SAS® GLOBAL FORUM 2020 #SASGF

Paper SAS4553-2020

Deploying Machine Learning Models in an Anti-Money Laundering (AML) Program

Beth Herron and Saurabh Duggal, SAS Institute Inc.

ABSTRACT

As expectations of artificial intelligence (AI) and deep learning have peaked in the financial services industry, anti-money laundering (AML) professionals are exploring advanced methods to more accurately identify suspicious activities that impact their institutions. Fueled by regulatory guidance in the 2018 Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing, many have set up pilot programs, but few have moved into production. Creating a modern AML platform that can support rapid and automated deployment of models and traditional rule-based scenarios ensures that banks will evolve to keep pace with sophisticated financial crimes. This paper explores use cases for machine learning models in AML and provides examples of how clients are promoting advanced analytics from the sandbox into their production SAS® Anti-Money Laundering software environment.

INTRODUCTION

Financial institutions dedicate substantial resources in the area of financial crimes compliance. In 2019, Celent estimated that spending reached $8.3 billion and $23.4 billion for technology and operations, respectively. This investment is allocated toward ensuring anti-money laundering (AML) and counter terrorist financing (CTF) compliance. Money Laundering is the process of making illegally gained earnings appear legal. Terrorist financing is the process of funding the use of violence or intimidation in the pursuit of political gain, regardless of the legitimacy of the source of funds.

The following bar chart shows projected global AML-KYC spend in 2019:

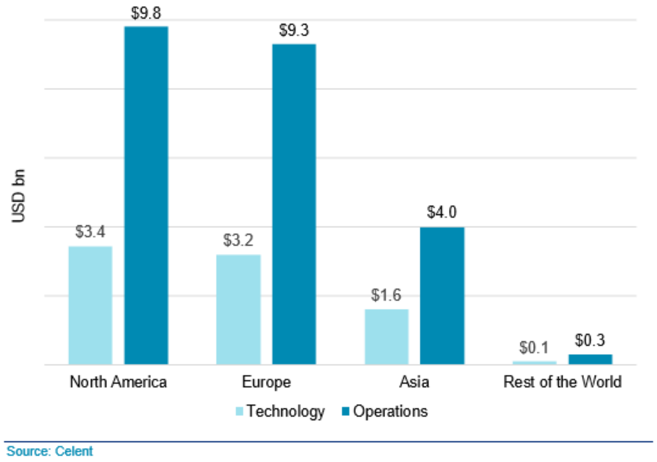


Figure 1. Global Spending on AML-KYC Operations (Celent)

Traditional methods of combating financial crimes are not keeping pace with the disruption occurring in the banking and financial sector. Faster payments, digital first strategies and the ever-rising cost of compliance are accelerating the change. Fueled by regulatory guidance in the 2018 Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing (Federal Reserve), many have set up pilot programs, but few have moved into production. Some innovation barriers observed within financial crimes compliance are:

- Limited Data Science skills
- Difficulty in explaining machine learning models
- Lack of integration capability with incumbent systems

With the help of SAS® Financial Crimes Analytics, everyone is enabled to build explainable machine learning models (including data scientists, business analysts, and developers) and operationalize so that the investigators can act quickly.

MACHINE LEARNING STRATEGIES

Financial institutions have been leveraging automated transaction monitoring strategies to combat money laundering for years, but machine learning techniques are gaining traction. Machine Learning is based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention. The most commonly used branches of machine learning are supervised and unsupervised learning.

SUPERVISED LEARNING

Supervised learning algorithms are trained to detect money laundering that leverage known historical outcomes. For example, the automated transaction monitoring rules that are run today are generating alerts reviewed by investigators. The machine can leverage the suspicious activity report (SAR) flag (‘yes’ or ‘no’) to predict future outcomes by considering complex relationships in the customers behavior.

UNSUPERVISED LEARNING

Unsupervised learning techniques leverage unlabeled data, learning from the structure of the data itself. In this case, there is no historical outcome or "right answer" to learn from. For example, client populations can be segmented into peer groups in order to understand deviations in their activity.

