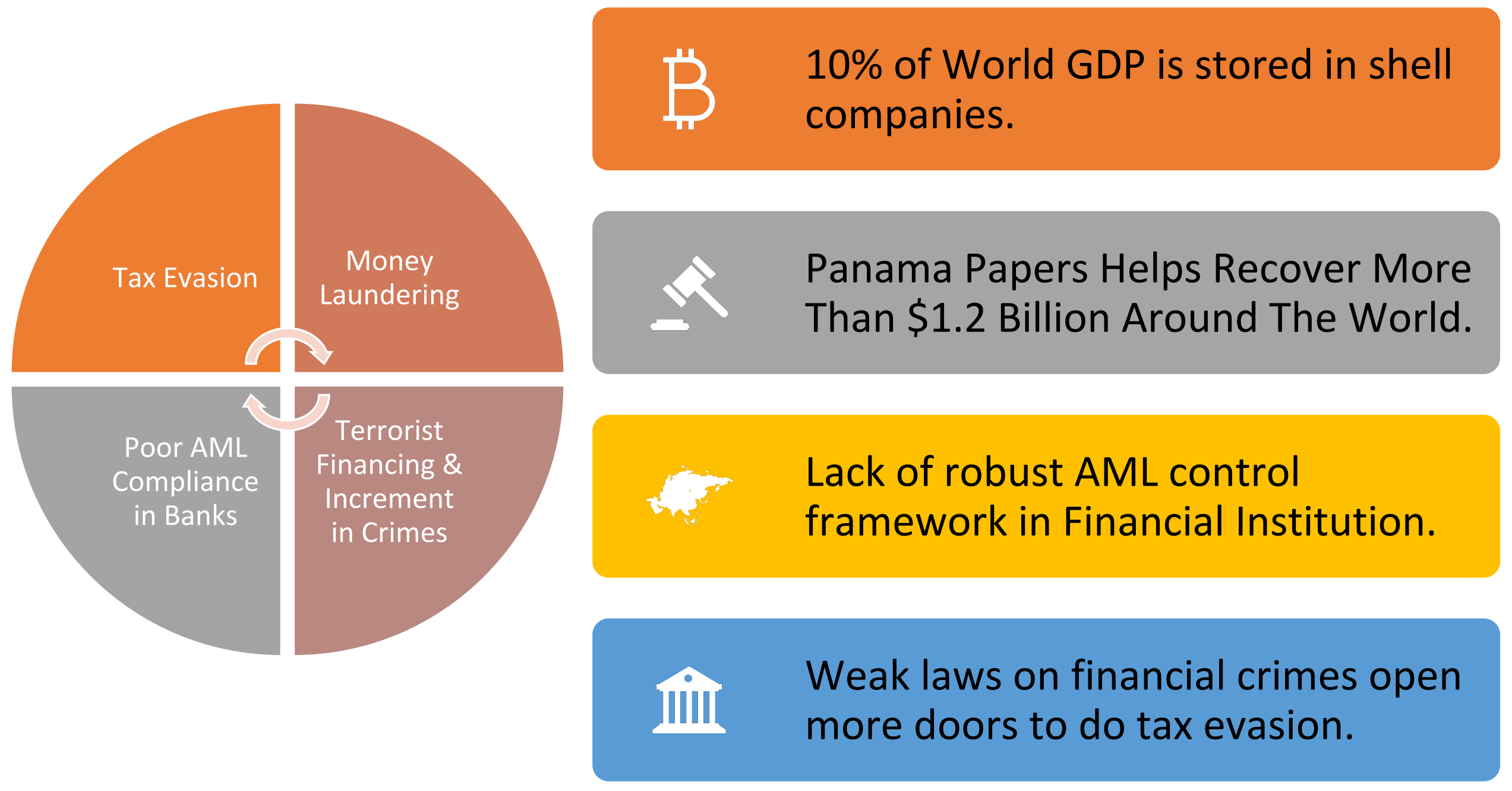
Identifying Suspicious Activities in Financial Data using Machine Learning

Authors: Pooja Pant, Kolby Viera, Monika Bagyal

UNC CHARLOTTE

Introduction

- Tax havens cost US around $70 billion a year, a fifth of its total corporate tax revenue.
- Billions of Dollars Stolen from an economy which impacts public and government programs. Concealing Income or Wealth from tax authorities using shell companies.These shell companies are further used by Criminals and Terrorists to access money worldwide.
- A staggering $26 billion in fines has been imposed on Financial Institutions for non-compliance with anti-money laundering (AML), know your customer (KYC) and sanctions regulations in the last decade.


