UNIVERSIDAD ANÁHUAC · VINCE IN BONO MALUM · MÉXICO

IDAA
INSTITUTO DE ANALÍTICA AVANZADA

A Statistical and Machine Learning Model to Detect Money Laundering: an Application

Dr. Miguel Agustín Villalobos
Director – Advanced Analytics Institute

Dr. Eliud Silva
Professor

Actuarial Sciences Department
Anahuac University

June 15th, 2017

Agenda

- Executive summary
- Background and key references
- Financial Institutions Needs
- The model
- Conclusions

The objective is to apply statistical learning methods to address the AML needs of medium size banks to comply to recent regulation

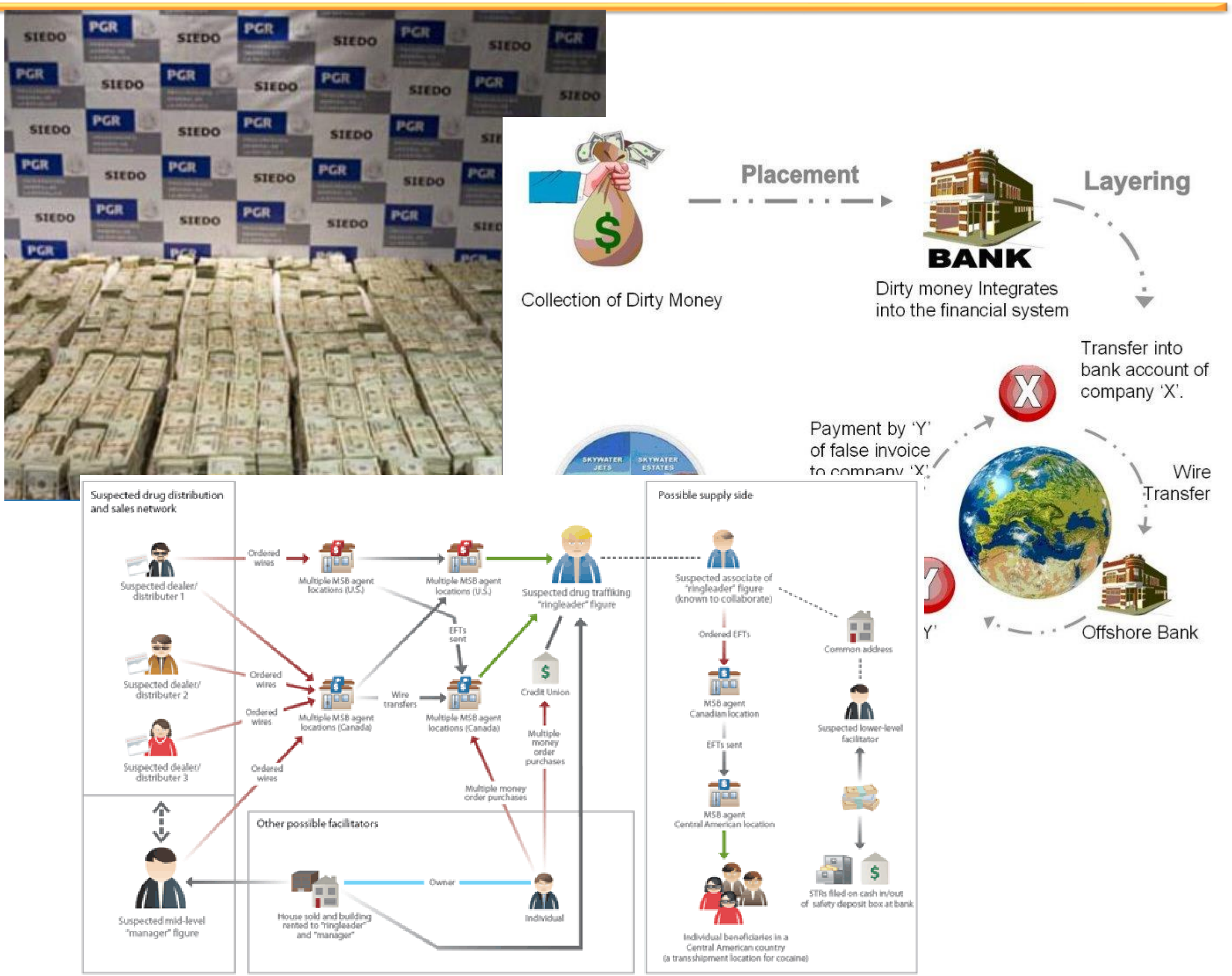
- Recent regulation by Central Banks is requiring financial institutions to implement a data & statistical models based approach to detect money laundering
- Medium and small sized banks seek to comply with regulation without or with minimal changes to their AML systems which are mostly deterministic rule-based
- We present a three module solution that responds to the medium size banks in the region to have an easy to implement solution to:
  - Complement their current AML systems
  - Minimize changes to their current AML and transactional systems
  - Provide a “best classification models combination” to detect suspect transactions
- In the particular case analyzed here, it is concluded that the developed model, provides a 99.6% correct classification rate on test data, leading to a reduction in the number of alerted cases from the current close to 30% of transactions to less than 1%