USE CASES IN AML

Machine learning can be valuable in improving the efficiency and effectiveness of your anti-money laundering program. In most cases, machine learning is augmenting rather than replacing traditional transaction monitoring methods. Below are several examples of use cases where banks are leveraging these techniques today:

- Segmentation **–** Transaction monitoring scenarios can generate high false positive rates due to the use of thresholds. To mitigate operational risk, thresholds are often set high, creating alerts that are disproportionately skewed toward high dollar transactions. As a result, an activity that is worth a review might be sitting below the threshold. Leveraging a behavior-based approach, customer groups can be created using unsupervised techniques. Transaction monitoring scenarios can then be segment aware, having different thresholds that are based on each customer group. The resulting effects are lower false positives and greater coverage of activity that was hiding ‘below the line’.

- **Alert Scoring and Hibernation –** Anti-money laundering typologies are often temporal in nature, making it difficult to identify on the basis of a single transaction or alert. Money laundering involves the layering of funds to officiate its source. Many alerts are closed during the triage phase of the investigations process due to lack of evidence. By aggregating alerts together, we can continuously risk rate customer activity to determine when it becomes investigations worthy. At that time, the group of alerts is promoted to case and valuable investigative resources are focused on the riskiest behaviors.
- **Scenario Replacement –** More progressive financial institutions have replaced several transaction monitoring scenarios with one machine learning model. Typologies that have historical case outcomes are well defined and result in high false positives and are generally the best candidates. In this paper, we will explore replacing Cash focused scenarios with a machine learning model.

EXAMPLE FRAMEWORK LEVERAGING SAS® FINANCIAL CRIMES ANALYTICS

This paper explores the process of building, registering and deploying a machine learning model to replace transaction monitoring scenario(s). To demonstrate the framework, we will leverage SAS® Financial Crimes Analytics as the underlying technology. SAS® Financial Crimes Analytics provides end-to-end capabilities, from data to decisioning.

- **Data Acquisition –** Data is acquired from the SAS® Anti-Money Laundering data model and loaded into the SAS® Financial Crimes Analytics environment.
- **Feature Engineering –** Data is joined to create a customer centric view with transactional, non-transactional, demographic, and other relevant features to create a base table for modeling.
- **Automated Model Development –** Machine Learning models are created leveraging the Automated Pipeline Generation feature.
- **Model Management and Governance –** The champion machine learning model is registered in a centralized repository.
- **Alert Generation –** Output from the machine learning model is assessed with a cutoff score to determine at what threshold to generate an alert. The decision flow is tested, then deployed to SAS® Anti-Money Laundering using the alert API.
- **Investigations –** Output from the machine learning model becomes actionable for investigations.

DATA ACQUISITION

The SAS® Anti-Money Laundering solution has a robust data model that links a wide variety of transaction, non-transaction, and demographic dimensions to the entity ("Customer") or external entity ("Customer's Customer"). This data is rich in information such as the primary instrument, secondary instrument, beneficiary, channel, branch, and other potential features that can be used to predict suspicious activity. The solution also houses information related to the disposition of prior investigations. Because information related to anti-money laundering programs are not publicly available, we have populated the SAS® Anti-Money Laundering data model with fictitious data to illustrate the framework.

This diagram shows the SAS® Anti-Money Laundering conceptual data model:

