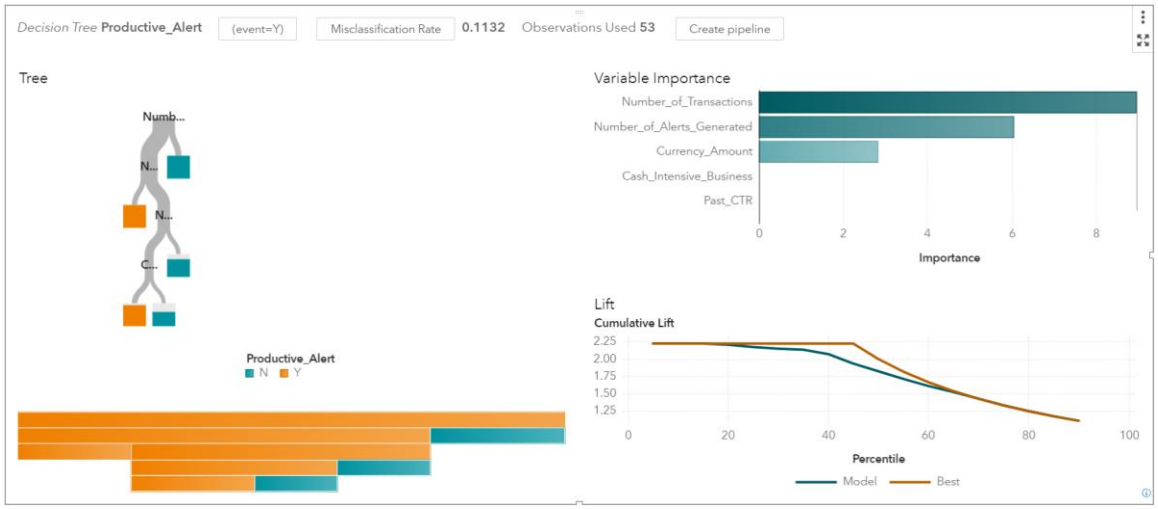

Figure 3 Decision Tree Output from SAS VDMML

Following the tree path, you can see groups of alerts that are classified as either productive or not productive. If you hover over a node a pop-up window appears as shown in figure 4.

Interpreting the window in figure 4, alerts with more than 5 transactions and that have generated more than 15 alerts, creates a group of observations that have 100% productive alerts. Coupling this type of logic generated through a decision tree with the rule-based scenario can add a great deal of insight into which alerts are most likely productive. Alternatively, an additional step could be to use this insight to create a risk rating for which alerts should be investigated.

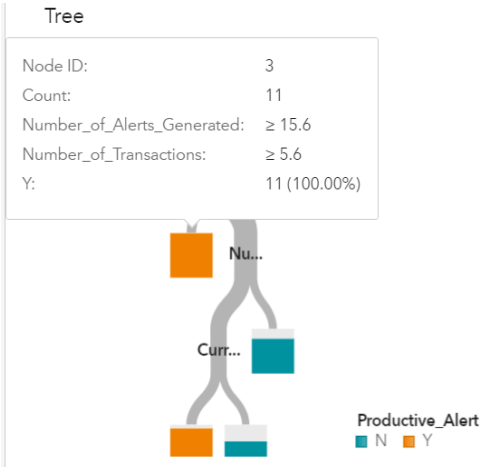

Figure 4 Decision Tree Pop-Up Window from SAS VDMML

USING GRADIENT BOOSTING

Gradient boosting models are some of the most powerful predictive models used today. While they are widely used, most people view them as a black-box solution. So, before we discuss the gradient boosting model we used to predict productive alerts, let's take a minute to understand what is going on under the hood. A straightforward way to think of gradient boosting models is depicted in one of my kids' favorite new movies: Disney's *Ralph Breaks the Internet* (spoiler alert – Ralph nearly breaks the internet). For those who haven't seen the movie: Ralph is a huge, strong guy who by himself is fairly powerful. In the movie, a weaker and less intelligent version of himself gets cloned about a million times. While each clone isn't much to worry about, when all the weaker and less intelligent clones converge together they become a massive unstoppable rage monster. It's quite intense.