Money laundering poses significant challenges to financial institutions

- For example, economic impact in Mexico is estimated to be between \$10 BUSD[1] and 25 BUSD[2])

- High volumes and complexity

- Dynamic and fast evolving


SIEDO
PGR
Collection of Dirty Money
Placement
BANK
Dirty money Integrates into the financial system
Layering
Transfer into bank account of company 'X'.
X
Payment by 'Y' of false invoice to company 'X'.
Wire Transfer
Offshore Bank
Y'
SKYWATER JETS
SKYWATER ESTATES
Suspected drug distribution and sales network
Suspected dealer/ distributer 1
Suspected dealer/ distributer 2
Suspected dealer/ distributer 3
Ordered wires
Ordered wires
Ordered wires
Ordered wires
Multiple MSB agent locations (U.S.)
Multiple MSB agent locations (U.S.)
EFTs sent
Multiple MSB agent locations (Canada)
Wire transfers
Multiple MSB agent locations (Canada)
Suspected drug traffiking "ringleader" figure
Credit Union
Multiple money order purchases
Multiple money order purchases
Suspected mid-level "manager" figure
Other possible facilitators
House sold and building rented to "ringleader" and "manager"
Owner
Individual
Possible supply side
Suspected associate of "ringleader" figure (known to collaborate)
Ordered EFTs
MSB agent Canadian location
EFTs sent
MSB agent Central American location
Individual beneficiaries in a Central American country (a transshipment location for cocaine)
Common address
Suspected lower-level facilitator
STRs filed on cash in/out of safety deposit box at bank

1. Angel, A. (11 de 11 de 2016). InSight Crime. Recovered from http://www.insightcrime.org/news-analysis/mexico-fight-against-money-laundering-comes-down-to-resources
2. Tamaño de lavado de dinero en México: Recovered from http://www.elfinanciero.com.mx/nacional/lavado-de-dinero-en-mexico-supera-los-25mil-mdd.html

Significant research has been done in the area of counter financial crimes based on data mining and statistical learning methods

- There has been a significant amount of research in the use of statistical learning and data mining to fight financial crimes. At the end of this presentation we include the most relevant recent references used to develop the model presented here

- The key objective is to develop a model simple enough for ease of implementation and sophisticated enough to protect the financial institution and comply with regulation