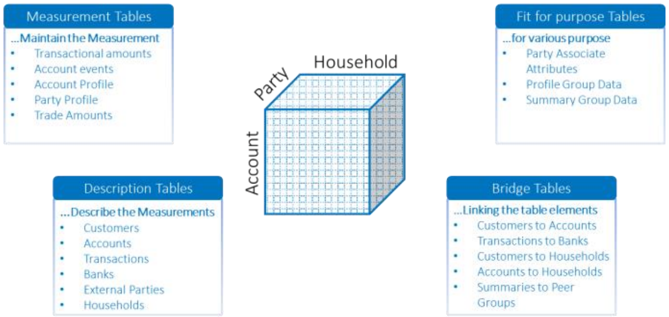
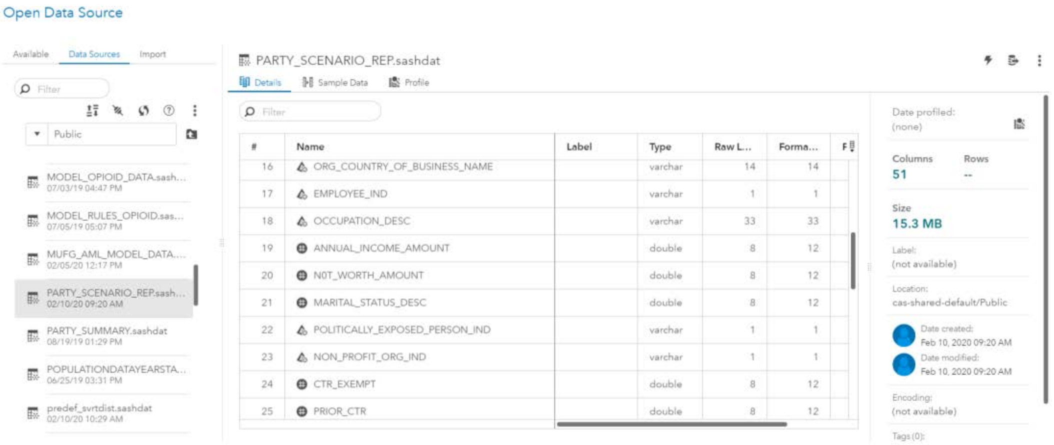


Figure 2: SAS® Anti-Money Laundering Core Dimensional Model

FEATURE ENGINEERING

After loading the data, the first step is to transform the raw data collected from the SAS® Anti-Money Laundering core dimensional model and create a base table for our model development. Transaction monitoring scenarios typically leverage 5 to 7 parameters, but machine learning techniques can consider many more features to more accurately identify patterns. We identified 50 potential features that were indicative of Cash related behaviors based on the FFIEC guidance (Federal Financial Institutions Examination Council) and domain knowledge. Next, the label (or target) was identified as "Good Alert" that represents alerts which were promoted from triage to case for cash typologies. This information was aggregated to the customer level in order to identify cross account behaviors such as Cash Structuring.

Display 1 shows the base modeling table created:


Open Data Source

MODEL_OPIOID_DATA.sash...
07/03/19 04:47 PM
MODEL_RULES_OPIOID.sas...
07/05/19 05:07 PM
MUFG_AML_MODEL_DATA....
02/05/20 12:17 PM
PARTY_SCENARIO_REP.sash...
02/10/20 09:20 AM
PARTY_SUMMARY.sashdat
08/19/19 01:29 PM
POPULATIONDATAYEARSTA...
06/25/19 03:31 PM
predef_svrtdist.sashdat
02/10/20 10:29 AM

PARTY_SCENARIO_REP.sashdat

#	Name	Label	Type	Raw L...	Forma...	F
16	ORG_COUNTRY_OF_BUSINESS_NAME		varchar	14	14	
17	EMPLOYEE_IND		varchar	1	1	
18	OCCUPATION_DESC		varchar	33	33	
19	ANNUAL_INCOME_AMOUNT		double	8	12	
20	N0T_WORTH_AMOUNT		double	8	12	
21	MARITAL_STATUS_DESC		double	8	12	
22	POLITICALLY_EXPOSED_PERSON_IND		varchar	1	1	
23	NON_PROFIT_ORG_IND		varchar	1	1	
24	CTR_EXEMPT		double	8	12	
25	PRIOR_CTR		double	8	12	

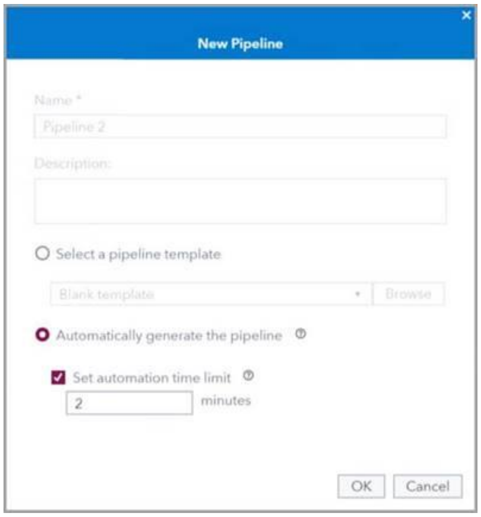
Date profiled:
(none)
Columns 51
Rows --
Size
15.3 MB
Label:
(not available)
Location:
cas-shared-default/Public
Date created:
Feb 10, 2020 09:20 AM
Date modified:
Feb 10, 2020 09:20 AM
Encoding:
(not available)
Tags (0):

