Workshop 1
Use Case

December 2022
Global Risk Management

Scotiabank®

Agenda

01 Case Theme

02 Background

03 Tasks

04 Data Sources

05 Summary

Agenda

01 **Case Theme**

02 Background

03 Tasks

04 Data Sources

05 Summary

Case Theme: Detect Money Laundering and criminal networks

Motivation:

Financial Institutions are used by criminals to “legitimize” money generated from criminal activities, hence banks are expected by society and regulators to act proactively in detecting and stopping criminals from using the financial system to access their money through effective monitoring and reporting to authorities.

Complication:

- **Number of clients** and **transaction volume** are **large** and only expected to grow in the future
- **Diminishing returns of adding more headcount**
- **Criminals** are constantly **innovating** on ways to use **financial system to “legitimize” their funds**

Ask:

Use suitable **data analytics tools and techniques** to help **Scotiabank detect financial crimes**:

- **Find known high risk people in our customer base using public data**
- **Score clients according to their likelihood of being involved in Money Laundering using transactional data**
- **Enhance scoring and visualize networks using connections between clients**

Agenda

01 ~~Case Theme~~

02 Background

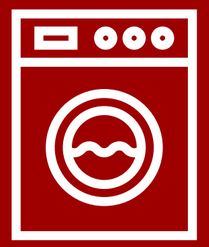
03 Tasks

04 Data Sources

05 Summary

Background

01



What is Money Laundering ?

02

How can we prevent Financial Crimes?

03

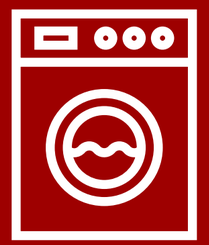
Risk Ratings

04

Customer Connections

Background

01



What is Money Laundering ?

02

How we can prevent Financial Crimes?

