

Risk Profiling LLD – 1.0

FINNet 2.0

**CONFIDENTIALITY CLAUSE**

This document is the property of Financial Intelligence Unit - India under Project FINnet 2.0. All ideas and information contained within the document is the intellectual property of FIU-IND. These documents are not for general distribution and are meant for use solely by the person/persons to whom it is specifically issued to. Copying or unauthorized distribution of these documents, in any form or means including electronic, mechanical, photocopying or otherwise is illegal.

DOCUMENT CONTROL

|  |  |
| --- | --- |
| Document Version Control Note | |
| Name of the Document | Software Design Document (SDD)– FINnet 2.0 Risk Scoring |
| Document Version Number | 1.0 |
| Effective Date | 22-02-2022 |
| Approved By | Tarun Gupta, Saurabh Shrivastava, Usha V |
| This Revision Supersedes | NA |
| Document Classification | Restricted |
| Distribution List | FIU-IND, PMU, MSP |
| Access Level | Read Only |

REVISION HISTORY

| Version | Release Date | Prepared By | Reviewed/  Approved By | Description of changes made | Section of document Impacted |
| --- | --- | --- | --- | --- | --- |
| 0.1 | 22-02-2022 | Robin Singh/ Kuldeep Rana | Tarun Gupta | 1.0 |  |

REFERENCE DOCUMENTS

| S. No. | Document Name | Document Description |
| --- | --- | --- |
| 1 | SRS 3.2 & 3.6 | Final Version |
|  |  |  |
|  |  |  |

Table of Contents

[1 Purpose of Document 5](#_Toc135125684)

[2 Scope of Document 5](#_Toc135125685)

[3 Technology Architecture 6](#_Toc135125686)

[3.1 Technology Landscape 6](#_Toc135125687)

[3.2 Fosfor Architecture 6](#_Toc135125688)

[3.3 FOSFOR connection 7](#_Toc135125689)

[4 Basic Assumptions and Limitations of the System 8](#_Toc135125690)

[4.1 Assumptions. 8](#_Toc135125691)

[5 Risk Details 8](#_Toc135125692)

[6 Solution Architecture 10](#_Toc135125693)

[6.1 ML models Workflow 10](#_Toc135125694)

[6.2 Entity (Inherent) Risk Data Flow 12](#_Toc135125695)

[7 Peer Grouping 12](#_Toc135125696)

[7.1 Peer Grouping 14](#_Toc135125697)

[7.2 Algorithm Used 14](#_Toc135125698)

[8 Location Risk 15](#_Toc135125699)

[8.1 Available data: 15](#_Toc135125700)

[8.2 Assumptions: 15](#_Toc135125701)

[8.3 Calculation matrix: 16](#_Toc135125702)

[8.4 Location Risk Flow 17](#_Toc135125703)

[9 Entity Risk: 18](#_Toc135125704)

[9.1 Risk Calculation: 20](#_Toc135125705)

[10 Transaction Risk: 21](#_Toc135125706)

[11 Network Risk 22](#_Toc135125707)

[12 Implementation 24](#_Toc135125708)

[Implementation Plan 25](#_Toc135125709)

[12.1 Location Risk 25](#_Toc135125710)

[12.2 Transaction Risk 25](#_Toc135125711)

[12.3 Entity Risk 25](#_Toc135125712)

[13 Conclusion 26](#_Toc135125713)