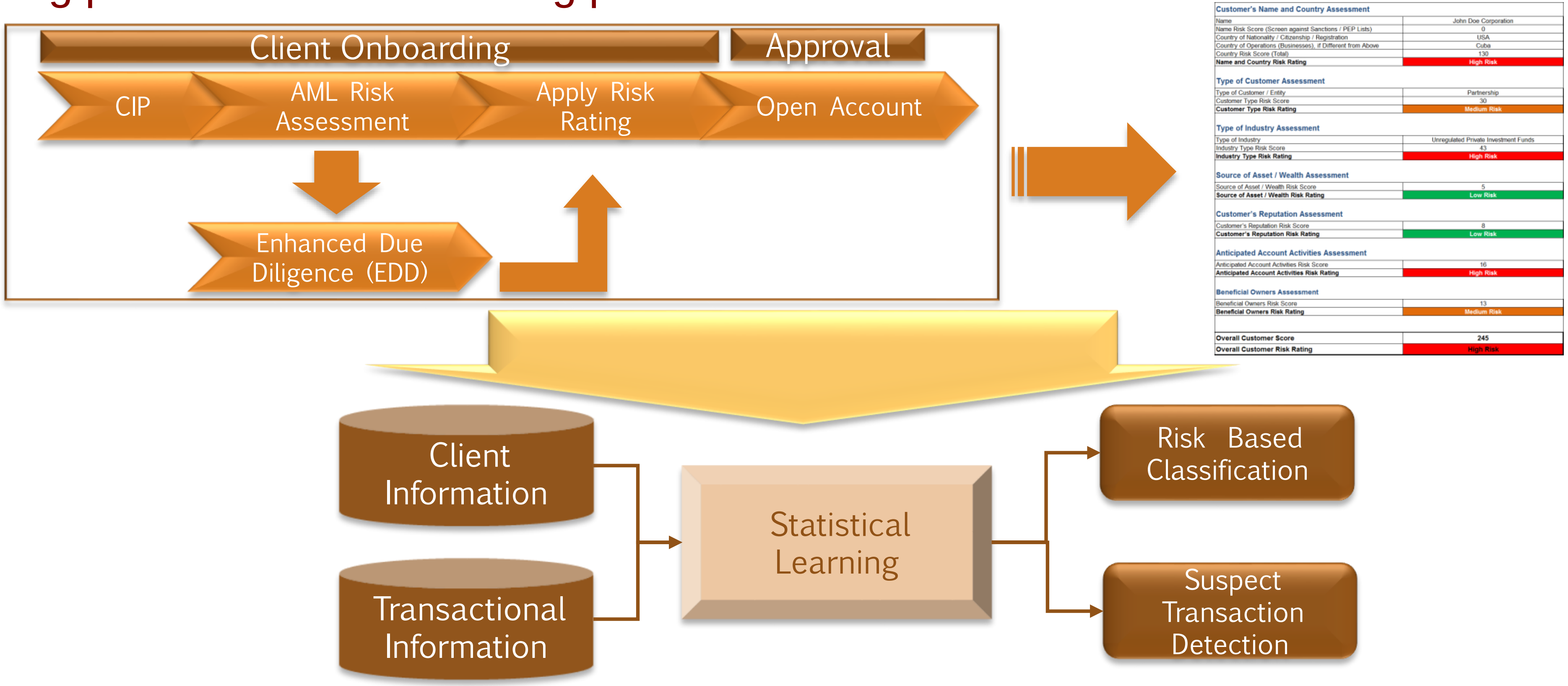
- Two main references to this applied work are:
  - Kharote & Kshirsagar (2014)
    - Clustering
    - Profile generation
    - Customer behavior extraction
  - Claudio & Joao (2016):
    - Clustering based client profiling
    - Classification rule generation

Due to the extended use of electronic transactions in different currencies, Central Banks in LA are starting to require implementation of more advanced risk models

- Current models, based on deterministic business rules need to be replaced by statistical learning models
- Risk Model Criteria
  - Inherent risk factors
  - Transactional Profile
- Data based client risk classification under well structured criteria
- Periodic validation
- Theoretical and empirical basis documentation

Most financial institutions already have an AML system in place, but need to modify it to comply with current regulations

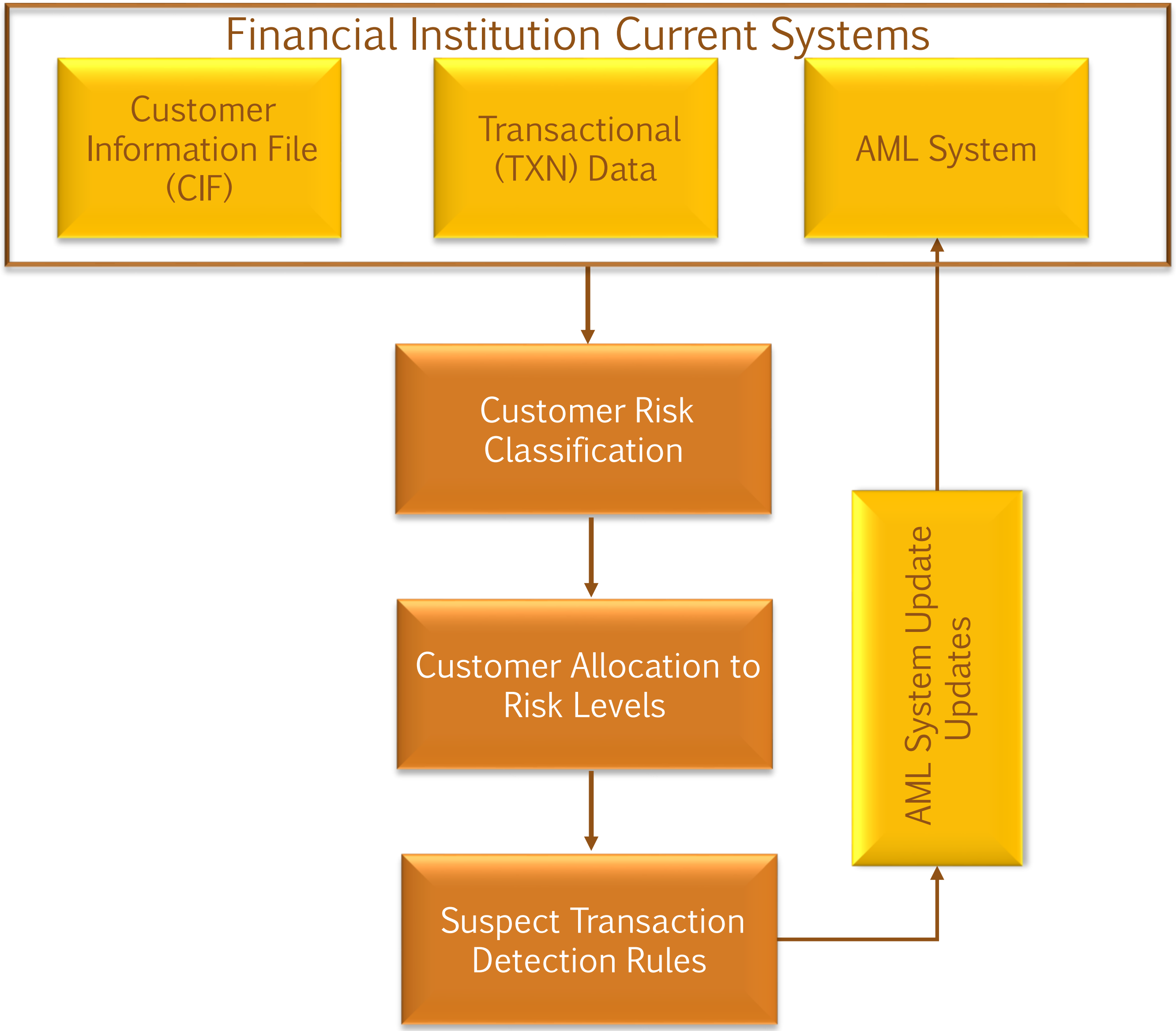
- For speedy implementation, a solution is needed to ***complement*** current AML business rule based systems
- Due to security issues faced by banks IT environment, a model with minimal impact in current systems and processes must be implemented
- First requirement imposed by new regulations is for banks to challenge their current customer risk rating process with a risk rating process based on data