03

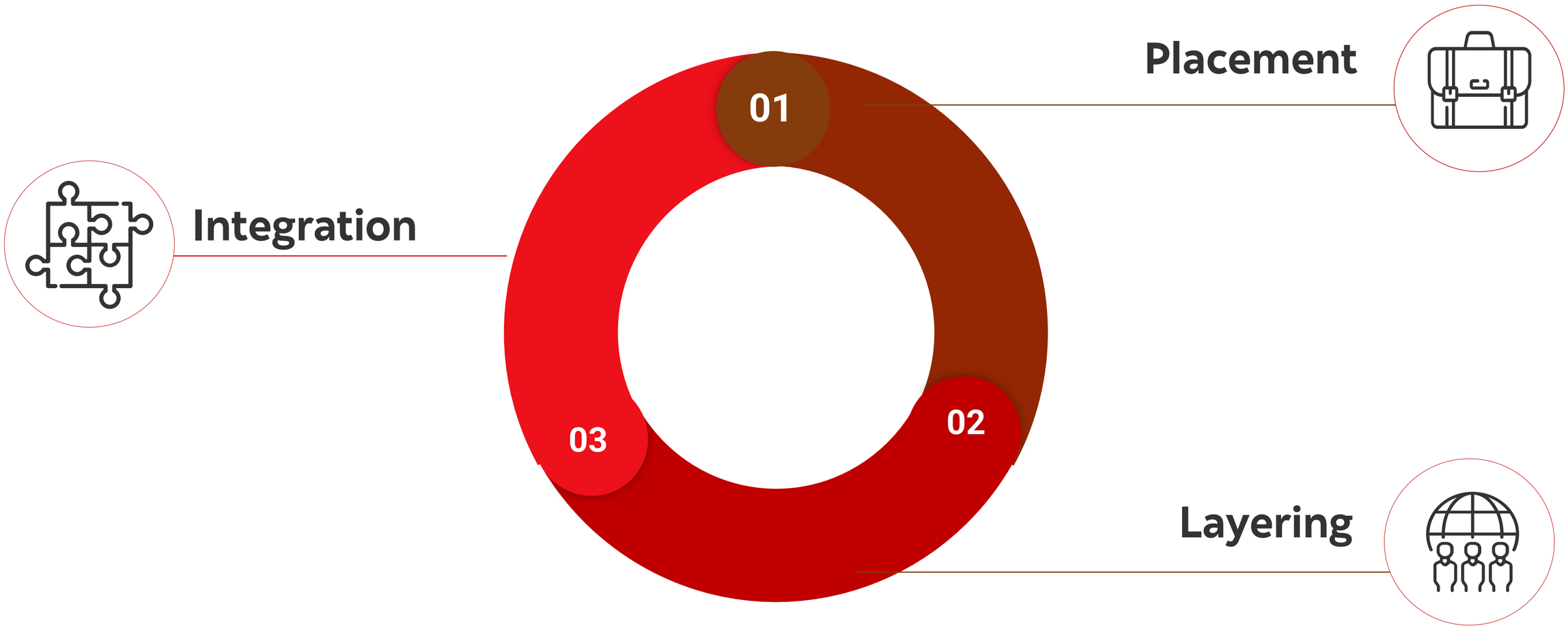
Risk Ratings

04

Customer Connections

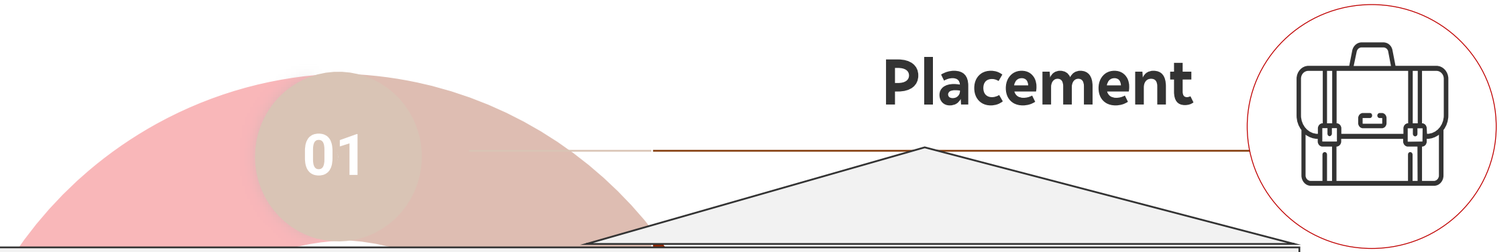
Background | What is Money Laundering?

Definition: Money laundering is the act of “turning the proceeds of crime into cash or property that looks legitimate and can be used without suspicion.”



What is Money Laundering? | Stage 1: Placement

Definition: Money laundering is the act of “turning the proceeds of crime into cash or property that looks legitimate and can be used without suspicion.”



Stage 1: Placement

Money laundering starts by placing funds derived from illegal activities into the financial system in the form of **cash deposits, cheque, money transfers or any type of transaction where money enters the Bank**.

Placement can be identified at the frontline by understanding the source of the funds.

Example: customer deposits a total of **$20,000 in cash** using **multiple transactions** to **avoid reporting and detection** of large cash deposits

What is Money Laundering? | Stage 2: Layering

Definition: Money laundering is the act of “turning the proceeds of crime into cash or property that looks legitimate and can be used without suspicion.”

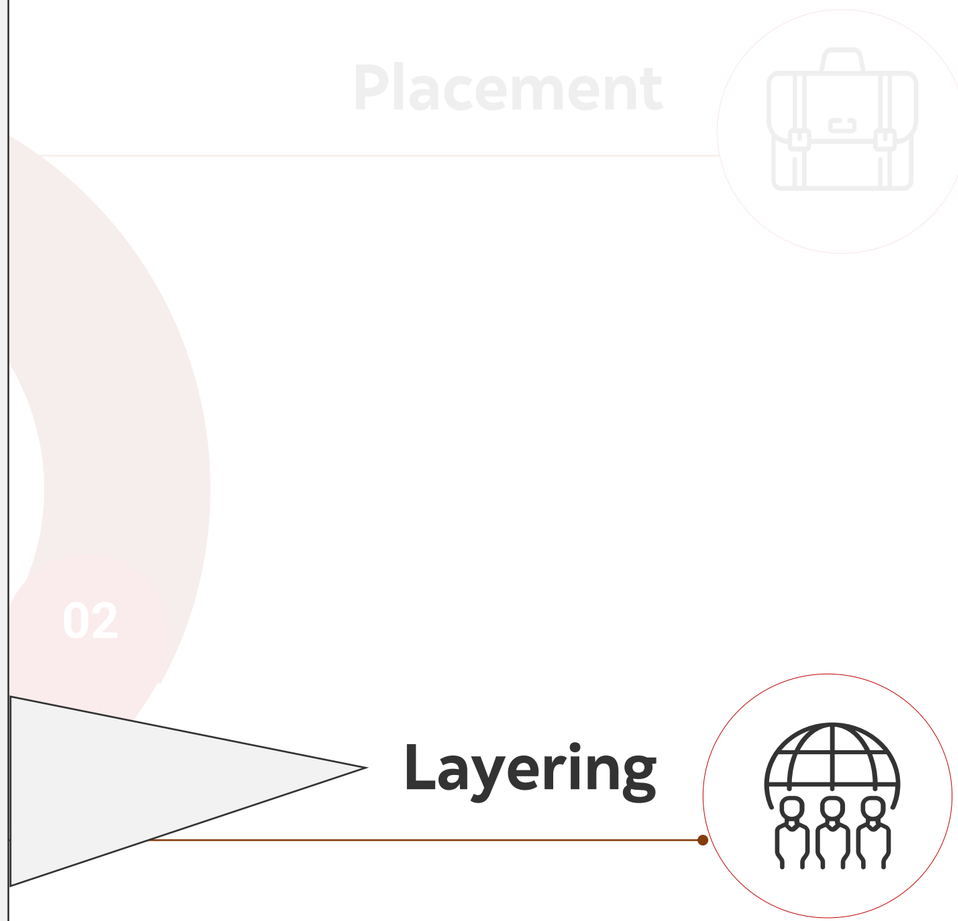
Stage 2: Layering

Second stage is for a criminal to move **funds around to hide their illegal origins**. It may consist of **multiple transactions without clear purpose** to move money between **products, clients, bank, corporations and geographies**

Layering can be detected by understanding transaction purpose and if it is unusual for the customer.