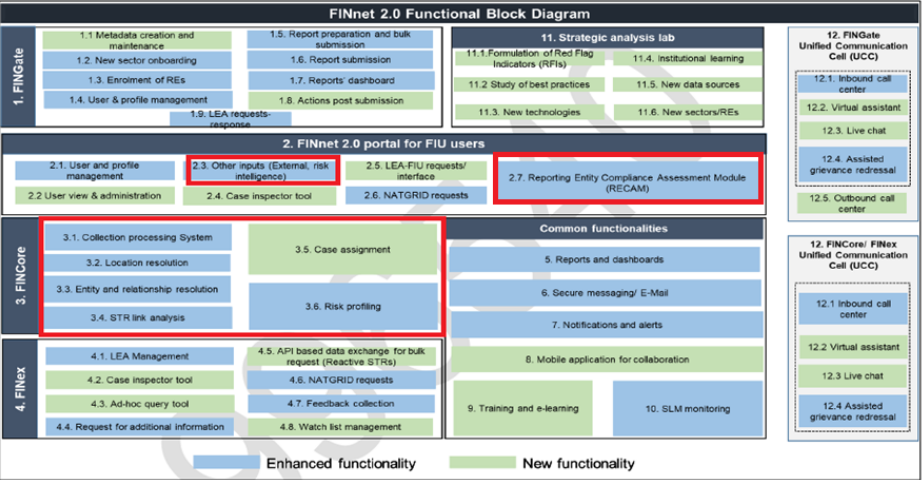
**Module Introduction**

# Purpose of Document

This document contains low level Design of FINCore 2.0\_Risk Scoring. In this document, design of analytics used in risk scoring has been described. This document would continue to evolve as the program progresses and with evolution of technologies and architecture in Industry, same would be reflected in this document on need basis.

# Scope of Document

The scope of the document is to cover FINCore Risk Scoring ML practises to build robust system to identify the probable risk factor and assign it to appropriate entity and legal has been covered. All the ML and rule-based model are designed in a way to reduce the manual intervention of FIU Analysts simultaneously providing them better insights on identifying the risk score of individual as well as organizational. Also, this new system aims to provide better results by incorporating new laws, change in fraud types and new scenarios as and when identified.



# Technology Architecture

## Technology Landscape

The below listed tools and technologies have been used in building the solution to provide insights on all the problems, below is the description for usage to technology:

* Fosfor Refract & Spectra – will be used as a processing engine which helps processing the data faster
* AI/ ML – models will be created as a backend engine to process risk scoring and GoS
* ETL servers – will be used to fetch and store data
* SQL – will be used to build and run queries to resolve the risk scoring dynamic rule engine

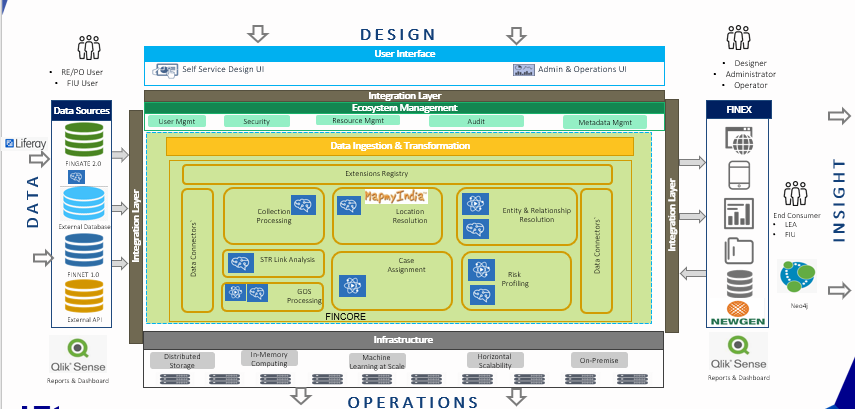
|  |  |  |  |
| --- | --- | --- | --- |
| **Layer** | **Components** | **Technology** | **Vendors** |
| API | API Gateway | WSO2 | Open source |
| Application | AI/ML Engine | Fosfor AI Logistics | LTI |
| Application | Data Visualization | Linkurious | Linkurio.us |
| Data | Data Streaming | Fosfor Decision | LTI |
| Data | Data Processing | Fosfor Decision | LTI |
| Data | Graph Data | Neo4j | Neo4j |
| Data | Transactional Database | MS SQL Server | Microsoft |
| Integration | ETL Server | Fosfor Decision | LTI |