Tax Evasion
Money Laundering
Poor AML Compliance in Banks
Terrorist Financing & Increment in Crimes

10% of World GDP is stored in shell companies.

Panama Papers Helps Recover More Than $1.2 Billion Around The World.

Lack of robust AML control framework in Financial Institution.

Weak laws on financial crimes open more doors to do tax evasion.

Fig. Problem Statement

Transaction Monitoring Model

Traditional ETL Transaction Monitoring

- Existing ETL model is rule based system which generates a lot of false positives.
- ETL system do not learn on its own from existing data pattern.

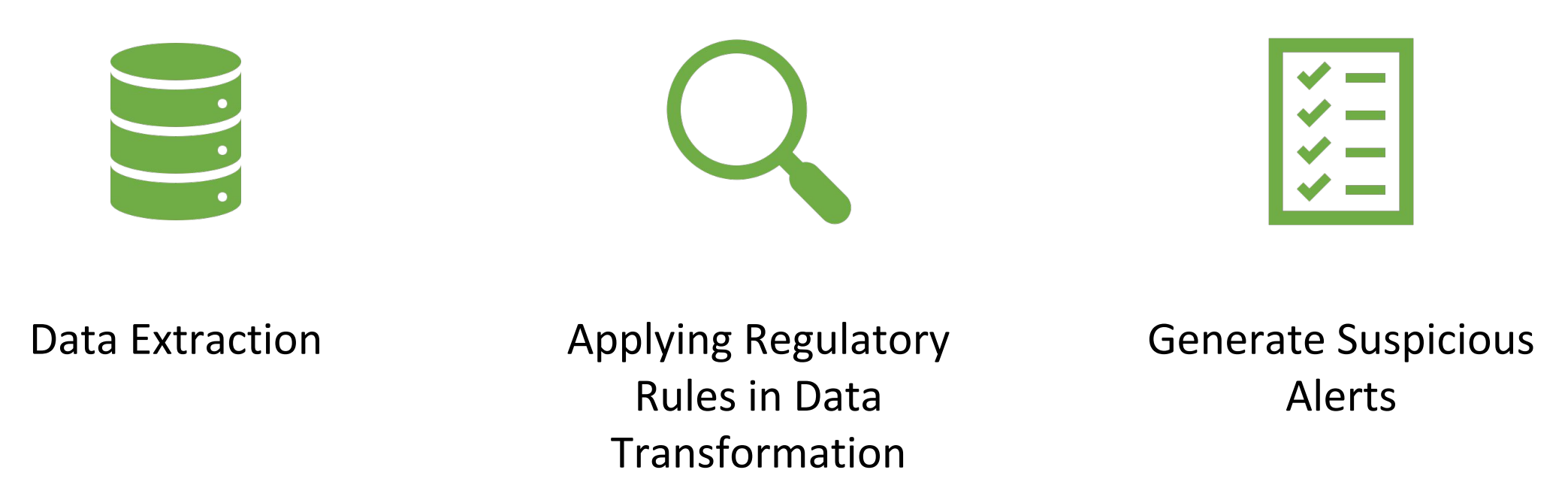
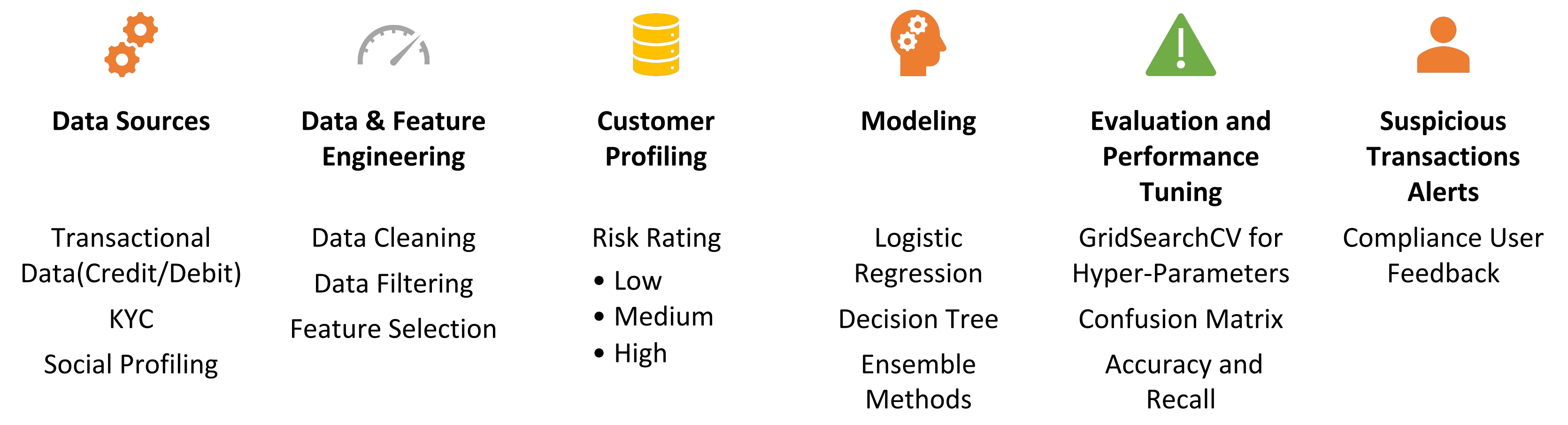