Display 1: Data Profile

MODEL DEVELOPMENT

Once the base modeling table is created, the second step is to create a new project. In order to accelerate our model development process, the automated pipeline generation feature was leveraged. This process automates iterative model development process, creating a dynamically generated pipeline from the input data. Once created, we will review the champion and challenger models, expose the underlying settings, parameters, and code. Though the system is creating the pipeline for the user, the model development process is not a black box and can be edited.

Display 2 shows settings to create an automatically generated pipeline:


New Pipeline

Name *

Pipeline 2

Description:

○ Select a pipeline template

Blank template

● Automatically generate the pipeline

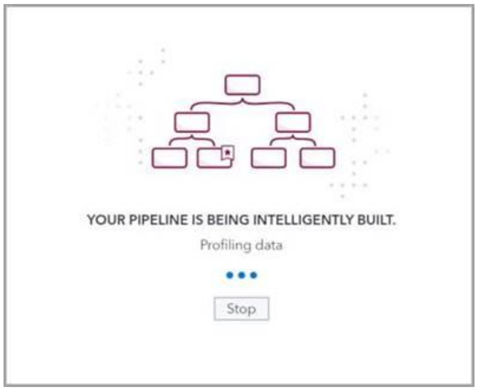
☑ Set automation time limit

2 minutes

Display 2: Automatically Generate the Pipeline

Once executed, a pipeline is created automatically based on the data that is provided. The system is intelligently determining the transformations and features needed to identify the best solution. Through an iterative process, models are assessed, hyper-parameters are tuned, and ensembles are created to predict the outcome of "good alert".

Display 3 shows the steps in real time as the pipeline is being created:

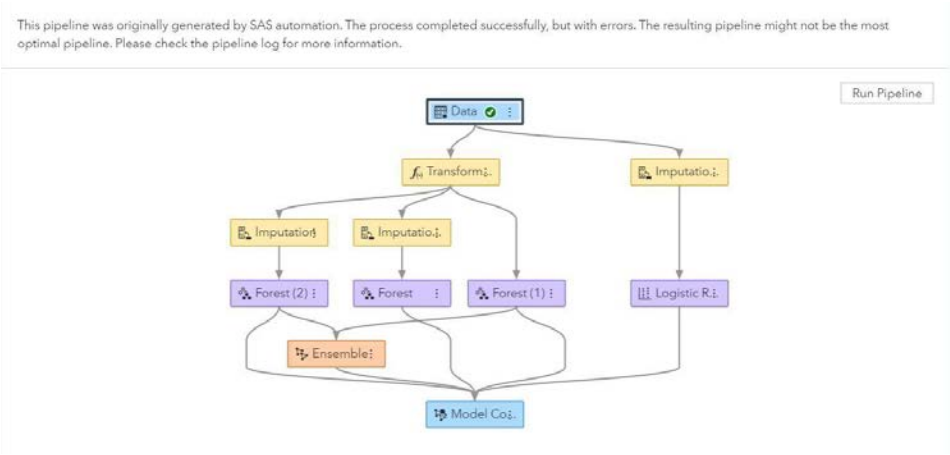

YOUR PIPELINE IS BEING INTELLIGENTLY BUILT.

Profiling data

Display 3: Pipeline is Being Intelligently Built

Once complete, the pipeline is displayed showing the nodes generated including variable imputation, feature selection, and models. Many modeling techniques will be executed in order to identify a champion model. Keep in mind that with different input data, the pipeline generated will be different.