Customer's Name and Country Assessment	
Name	John Doe Corporation
Name Risk Score (Screen against Sanctions / PEP Lists)	0
Country of Nationality / Citizenship / Registration	USA
Country of Operations (Businesses), if Different from Above	Cuba
Country Risk Score (Total)	130
Name and Country Risk Rating	High Risk
Type of Customer Assessment	
Type of Customer / Entity	Partnership
Customer Type Risk Score	30
Customer Type Risk Rating	Medium Risk
Type of Industry Assessment	
Type of Industry	Unregulated Private Investment Funds
Industry Type Risk Score	43
Industry Type Risk Rating	High Risk
Source of Asset / Wealth Assessment	
Source of Asset / Wealth Risk Score	5
Source of Asset / Wealth Risk Rating	Low Risk
Customer's Reputation Assessment	
Customer's Reputation Risk Score	8
Customer's Reputation Risk Rating	Low Risk
Anticipated Account Activities Assessment	
Anticipated Account Activities Risk Score	16
Anticipated Account Activities Risk Rating	High Risk
Beneficial Owners Assessment	
Beneficial Owners Risk Score	13
Beneficial Owners Risk Rating	Medium Risk
Overall Customer Score	**245**
Overall Customer Risk Rating	High Risk

The model is structured in three modules defined to achieve simplicity and minimal financial institution systems

- Based on the premises simplicity and minimal bank systems intrusion our model is structured in three modules and assumes no on-line interphases with current systems
- The updates to the institution´s AML system are assumed to be made by their personnel, according to their standards and procedures
- The process will be illustrated using transformed data from a Latin American financial institution
- Two types of events are considered:
  - Alerts: transactions flagged for investigation
  - Suspicious: Reported to regulator after investigation



Financial Institution Current Systems

Customer Information File (CIF)