A simple way to think of a gradient boosting model is to picture a collection of weaker decision trees put on steroids. Gradient boosting models are a collection of miniature, weaker decision trees, each built on a different subset of your total data. The model iterates through the entire data set and takes weighted samples for each model. The goal is to give higher weights to observations that are difficult to predict and lower weights to observations that are easier to predict. The final model is essentially an ensemble of all the "weak" prediction models which creates an unstoppable rage monster that can predict both easy and difficult observations with incredible accuracy. As a result, gradient boosting is one of the most powerful predictive models used today.

By creating several "weak" decision trees on weighted samples of the same data used as the other two models, the first attempt for our gradient boosting model has a misclassification rate of 3% (8% lower than the other models) as can be seen in figure 5.

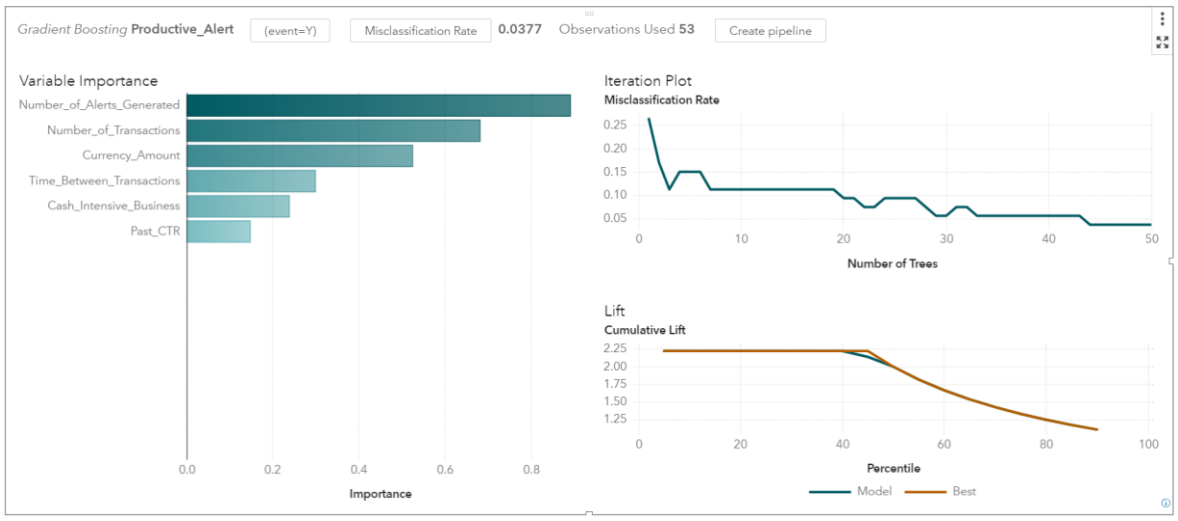

Figure 5 Gradient Boosting Output from SAS VDMML

One of the tradeoffs of gradient boosting is that while you can get a much lower misclassification rate, they do not have the nice interpretability of logistic regression or a decision tree model. SAS VDMML, however, provides variable importance showing which variables had the most impact on the actual model. Gradient boosting models would lie in the bottom right quadrant of the analytic spectrum discussed earlier, offering high predictability but little interpretability behind the prediction.

USING MODEL GOVERNANCE

Another nice feature SAS VDMML provides is the model governance feature. Most projects don't have a predefined model – the user does not often know ahead of time which

predictor variables to use. The process many projects follow results in several candidate models being built, and then a champion model is selected based on each candidate model's predictive ability on *hold-out* data. Hold out data is critical to the modeling process to make sure you aren't just choosing a model that was overfitted to the specific data it was built on. You want a model that will provide accurate predictions for any new incoming data.

SAS VDMML can perform model comparisons on hold-out data and select the champion model based on the preferred model performance statistic and export SAS code for the selected model to score new incoming data. The model comparison output can be seen in figure 6.