Display 4 shows the pipeline that was automatically generated:

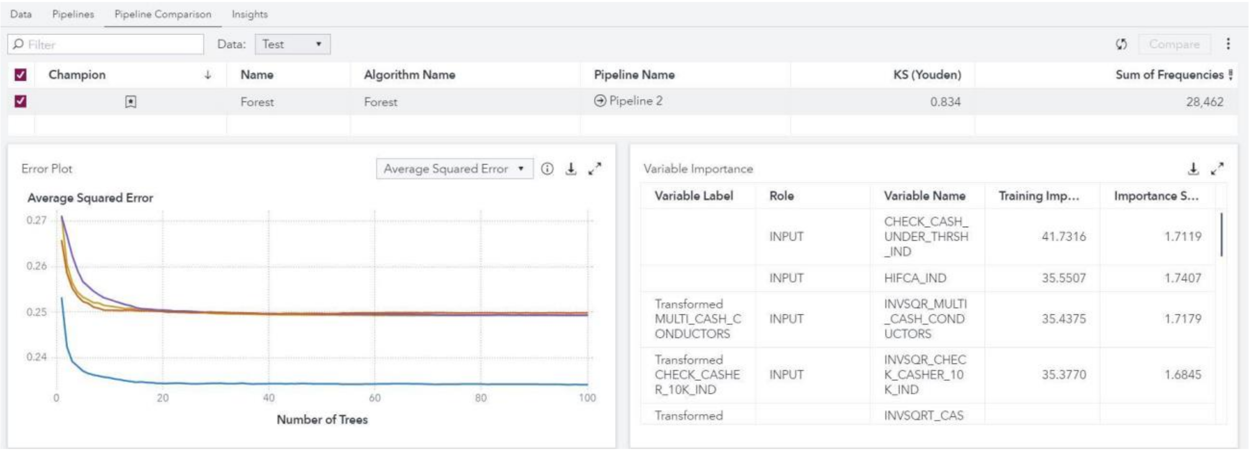


This pipeline was originally generated by SAS automation. The process completed successfully, but with errors. The resulting pipeline might not be the most optimal pipeline. Please check the pipeline log for more information.

Display 4: Automatically Generated the Pipeline

The automatically generated pipeline includes a model comparison node, which compares each technique to determine which model was best at predicting past cash activity and was promoted from triage to case. Based on misclassification rate, the Gradient Boosting model was selected as the champion.

Display 5 shows the results of our automatically generated pipeline:



☑ Champion	Name	Algorithm Name	Pipeline Name	KS (Youden)	Sum of Frequencies
☑ ★	Forest	Forest	Pipeline 2	0.834	28,462