Transactional (TXN) Data

AML System

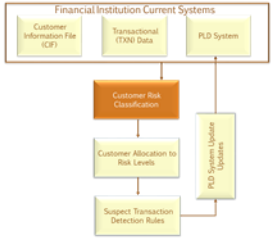
Customer Risk Classification

Customer Allocation to Risk Levels

Suspect Transaction Detection Rules

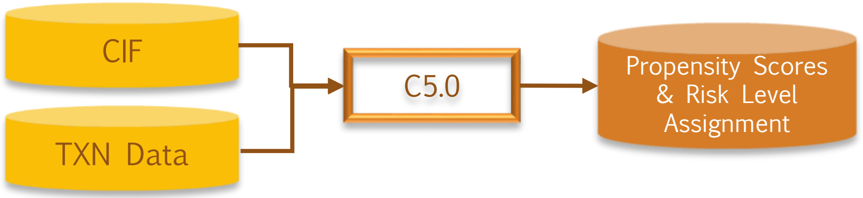
AML System Update Updates

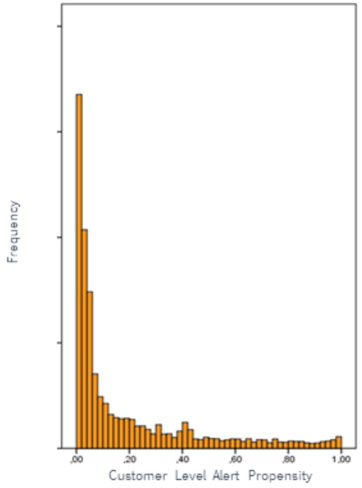
The Customer Risk Classification seeks to assess customer risk based on historical data to estimate alert propensity and define customer risk levels

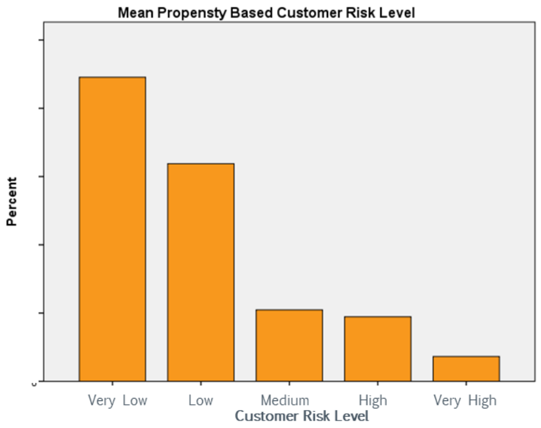


Financial Institution Current Systems
Customer Information File (CIF)
Transactional (TXN) Data
PLD System
Customer Risk Classification
Customer Allocation to Risk Levels
Suspect Transaction Detection Rules
PLD System Update Updates

- Based on historical transactional information containing an indicator of whether a transaction was flagged for investigation or not, Alert Propensity Scores are developed
- Several algorithms were used being C5.0[3] the one that provided the best correct classification (98.41% with the testing set)





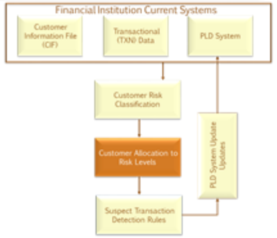


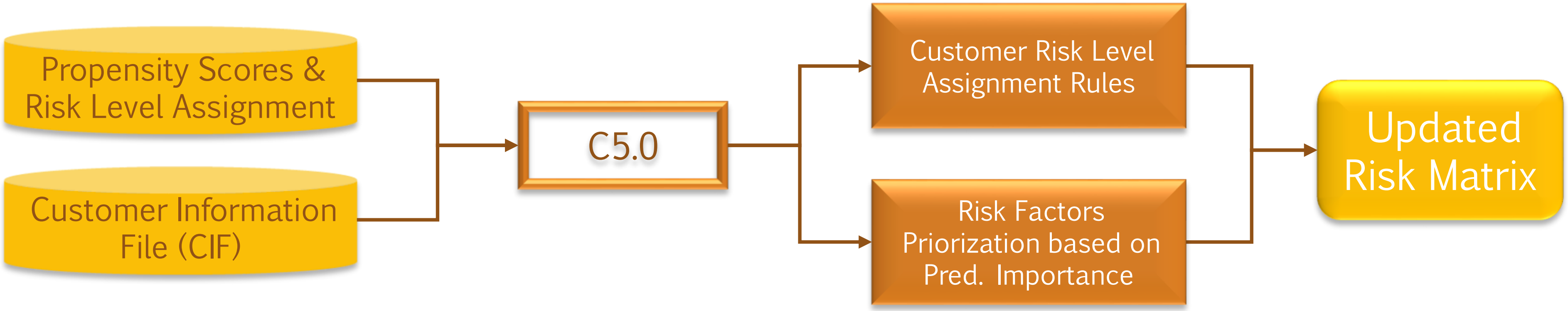
- Propensity Scores are then used to define different levels of risk (boundaries and number of risk levels depend on the organization's risk appetite (in this application 5 levels are used)

[3] Quinlan, R. (2004). C5.0, www.rulequest.com

Customer allocation to risk levels is done utilizing customer "static" risk factors as predictors of risk level class and updated risk classification rules are developed

- A C5.0 algorithm was elected to predict Risk Level based on the different customer "static" potential risk factors, to determine relative importance of each and create the set of rules for cluster allocation (68% Correct Classification)