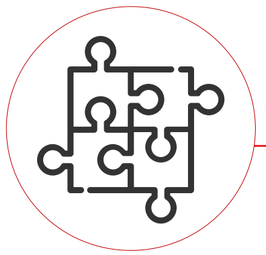
Examples:

- Customer receives an incoming wire from Country A for $20,000, then sends an outgoing wire in the same amount to Country B, or
- Customer makes a deposit to their account, then transfers the funds to another of their accounts with another financial institution



What is Money Laundering? | Stage 3: Integration

Definition: Money laundering is the act of “turning the proceeds of crime into cash or property that looks legitimate and can be used without suspicion.”



Integration

03

Stage 3: Integration

The final stage **integrates the illicit proceeds**, that now appear to be clean funds, **into the economy as "normal"** personal or business transactions. By this stage is **hard to distinguish between legal and illegal money**, and the **criminals** can now **use** these **funds without suspicion.**

Integration can be detected by knowing the client and monitor for odd and unusual transactions.

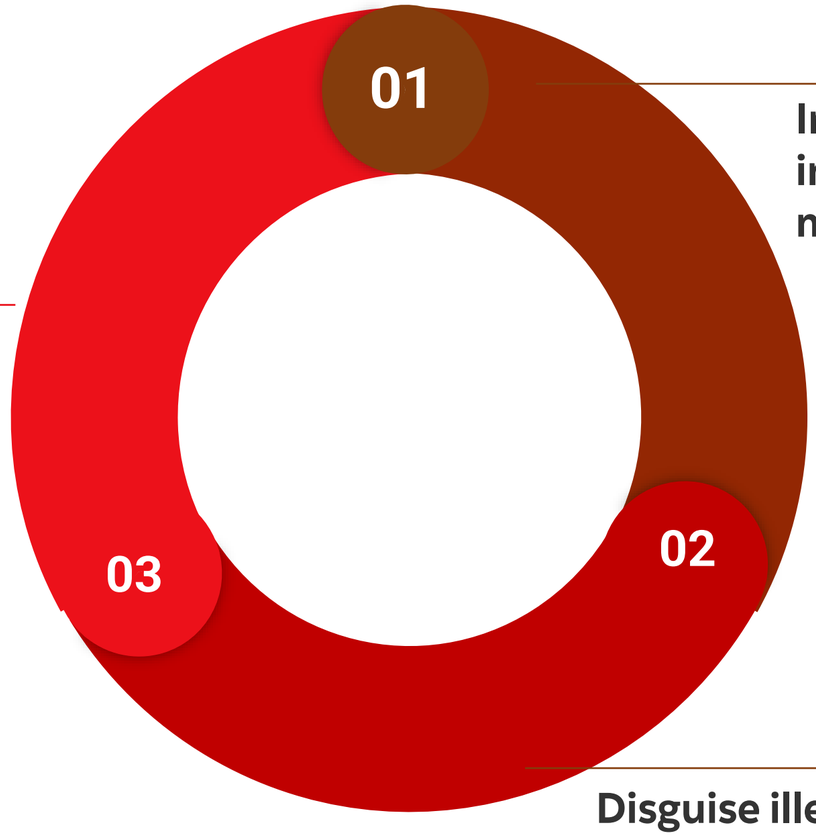
Examples:

- Use illicit proceeds to purchase luxury assets such as real estate, high end artwork, jewelry or vehicles

- A client opens an offshore bank account with a debit or credit card to access the account. They wire illicit proceeds to this offshore account, which are used for withdrawals or purchases

What is Money Laundering? | Summary

Definition: Money laundering is the act of “turning the proceeds of crime into cash or property that looks legitimate and can be used without suspicion.”



Placement

Insert the illegal money into Financial System via multiple deposits

Layering

Disguise illegal funds by moving the money around by transferring between banks, customers, geographies etc.

Integration

Use the “clean money” to buy goods and do “normal” business transactions

Background

01

What is Money Laundering ?

02

How we can prevent Financial Crimes?

03

Risk Ratings

04

Customer Connections

Know Your Customer (KYC)

Ongoing Monitoring

Background | How can we prevent Financial Crimes?