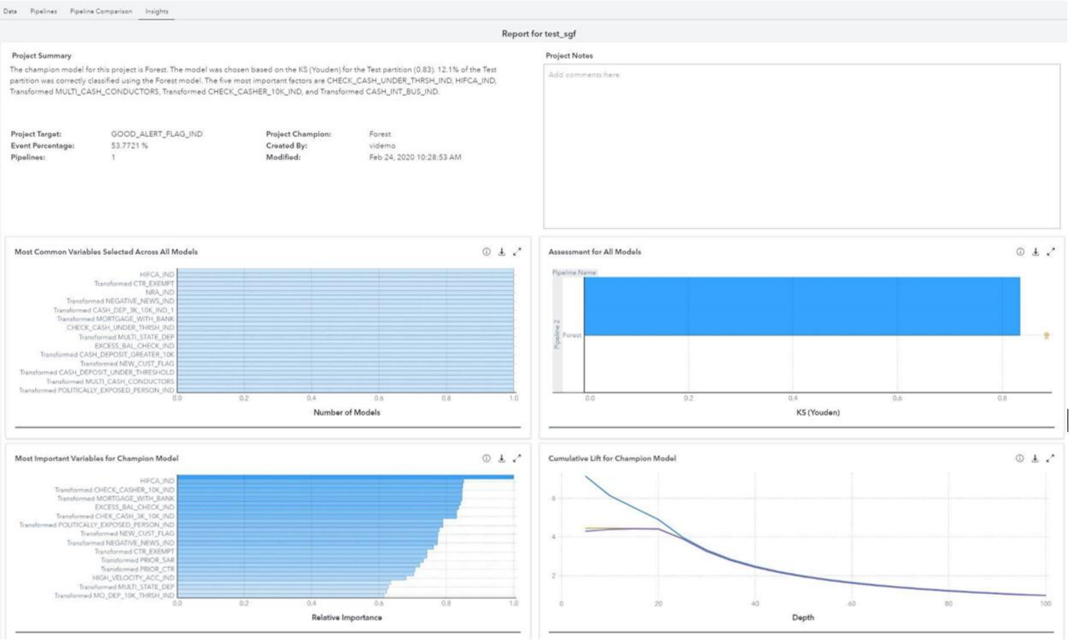
Variable Importance

Variable Label	Role	Variable Name	Training Imp...	Importance S...
	INPUT	CHECK_CASH_UNDER_THRSH_IND	41.7316	1.7119
	INPUT	HIFCA_IND	35.5507	1.7407
Transformed MULTI_CASH_CONDUCTORS	INPUT	INVSQR_MULTI_CASH_CONDUCTORS	35.4375	1.7179
Transformed CHECK_CASHER_10K_IND	INPUT	INVSQR_CHECK_CASHER_10K_IND	35.3770	1.6845
Transformed		INVSQRT_CAS		

Display 5: Model Comparison Results

Model documentation is automatically generated in a project summary report called an insight. Insights provide details for the project such as project summary created using natural language generation, model assessment, cumulative lift for champion model, most important variables for champion model, and many other key visualizations.

Display 6 shows a sample of the model assessment results:



Report for test_sgf

Project Summary

The champion model for this project is Forest. The model was chosen based on the KS (Youden) for the Test partition (0.83). 12.1% of the Test partition was correctly classified using the Forest model. The five most important factors are CHECK_CASH_UNDER_THRSH_IND, HIFCA_IND, Transformed MULTI_CASH_CONDUCTORS, Transformed CHECK_CASHER_10K_IND, and Transformed CASH_INT_BUS_IND.

Project Target:	GOOD_ALERT_FLAG_IND	**Project Champion:**	Forest
Event Percentage:	53.7721 %	**Created By:**	videmo
Pipelines:	1	**Modified:**	Feb 24, 2020 10:28:53 AM

Project Notes

Add comments here.

Most Common Variables Selected Across All Models
HIFCA_IND
Transformed CTR_EXEMPT
NRA_IND
Transformed NEGATIVE_NEWS_IND
Transformed CASH_DEP_3K_10K_IND_1
Transformed MORTGAGE_WITH_BANK
CHECK_CASH_UNDER_THRSH_IND
Transformed MULTI_STATE_DEP
EXCESS_BAL_CHECK_IND
Transformed CASH_DEPOSIT_GREATER_10K
Transformed NEW_CUST_FLAG
Transformed CASH_DEPOSIT_UNDER_THRESHOLD
Transformed MULTI_CASH_CONDUCTORS
Transformed POLITICALLY_EXPOSED_PERSON_IND
0.0
0.2
0.4
0.6
0.8
1.0
Number of Models
Assessment for All Models
Pipeline Name
Pipeline 2
Forest
0.0
0.2
0.4
0.6
0.8
KS (Youden)
Most Important Variables for Champion Model
HIFCA_IND
Transformed CHECK_CASHER_10K_IND
Transformed MORTGAGE_WITH_BANK
EXCESS_BAL_CHECK_IND
Transformed CHEK_CASH_3K_10K_IND
Transformed POLITICALLY_EXPOSED_PERSON_IND
Transformed NEW_CUST_FLAG
Transformed NEGATIVE_NEWS_IND
Transformed CTR_EXEMPT
Transformed PRIOR_SAR
Transformed PRIOR_CTR
HIGH_VELOCITY_ACC_IND
Transformed MULTI_STATE_DEP
Transformed MO_DEP_10K_THRSH_IND
0.0
0.2
0.4
0.6
0.8
1.0
Relative Importance
Cumulative Lift for Champion Model
6
4
2
0
20
40
60
80
100
Depth