Data Extraction
Applying Regulatory Rules in Data Transformation
Generate Suspicious Alerts

Fig. ETL design

Machine Learning Model

- Machine Learning Algorithms perform better over time as it holds more and more data which drastically reduce False Positives.



Data Sources: Transactional Data(Credit/Debit); KYC; Social Profiling

Data & Feature Engineering: Data Cleaning; Data Filtering; Feature Selection

Customer Profiling: Risk Rating
- Low
- Medium
- High

Modeling: Logistic Regression; Decision Tree; Ensemble Methods

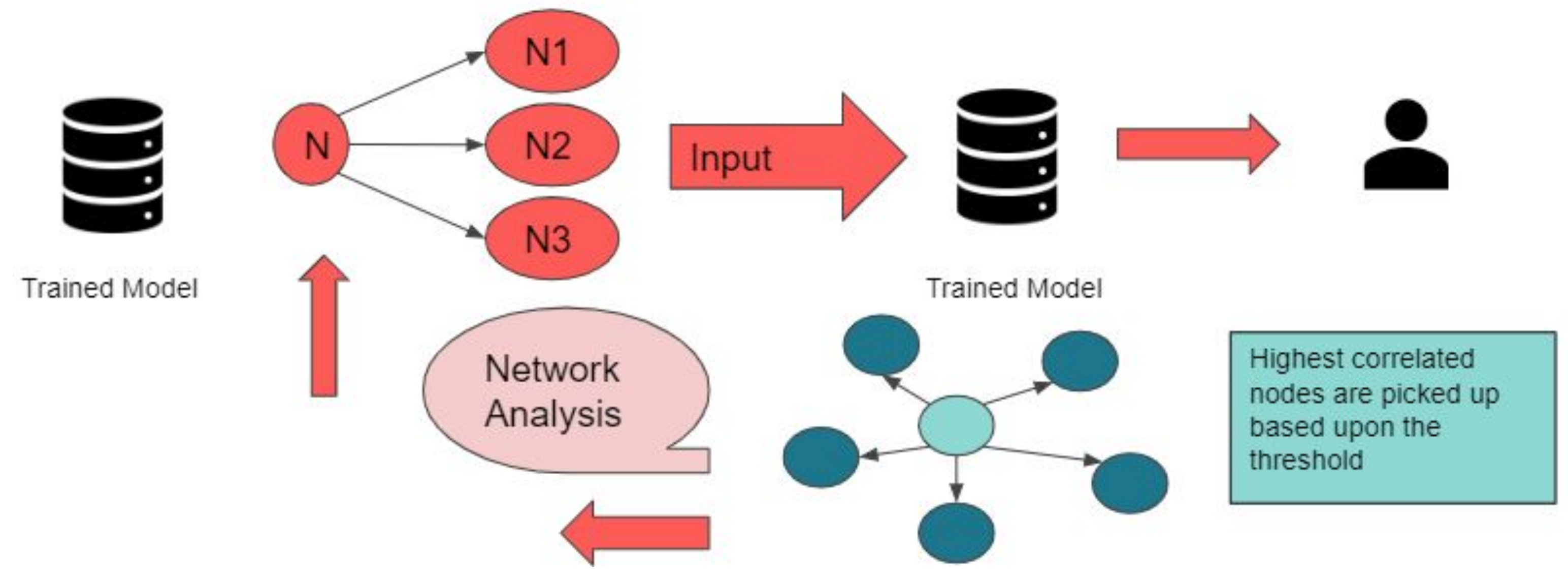
Evaluation and Performance Tuning: GridSearchCV for Hyper-Parameters; Confusion Matrix; Accuracy and Recall

Suspicious Transactions Alerts: Compliance User Feedback

Fig. Transactional Monitoring Design

Network Analysis and Machine Learning

- After system alerts for specific transactions, these will be checked for detailing.
- NA will be used to explore the networks and to identify the strongest correlation between different nodes based upon a predefined threshold.
- Correlation between the nodes will vary based upon the type of data.
- These newly identified nodes with Network analysis are input to the alert generation system.
- The idea is to add additional coverage to alert generation system for increased accuracy and better detection.