Know Your Customer (KYC)

Ongoing Monitoring

How can we prevent Financial Crimes? | Know Your Customer

Know Your Customer (KYC)

Collect Data

Accurately collect all (KYC) information at on-boarding

- Name, Address, Occupation
- Nature of Business
- Transactional Counterparties
- Source of funds and wealth
- Account activity
- Product Usage

Name Screening

Systematically check new clients name against watchlists

Prohibited of conducting business with

- **Known criminals**
- **Terrorists**
- **Sanctioned parties and countries**

Clients that match any criteria will not be onboard or; assets and accounts will be frozen

Client Risk Rating

Initial Risk Rating is determined at onboard and is either: Low, Medium or High

It is computed on risk factors and takes into account several customer data points, such as KYC information.

Ratings are not static and can change throughout the lifecycle as more information is gathered about the client.

~~Know Your Customer (KYC)~~

Ongoing Monitoring

Ongoing Monitoring

Up-to-date Data

Keeping client information up-to-date

- Recurrently update customer information
- Reassess client data when changes occur, such as opening new accounts, requesting products, adding account holder, change of address etc.

Transaction Monitoring

Systematically monitor transactions volume and frequency

- Use historical data and analytics to deploy automated models to detect and flag abnormal behaviour
- File for further investigations if an alert is raised
- Cash transactions > $10k are reported to FINTRAC

Payment/Name Screening

Continuously monitor client names and payment data with most updated Watchlist

- Watchlists and news are constantly developing, and the bank must be vigilant to detect any recent criminal activity going on
- Use NLP techniques to compare names, extract funds source etc.

Background | How can we prevent Financial Crimes?

~~Know Your Customer (KYC)~~

~~Ongoing Monitoring~~

Background

01

What is Money Laundering ?

02

How we can prevent Financial Crimes?

03

Risk Ratings

04

Customer Connections

Background | Risk Ratings

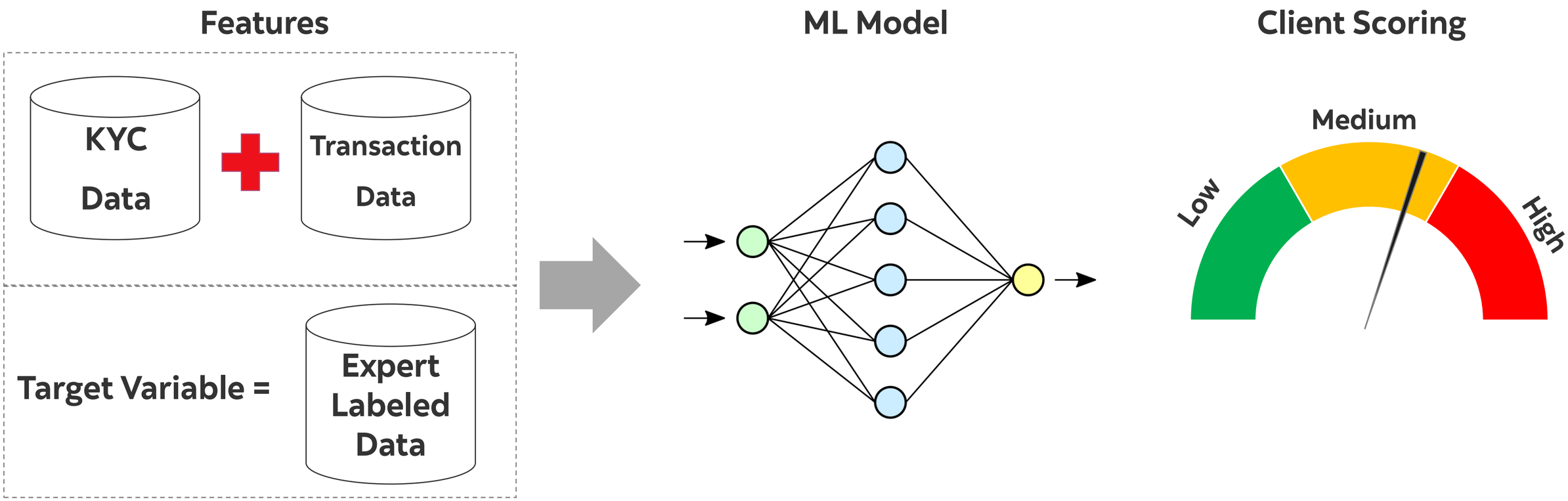
Definition: Risk Ratings is a scale that represents how likely a client is involved in money laundering activities.

High Risk: very high likelihood of engaging in money laundering

Low Risk: highly unlikely of being involved in financial crimes

Medium risk: anywhere between low and high risk

How ratings are computed:



Background

01

What is Money Laundering ?

02

How we can prevent Financial Crimes?