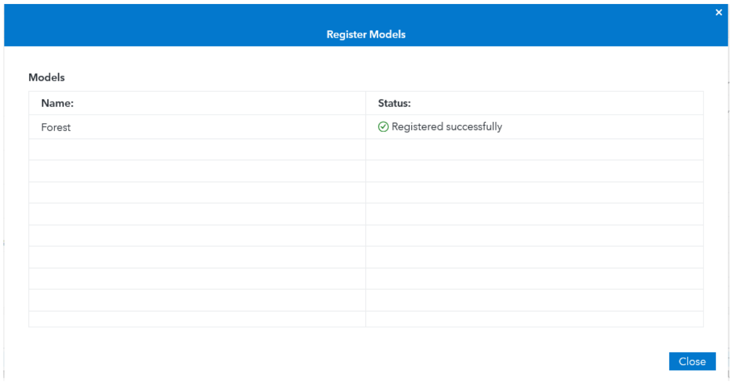
Display 6: Model Assessment Results

MODEL MANAGEMENT AND GOVERNANCE

Once you have identified your champion model, register the model. Registration of the model provides four benefits:

1. Organization and Management of Models: A centralized repository allows users to capture data and model lineage in one place, compare open source and SAS models, and create life cycle templates to expedite deployment.
2. Test and Validate Models: Evaluate whether the models were performant against investigative results to determine if they are capturing the money laundering behaviors intended in the definition.
3. Publish: Automatically deploy in batch, streaming, or cloud. For our example, we will leverage the integration with SAS Decision Manager to ultimately create an alert work item.
4. Monitor Performance over Time: Performance monitoring and alerting automate the model updating process to address model degradation.

Display 7 shows registration of the new model:



Register Models

Models

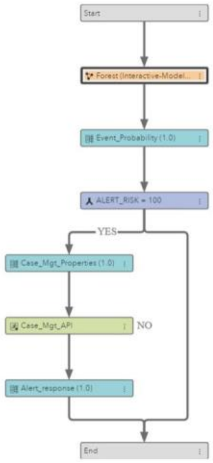
Name:	Status:
Forest	✓ Registered successfully

Display 7: Registering the Model

ALERT GENERATION

To derive value from our registered machine learning model, it is required to determine when to create an alert for investigations to review. A decision flow can be designed leveraging the drag and drop decision authoring interface. This flow automatically assesses the output of the machine learning model and leverage rules to determine when to generate an alert. Decisions can range from simple to complex. To illustrate the framework, we will create a simple flow which creates an alert if the event probability from the model has exceeded a cut off score.

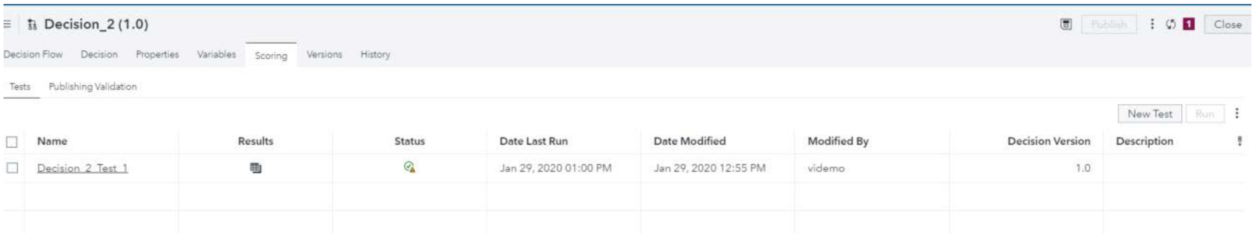
Display 8 shows the decision flow for creating an alert:



Display 8: Decision Flow

Based on the preceding flow, if the customer receives an event probability score of above .05, we structure the variables required for the SAS® Anti-Money Laundering Solution, send the alert via API and document each record with a message that indicates if an alert was generated or not. Next, you test the decision flow to ensure it is working as intended. The results can be reviewed to determine the operational impact, rules fired, and the alerts that would be generated once we publish.