Highest correlated nodes are picked up based upon the threshold

Fig. Network Analysis Design

Objectives

- Building a client's Social Profiling data hub using the client's provided Social account information and blending it with KYC data to build a robust system to effectively assess Risk Ratings of clients.
- Implementing Machine Learning models over ETL systems to improve overall Transaction Monitoring.
- Anomaly Detection model to detect real-time unusual activity appearing at the banking online platform which will result in faster reporting to authorities.
- Implementing Network Analysis to link similar-looking accounts to detect layers of money laundering.

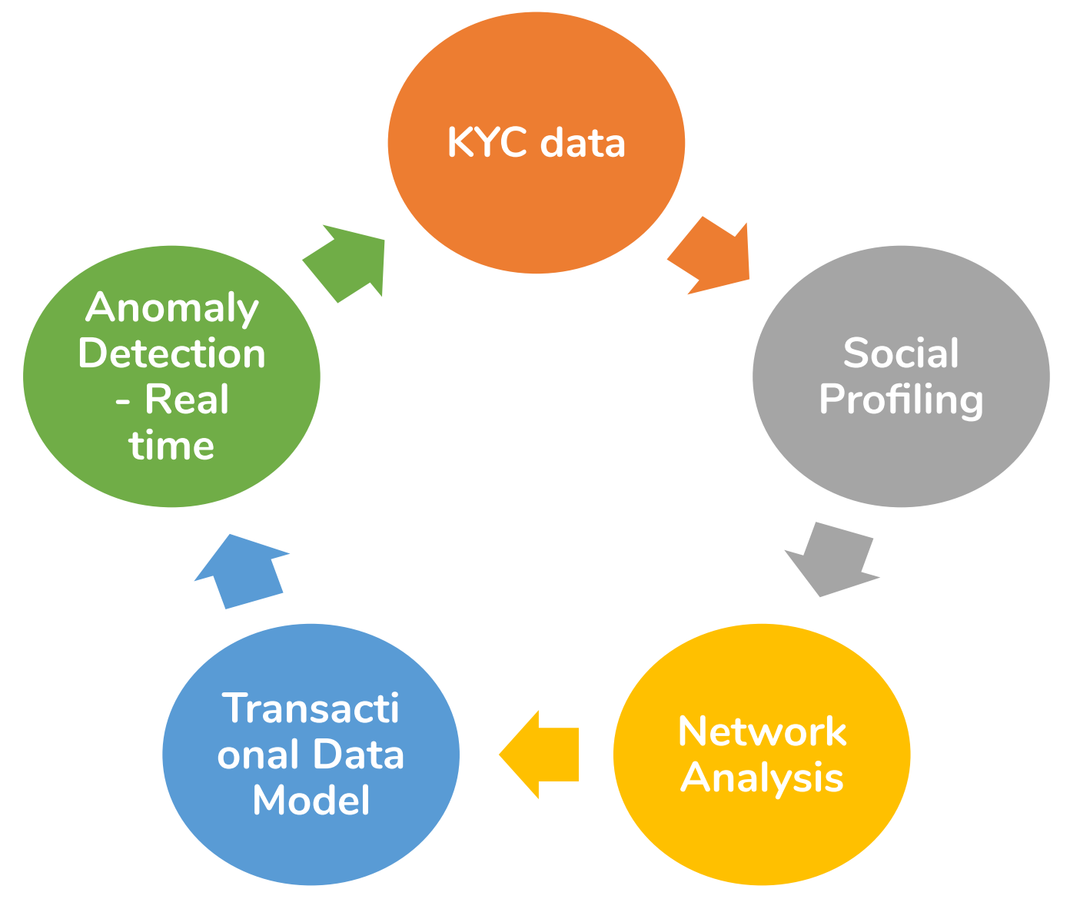
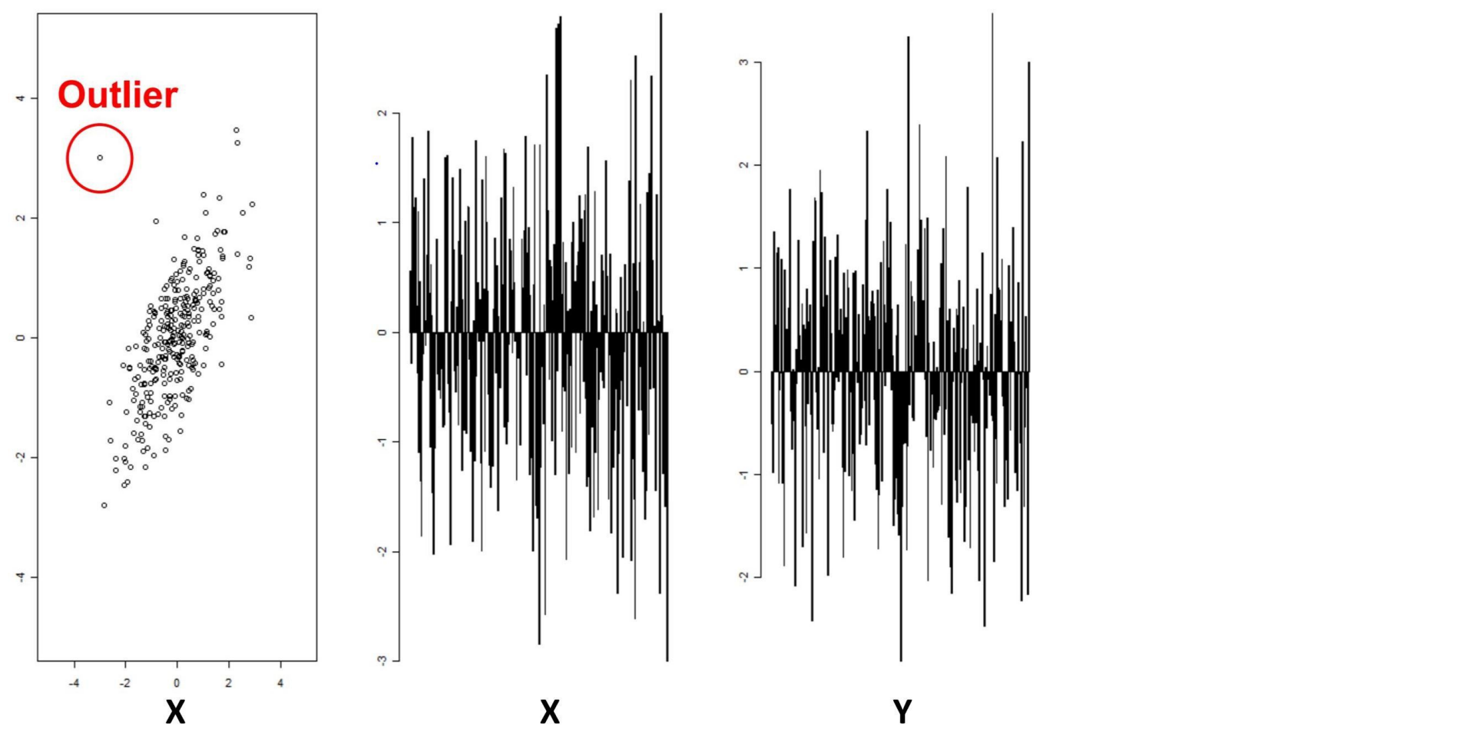
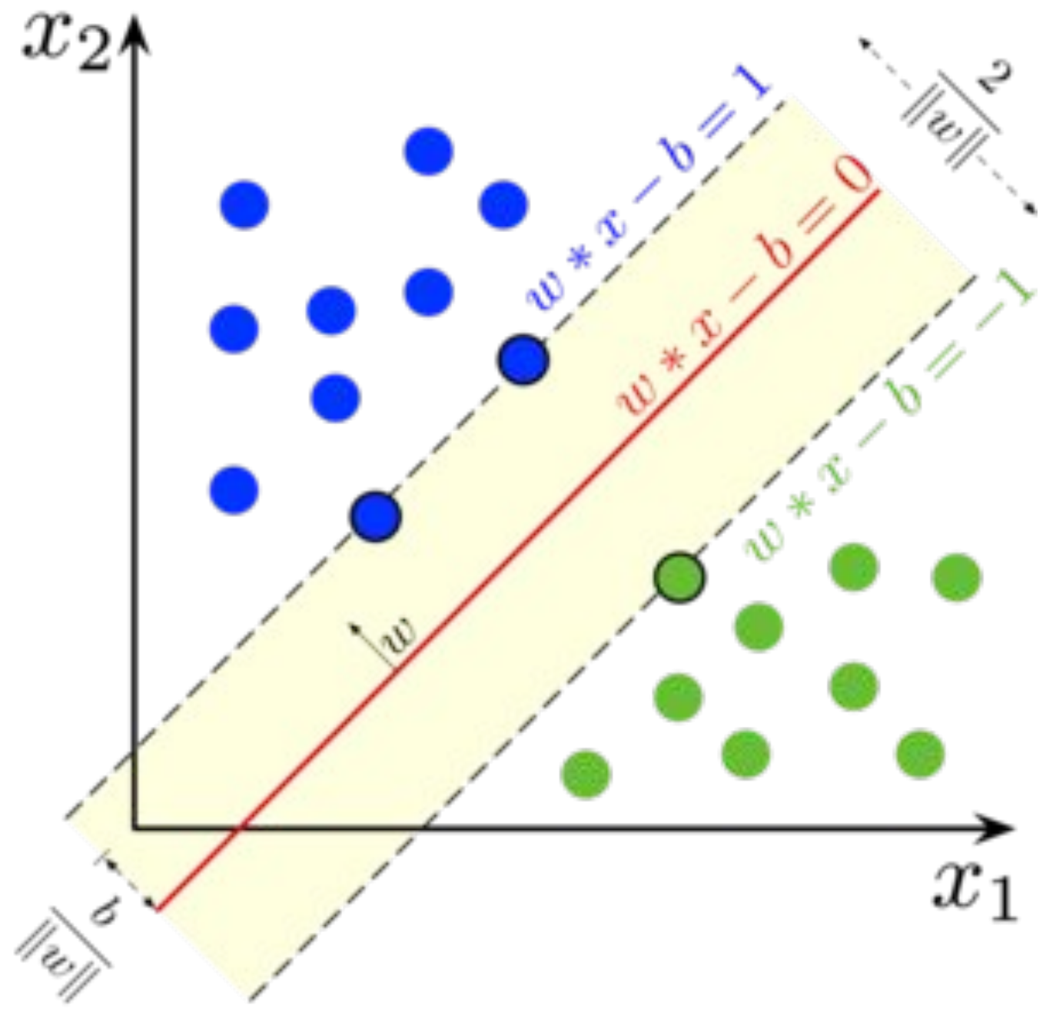