03

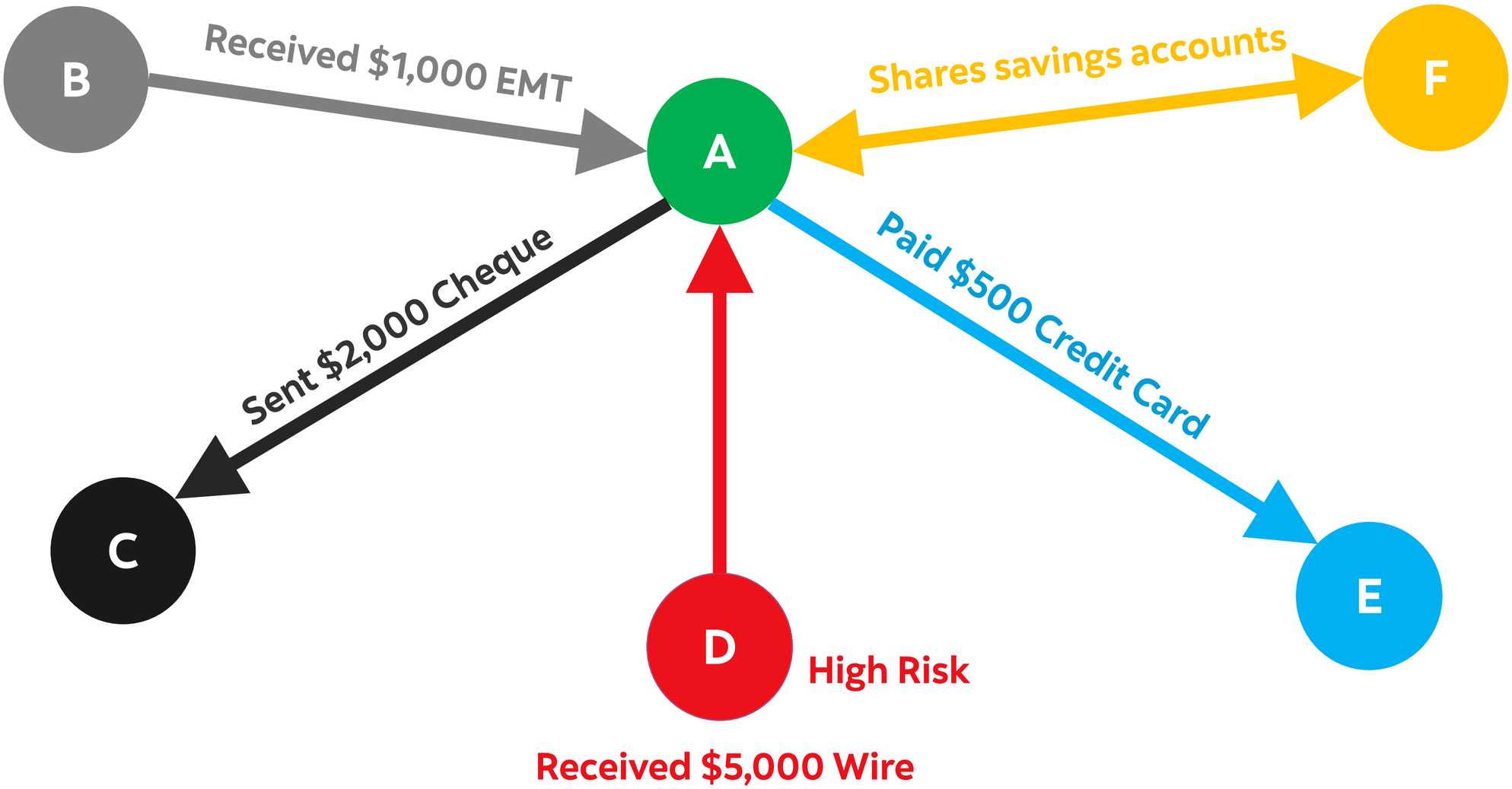
Risk Ratings

04

Customer Connections

Background | Customer Connections

Definition: customer connection is any relationship between clients, either shared account ownership, customer information (same address, email, phone number etc.) or **money transfer**



Agenda

Tasks

1 Name Screening

Detect 50 Bad actors in our customer base using public data sources

Ask:

Find as many bad actors as possible using NLP techniques to match customer names with watchlist and other relevant information given

2 Risk Rating

A. **Classify customers into Low, Medium and High risk.**
B. **Predict Bad actors using results from Task 1 as target variable**

Ask:

Using transactional and KYC data, create models to

A. **Classify clients into Low, Medium or Risk**

B. **Predict bad actors using your results from Task 1 as dependent variable**

3 Improve model using Graph data

Add customer connections information to improve Task 2 models or use a graph model directly

Ask:

Improve your Task 2 models by using client connections to either:

- **Extract new features**
- **Fit graph models directly**
- **Visualize interesting networks**

Tasks

Task 1: Name Screening

Task 2: Supervised Learning

2A: Risk Rating

2B: Predicting Bad Actors

Task 3: Improve your models

Task 1: Name Screening

Task 2: Supervised Learning

2A: Risk Rating

2B: Predicting Bad Actors

Task 3: Improve your models

Tasks | Task 1: Name Screening

Problem: there is a total of 50 bad actors in our customer base that were extracted from OpenSanctions (watchlist), amongst them we have warmongers, sex trafficker, drug dealers etc.