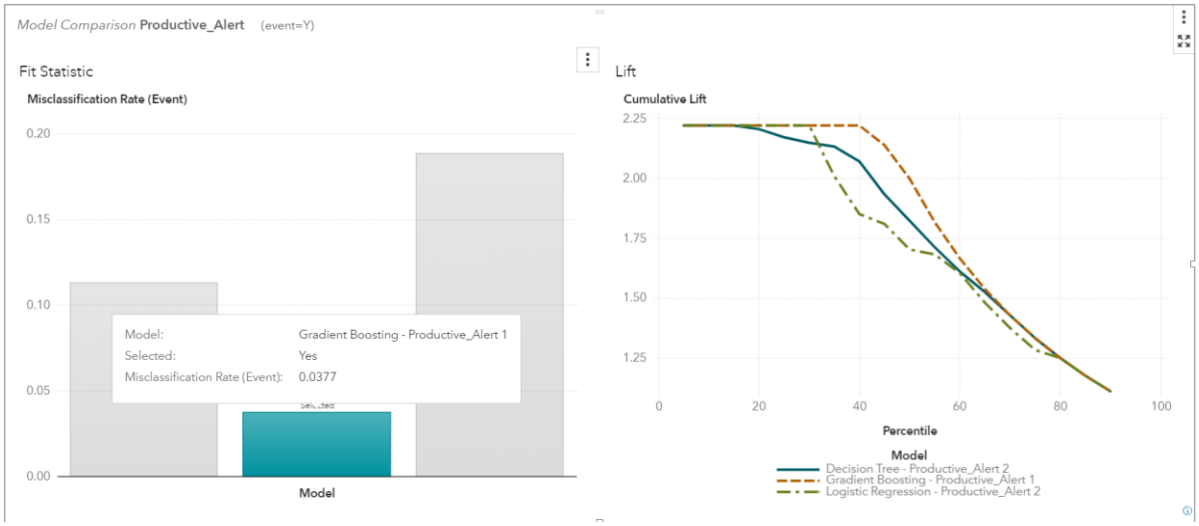

Figure 6 Model Comparison Output from SAS VDMML

After scoring each of the models on new hold-out data not previously used, the gradient boosting model still performs best while the logistic regression model (bar chart on the right) performs slightly worse.

Once you have a model you are comfortable with and that you are ready to put into production, SAS VDMML generates and exports the SAS code for the model so that you can score new alerts generated. You can dramatically save time and resources not investigating unproductive alerts. An example of the code export capabilities can be seen in figure 7.