Fig. Design of Machine Learning Approach

Anomaly Detection

- Anomaly detection is a vital part of detecting banking suspicious transactions..
- In higher dimensional problems we create a probability distribution of the data points, the when a new data point comes in we compare its probability distribution to a threshold. If its probability is less than the threshold then it is a anomaly.
- We can use training data to find distances of new points that are considered normal and then detect new data points that stray too far from out threshold
- Using support vector machines, we can classify our data into suspicious and normal activity. Support vector machines work by calculating relationships between data in higher dimensions It does this using kernel functions. It can then create a hyperplane in order to separate the data into two different classes. This will allow our algorithm to classify suspicious activity versus normal activity.





Conclusion

- New alert generation system will provide quality improvement in this complex process and can reduce upto 50% of false positives in compare to ETL system.
- Machine learning models can be trained on historical data to learn patterns and predict suspicious transactions.
- Machine learning Models can be trained on Know your customer(KYC) data to do risk analysis for a customer risk rating as per BSA recommendation on risk assessment.
- Scoring algorithm developed using logistic regression can perform much better compared to the naïve approach.
- Better detection systems in financial institutions can make money laundering difficult for people who exploit it.

References

http://files.acams.org/pdfs/2016/Guidelines_for_Effectively_Auditing_Tax_Evasion_Controls_S_Abraham.pdf
https://www.researchgate.net/publication/323157938_Finding_shell_company_accounts_using_anomaly_detection
https://www.acamstoday.org/using-data-analytics-to-identify-aml-risk/
https://www.moneylaundering.com/news/us-aml-enforcement-continued-climbing-in-2018/
https://www.researchgate.net/publication/251236864_Machine_Learning_for_Social_Network_Analysis_A_Systematic_Literature_Review
https://www.sciencedirect.com/science/article/abs/pii/S1568494619305605
https://ieeexplore.ieee.org/abstract/document/8823264