Financial Institution Current Systems
Customer Information File (CIF)
Transactional (TXN) Data
PLD System
Customer Risk Classification
Customer Allocation to Risk Levels
Suspect Transaction Detection Rules
PLD System Update Updates


Propensity Scores & Risk Level Assignment

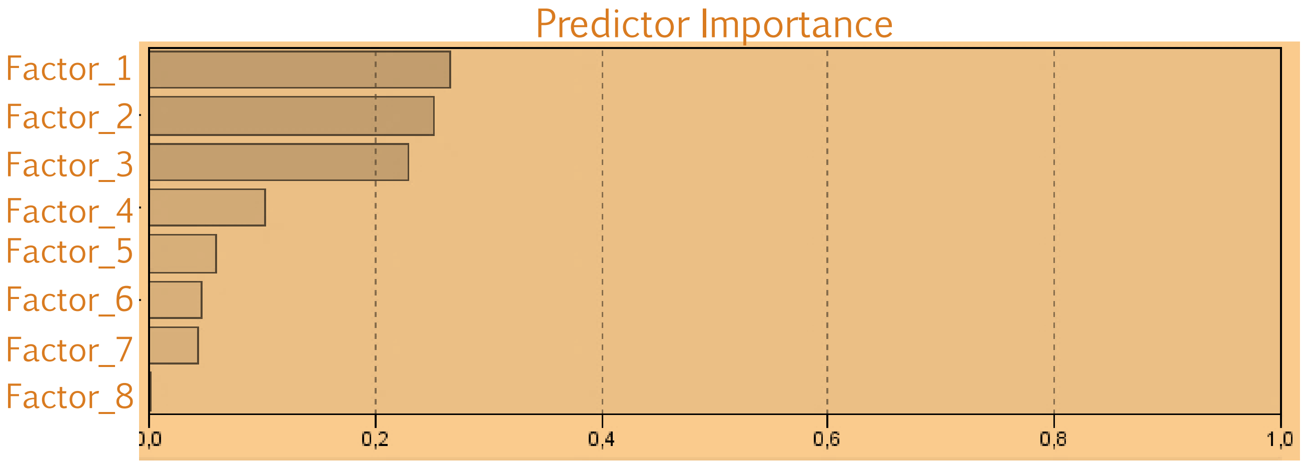
Customer Information File (CIF)

C5.0

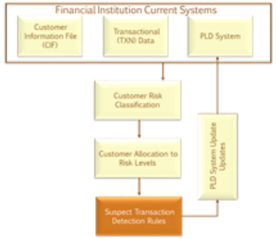
Customer Risk Level Assignment Rules

Risk Factors Priorization based on Pred. Importance

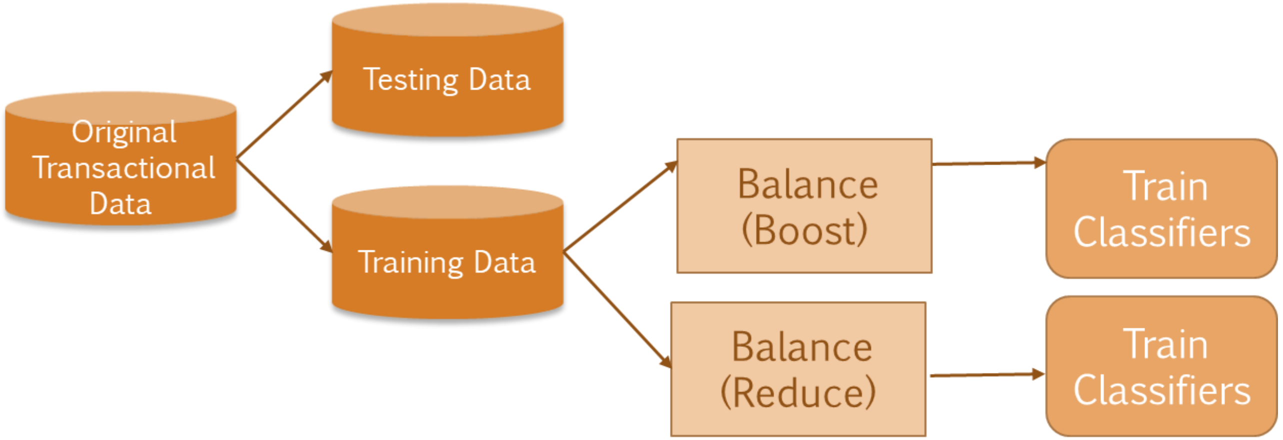
Updated Risk Matrix



Developing the classification model for suspicious transactions initiates by defining key transaction profile risk factors, balancing the data and splitting into training and testing data sets