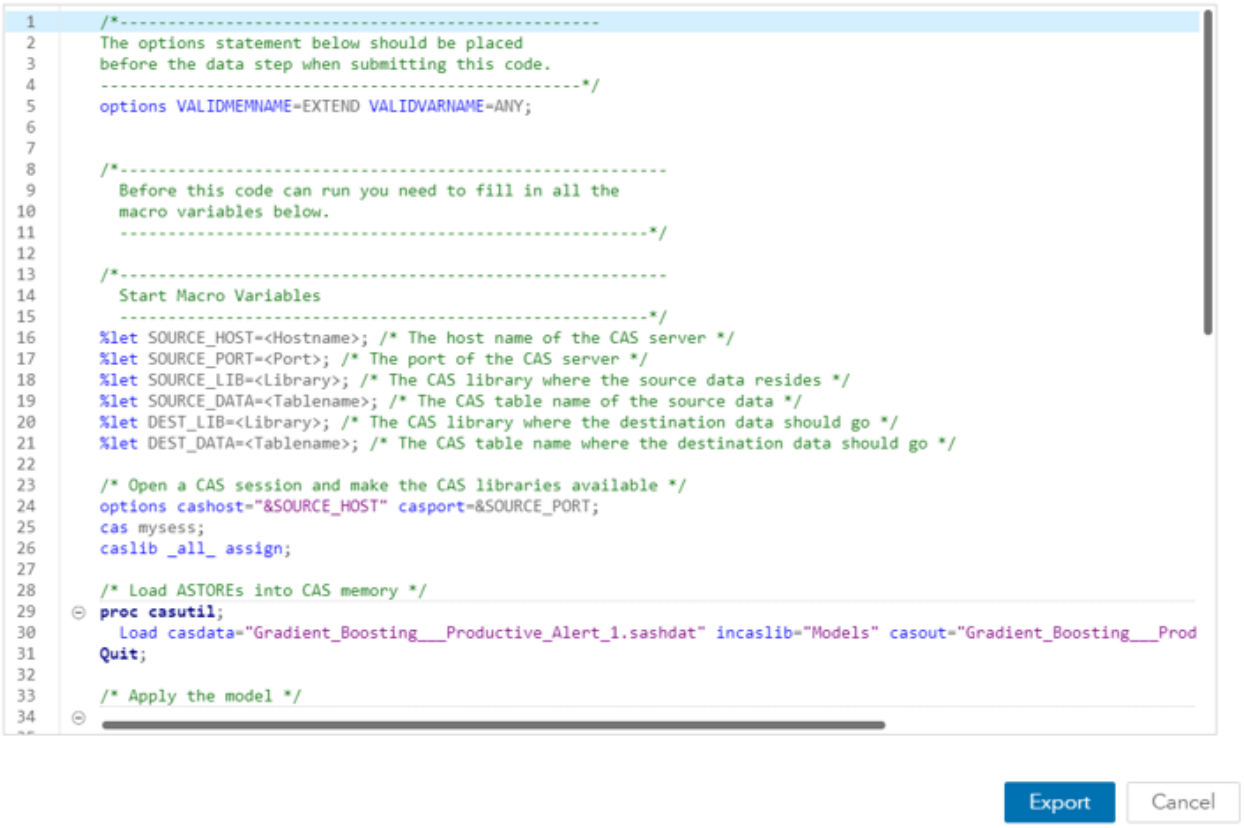

Figure 7 Model Code Export Window from SAS VDMML

CONCLUSION

While the standard rule-based logic for identifying potential money launderers is a critical aspect of quality BSA-AML practices, businesses need to find additional methods to supplement the investigation process. The powerful modeling capabilities of SAS VDMML allow financial institutions to quickly create predictive models to complement the rule-based logic.

When building predictive models, it is essential to keep in mind the end user and how they will use the model being built. For regulated industries, like those using AML monitoring, stronger consideration should be given to the overall interpretability of the model. While this may lead to less predictive models, even small gains in the alert process can save considerable resources in not chasing ghosts through the noise of the thousands of alerts currently being generated.

Financial institutions can use the output from the models to reduce the high levels of false positive rates. Over time, coupling standard scenario tuning with predictive models, financial institutions can dramatically decrease the amount of time and resources spent chasing false positive alerts. Companies that set achievable benchmark goals will make significant strides in their BSA-AML monitoring process and reset the bar for industry standards.

REFERENCES

FFIEC. (2004, December 23). *MONEY AND FINANCE: TREASURY*. Retrieved from Federal Financial Institutions Examination Council: https://www.ffiec.gov/bsa_aml_infobase/pages_manual/regulations/31cfr103.htm

Financial Crimes Enforcement Network. (2019). *What is money laundering?* Retrieved from https://www.fincen.gov/what-money-laundering

Oxford University Press. (2019). *English Oxford Living Dictionaries*. Retrieved from https://en.oxforddictionaries.com/definition/launder

PWC. (2010, September). *From Source to Surveillance* . Retrieved from Price Waterhouse Cooper: https://www.pwc.com/us/en/anti-money-laundering/publications/assets/aml-monitoring-system-risks.pdf

Stuart Breslow, M. H. (2017, November). *The new frontier in anti–money laundering*. Retrieved from McKinsey & Company: https://www.mckinsey.com/business-functions/risk/our-insights/the-new-frontier-in-anti-money-laundering

ACKNOWLEDGMENTS

All papers need collaboration and insights from talented and intelligent individuals. This one was no exception. We appreciate the aid of our thoughtful Zencos co-workers. Any oversights belong to the authors.

RECOMMENDED READING

- SAS VDMML Documentation: https://support.sas.com/documentation/prod-p/vdmml/index.html
- SAS AML Documentation: https://www.sas.com/en_us/software/anti-money-laundering.html
- Data Science for Business: What you need to know about data mining and data-analytic thinking: https://www.goodreads.com/book/show/17912916-data-science-for-business

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Chris St Jeor
cstjeor@zencos.com

Calvin Crase
ccrase@zencos.com

Zencos
919-460-5500
http://www.zencos.com