Ask: using appropriate NLP techniques for name matching and other relevant KYC data, find as many bad actors as possible.
Hint: formulate a similarity scoring methodology, and assign bad actors based on a cutoff

Example:

Customer KYC

Name	DOB	Gender
Pereira, Alex	06-Jan-92	Male
John Smith	13-Jul-79	Male
Lara Goldstein	1989-08-21	Female
…	…	…
Prsa Vafai	1994-05-28	Male
…	…	…

Bad Actor List

Name	DOB	Gender
Alexys Pereira	06-Jan-92	Male
Parsa Vafaie	1994-05-28	Male
…	…	…
…	…	…
…	…	…

Compare names

Compare Date of Birth

~~Task 1: Name Screening~~

Task 2: Supervised Learning

2A: Risk Rating

2B: Predicting Bad Actors

Task 3: Improve your models

~~Task 1: Name Screening~~

Task 2: Supervised Learning

2A: Risk Rating

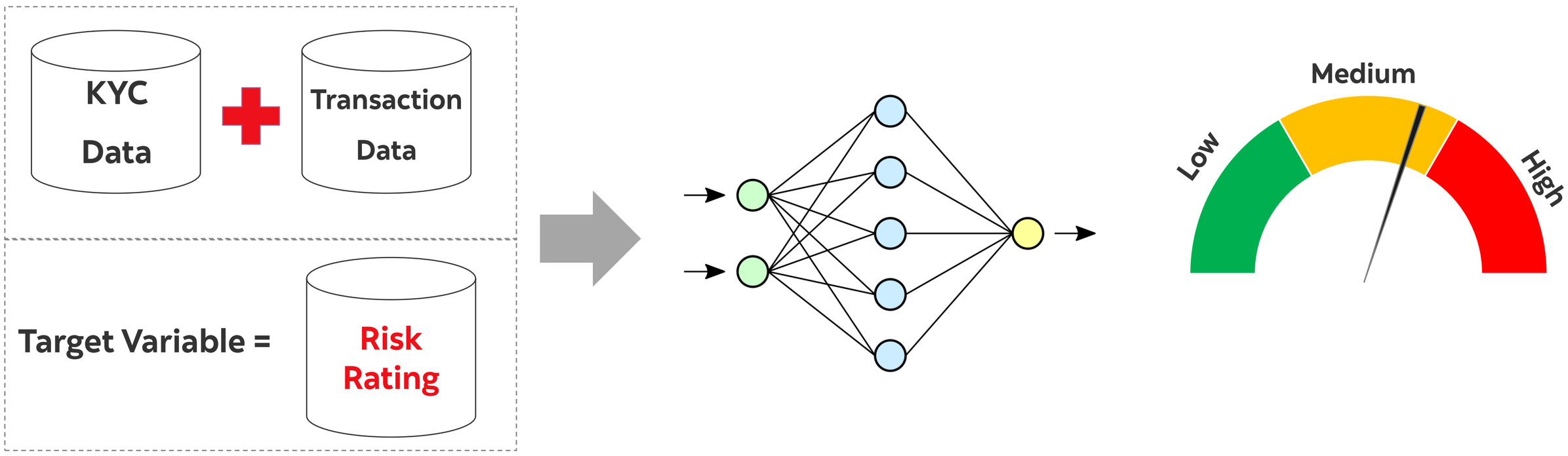
2B: Predicting Bad Actors

Task 3: Improve your models

Tasks | Task 2A: Risk Rating

Problem: there are too many clients in our customer database and heterogenous, we would like to automate this process to be as efficient and reliable as possible

Ask: using predictive modelling classify all 1M customers into three risk buckets: Low, Medium and High. As features you must use KYC data and transactional data, your target variable is the risk rating provided



Similar to the Risk Ratings presented earlier, we would like you to produce your own solution to this problem