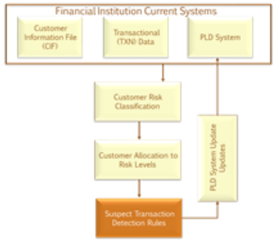

Financial Institution Current Systems
Customer Information File (CIF)
Transactional (TXN) Data
PLD System
Customer Risk Classification
Customer Allocation to Risk Levels
Suspect Transaction Detection Rules
PLD System Update Updates

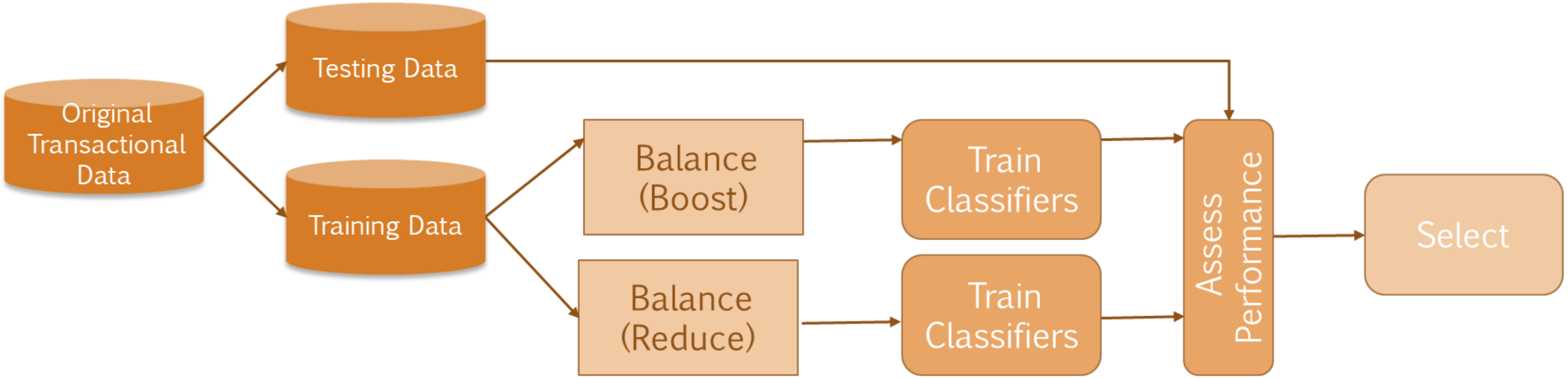
- After descriptive analysis of customer transaction data over a year of history, the following key factors were computed to characterize transactional profile both at customer and customer/month of the year levels, both for transaction amounts and number of transactions:
  - Mean, Median
  - Standard deviation, Q1, Q3, IQR, Minimum and Maximum
  - Q1 - 1.5IQR, Q1 - 3IQR, Q3 + 1.5IQR, Q3 + 3IQR,
- Data is split into a training and testing sets (50%/50%)
- Since suspicious transactions represent less than .4% of the training data, it is needed to balance the data



After adjusting for data unbalance by two different methods, several classification models were used and compared in terms of correct classification, AUC and GINI

- Five different classifiers were used with both the Boosted Balanced Data and the Reduced Balanced data, their performance was compared and the best three were selected utilizing the Boosting Balancing method to obtain a final model best combination


Financial Institution Current Systems
Customer Information File (CIF)
Transactional (TXN) Data
PLD System
Customer Risk Classification
Customer Allocation to Risk Levels
Suspect Transaction Detection Rules
PLD System Update Updates



Model	BOOST		REDUCE x 2	
	% Correct Classification	AUC	% Correct Classification	AUC
CHAID	87.20%	0.981	90.27%	0.982
C&RT	90.03%	0.956	87.59%	0.900
C5.0	**99.72%**	**0.991**	**96.74%**	**0.991**
Neural Network	88.70%	0.992	20.93%	0.547
SVM	89.80%	0.666	22.33%	0.552

The best three models were then combined under different criteria to arrive at a “best model”

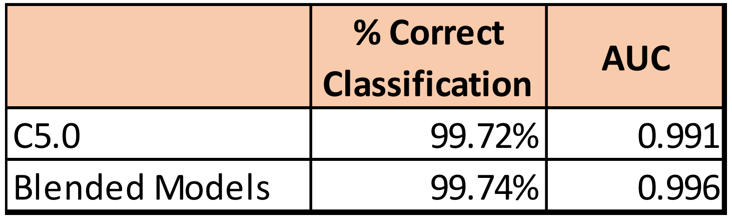
- The best three models were combined under the following criteria, to arrive at a “best model”