Display 9 shows testing of the decision flow:

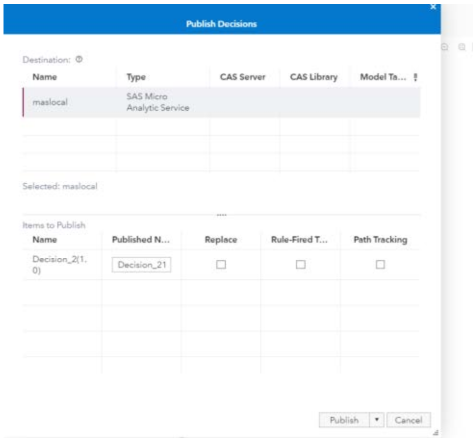


□	Name	Results	Status	Date Last Run	Date Modified	Modified By	Decision Version	Description
□	Decision_2_Test_1			Jan 29, 2020 01:00 PM	Jan 29, 2020 12:55 PM	videmo	1.0	

Display 9: Testing Decision Flow

Once the decision flow is tested and validated, publish the flow. There is an option of manually publishing on demand or scheduling the flow to run automatically.

Display 10 shows publishing the flow:



Destination:

Name	Type	CAS Server	CAS Library	Model Ta...
maslocal	SAS Micro Analytic Service			

Selected: maslocal

Items to Publish

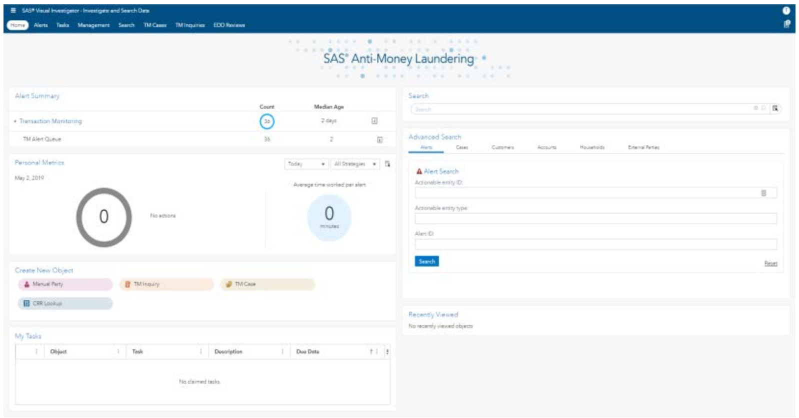
Name	Published N...	Replace	Rule-Fired T...	Path Tracking
Decision_2(1.0)	Decision_21	□	□	□

Display 10: Publishing the Flow

INVESTIGATIONS

After the alerts have been published, they are available in SAS® Anti-Money Laundering for investigators to act. The alerts can now be searched, triaged, and investigated to determine if a regulatory report is required. The alerts generated by the newly created cash machine learning model are grouped with other alerting channels, such as transaction monitoring and human rereferrals, so that analyst can have a compressive review of the entity. Lastly, you can surface descriptive and understandable reasons as to why the alerts were created to assist investigations in reaching a decision.

Display 11 shows alerts generated in SAS® Anti-Money Laundering:



	Count	Median Age
Transaction Monitoring	36	2 days
TM Alert Queue	36	2

Display 11: SAS Anti-Money Laundering Homepage

CONCLUSION

There is a lot of excitement in the Financial Crime and Compliance industry around the application of machine learning techniques. We see many opportunities available today to apply these methods to improve the efficiency and effectiveness of AML and KYC programs. Replacement of low performing high volume transaction monitoring scenarios can be an area where machine learning techniques can provide significant value, freeing up resources to focus on investigation worthy activity. With SAS® Financial Crimes Analytics, banks can create, manage, and publish machine learning models rapidly within an integrated framework.

REFERENCES

Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Financial Crimes Enforcement Network, National Credit Union Administration, and Office of the Comptroller of the Currency. 2018. "Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing" Available at https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20181203a1.pdf

Federal Reserve Bank of St. Louis. 2012. "Economic Research." Accessed November 7, 2019. http://research.stlouisfed.org.

Ray, Arin. and Katkov, Neil. 2019. "IT and Operational Spending in AML-KYC A Global Perspective." *CELENT*.

ACKNOWLEDGMENTS

The authors would like to thank Charlotte Crain and Mason Wheeless for their technical guidance and thought leadership.

CONTACT INFORMATION

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

Beth Herron
SAS Institute, Inc.
Beth.Herron@sas.com
Fraud and Security Intelligence Division

Saurabh Duggal
SAS Institute, Inc.
Saurabh.Duggal@sas.com
Fraud and Security Intelligence Division