~~Task 1: Name Screening~~

Task 2: Supervised Learning

~~2A: Risk Rating~~

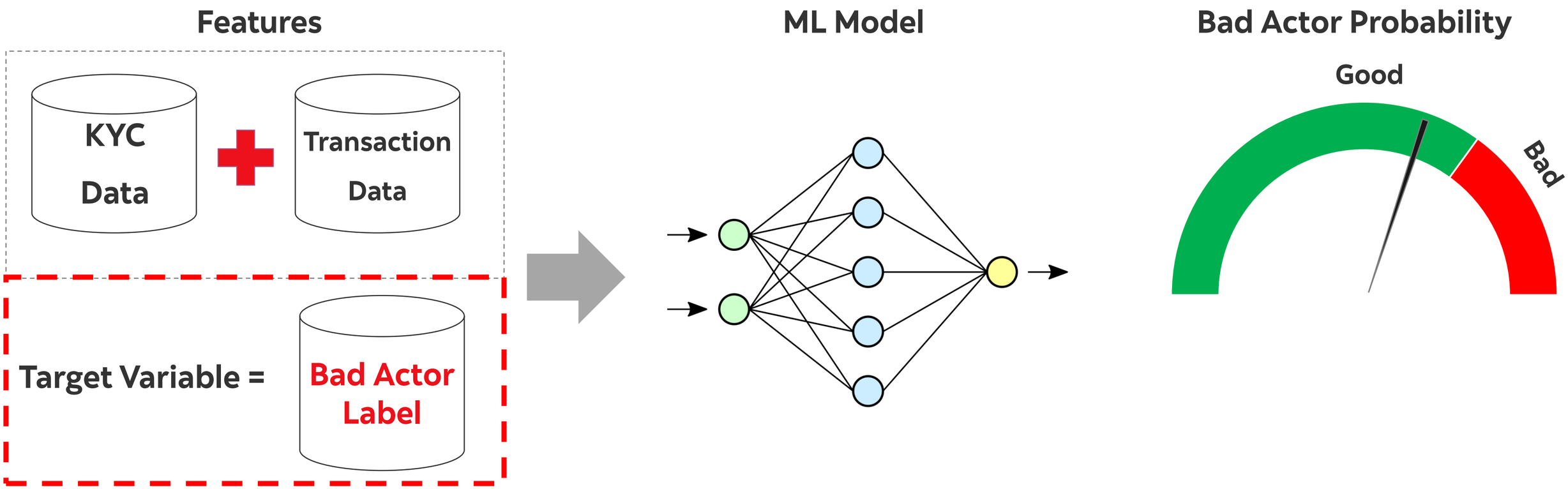
2B: Predicting Bad Actors

Task 3: Improve your models

Tasks | Task 2B: Predicting Bad Actors

Problem: bad actors are hard to detect, and their presence is a high risk to the bank, we also would like to have an automated process for screening

Ask: similar to Task 2A you must use KYC and Transactional data to produce a model to estimate the likelihood of a customer being a bad actor based on your results of Task 1.



The difference between 2A and 2B is the target variable: replace Risk Rating with your own Bad Actor Label