Model (CHAID + C&RT + C5.0)	% Correct Classification	AUC	GINI
Highest Confidence	99.74%	0.996	0.990
Confidence Weighted Voting	95.84%	0.998	0.995
Average Propensity	97.80%	0.986	0.974
Raw Propensity	96.11%	0.987	0.977
Adjusted Propensity	94.26%	0.987	0.976
Voting	96.15%	0.953	0.973

- It can be seen that the “Highest Confidence” criteria yields the best results

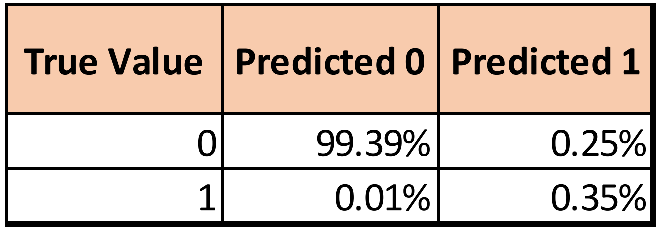
Given the results we may decide to use the C5.0 algorithm alone since the rules generated can be easily incorporated into the Institution's AML system

- We can see that the differences between the blended model and de C5.0 model are minimal



	% Correct Classification	AUC
C5.0	99.72%	0.991
Blended Models	99.74%	0.996

- Given the difference, due to ease of implementation in the financial institution, it may be decided to use the C5.0 model alone
- Currently, the number of transactions that are alerted amount to close to 30% of total number of transactions, whereas with the approach presented here, using the blended model, flagged transactions represent less than 1%



True Value	Predicted 0	Predicted 1
0	99.39%	0.25%
1	0.01%	0.35%

Bibliography

- Abdelhamid, D., Khaoula, S., & Atika, O. (2014). Automatic bank fraud detection using support vector machines. Paper presented at the 10-17. Retrieved from https://search.proquest.com/docview/1702793030?accountid=26252
- Bolton, R., & Hand, D. (2002). Statistical Fraud Detection: A Review. *Statistical Science, 17*(3), 235-249. Retrieved from http://www.jstor.org/stable/3182781
- Clunan, A. (2006). The Fight against Terrorist Financing. *Political Science Quarterly, 121*(4), 569-596. Retrieved from http://www.jstor.org/stable/20202763
- Deng, X., Joseph, V., Sudjianto, A., & Wu, C. (2009). Active Learning Through Sequential Design, With Applications to Detection of Money Laundering. *Journal of the American Statistical Association, 104*(487), 969-981. Retrieved from http://www.jstor.org/stable/40592268
- Helmy, T. H., Zaki, M., Salah, T., & Badran, K. (2016). DESIGN OF A MONITOR FOR DETECTING MONEY LAUNDERING AND TERRORIST FINANCING. Journal of Theoretical and Applied Information Technology, 85(3), 425-436. Retrieved from https://search.proquest.com/docview/1823379128?accountid=26252
- Heidarinia, N., Harounabadi, A., & Sadeghzadeh, M. (2014). An intelligent anti-money laundering method for detecting risky users in the banking systems. International Journal of Computer Applications, 97(22) doi:http://dx.doi.org/10.5120/17141-7780
- Moustafa, T. H., Abd El-Megied, M. Z., Sobh, T. S., & Shafea, K. M. (2015). Anti money laundering
- using a two-phase system. Journal of Money Laundering Control, 18(3), 304-329. Retrieved from
- https://search.proquest.com/docview/1690364822?accountid=26252
- Sudjianto, A., Yuan, M., Kern, D., Nair, S., Zhang, A., & Cela-Díaz, F. (2010). Statistical Methods for Fighting Financial Crimes. *Technometrics, 52*(1), 5-19. Retrieved from http://www.jstor.org/stable/40586676
- Verhage, A. (2009). Between the hammer and the anvil? The anti-money laundering-complex and its interactions with the compliance industry. Crime, Law & Social Change, 52(1), 9-32. doi:10.1007/s10611-008-9174-9
- Viquaruddin, M. (2010). Global Combat against Terrorism and Money Laundering: A Historical Perspective with Assessment and Strategy. Alternatives: Turkish Journal Of International Relations, 9(3), 73-80.
- Wood, K. P. (2014). Anti-money laundering in banking an enterprise-wide risk approach (Order No. 1565875). Available from ABI/INFORM Collection. (1618231900). Retrieved from https://search.proquest.com/docview/1618231900?accountid=26252

Thank you!