~~Task 1: Name Screening~~

~~Task 2: Supervised Learning~~

~~2A: Risk Rating~~

~~2B: Predicting Bad Actors~~

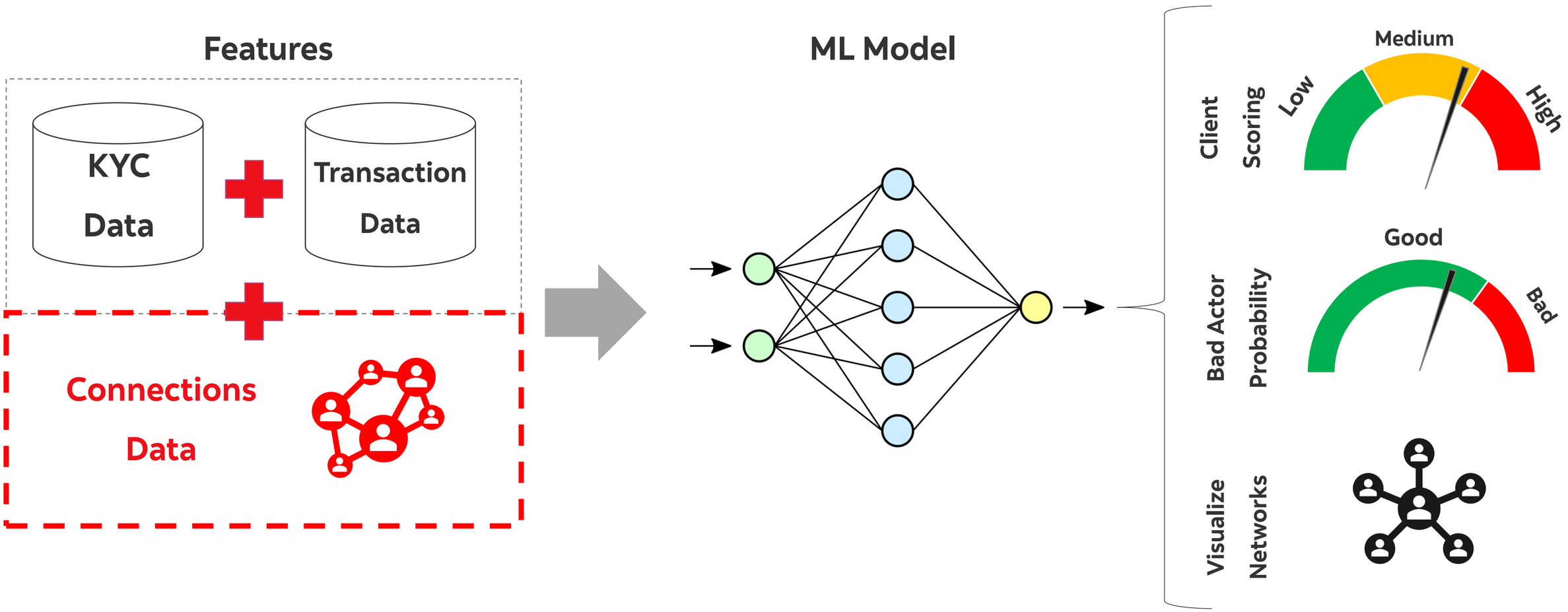
Task 3: Improve your models

Tasks | Task 3: Improve your models

Problem: KYC and Transactional data provide useful information but fails to capture one of the most important components of detecting financial crimes – **money flow**

Ask: improve your Task 2 models by using client connections data to either

a. Engineer new features
b. Build a Graph model i.e. model that ingests graph data directly
c. Visualize and detect interesting networks using unsupervised learning



Tasks | Task 3: Improve your models

~~Task 1: Name Screening~~

~~Task 2: Supervised Learning~~

~~2A: Risk Rating~~

~~2B: Predicting Bad Actors~~

~~Task 3: Improve your models~~

Agenda

~~01 Case Theme~~

~~02 Background~~

~~03 Tasks~~

04 Data Sources

05 Summary

Data Sources | Overview

Overview: you will be given 3 Scotiabank data sets and one open-source data for Bad Actors matching

Open Source: you should explore the following link to get a list of Bad Actors
https://www.opensanctions.org/datasets/default/

Scotiabank Data: synthetic data sets generated using real bank data for this case competition

- **Nodes Data**: main data source for this competition, contains KYC, Transactional data and Risk Rating
  File name: UofT_nodes.csv
  Number of Rows: 1M, one row per CUSTOMER_ID
  Number of Columns: 20 columns (1 target column); RISK = Target variable for Task 2

- **Edges Data:** shows the connections between clients i.e. amount of money sent via EMT from one customer (source) to another (target)
  File name: UofT_edges.csv
  Number of Rows: 466k

- **Occupation Data:** lookup table that maps an occupation (code) to their risk level of being involved in financial crimes
  File name: UofT_occupation_risk.csv
  Number of rows: 841

Agenda

~~01 Case Theme~~

~~02 Background~~

~~03 Tasks~~

~~04 Data Sources~~

05 Summary

Summary

Overview: use appropriate modelling techniques to name screen bad actors and classify customers according to their likelihood of being involved in financial crimes using KYC, Transactional and client connections data

Data: Three synthetic Scotiabank data sets + One open-source data for Bad Actors matching

- **Open Source:** https://www.opensanctions.org/datasets/default/
- **Nodes Data:** KYC + transactional data **-** UofT_nodes.csv
- **Edges Data:** client connections - UofT_edges.csv
- **Occupation Data:** lookup table - UofT_occupation_risk.csv

Tasks:

- **Task 1 (Name screening)**: find bad actors in Nodes data using OpenSanctions data

- **Task 2 (Supervised Learning):** use Nodes Data + Occupations data to build models
  - **2A (Risk Ratings):** assign each client to a risk bucket Low, Medium or High
  - **2B (Bad Actors):** compute the likelihood a client being a Bad Actor **- use your Task 1 output as target variable**

- **Task 3 (Improve model):** add Connections Data to Task 2 models or build a model that ingests Graph Data natively.

Summary | Suggested Approach

1. **Get access and familiarize yourself with the data provided**

2. Do an extensive Exploratory Data Analysis to understand correlations and patterns

3. Research papers that use machine learning/statistical techniques to solve similar problem – leverage UofT library access, seek for insights from other domains other than Finance

4. Define a robust validation methodology to assess your models' performance

5. Select relevant evaluation metrics for each problem

6. **Start with simple models, learn from their outputs and leverage those insights to:**
    1. Understand the patterns in the data
    2. Draw sensible conclusions for your presentation
    3. And/or have ideas for more complex models

7. **Plan yourself ahead of time:**
    1. Define by when you want to have presentation and video ready for submission
    2. Make weekly goals to keep momentum going
    3. Split tasks among team members – everyone has something to contribute, play to their strengths

Disclaimer: we are interested in sensible, robust methodologies, insights derived and how well you communicate your work to an audience.

Agenda

~~01 Case Theme~~

~~02 Background~~

~~03 Tasks~~

~~04 Data Sources~~

~~05 Summary~~

Q & A

Thank you

Good Luck!

December 2022
Global Risk Management

Scotiabank®