## Fosfor Architecture

Fosfor Refract & Spectra will be used to perform all the high-end processing of data using risk rule engine and ML based models, it provided high processing capability along with scheduling features which enables the opportunity to keep things going by itself in the production environment, Scheduling feature will have a certain pre-defined time frequency on which the Fosfor will check the new data availability in FINGate 2.0 for further processing

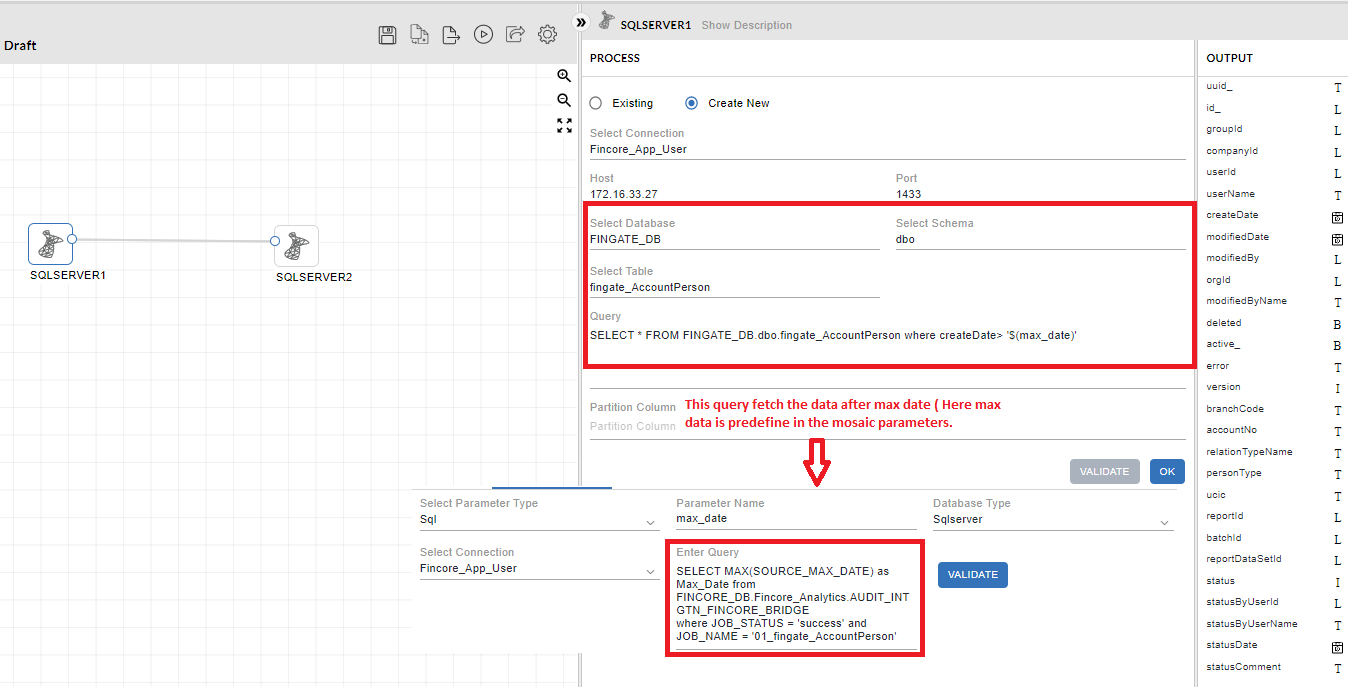
Also, the ML models evolution will be taken care by Fosfor.Ai in the pre-prod environment

In the below listed diagram, the highlighted cell “Risk Profiling” is the area where the entire exercise is getting operated. All the rules-based models & risk-based ML models will be available there only



## FOSFOR connection

The below listed diagram explains how the Fosfor will help in fetching the data from FINGate 2.0 & FINCore 2.0 table, the highlighted box talks about the table selection and standard connection query of Fosfor with the data base. This is an important exercise which will help in getting all the required data in Fosfor for further processing.



# Basic Assumptions and Limitations of the System

## Assumptions.

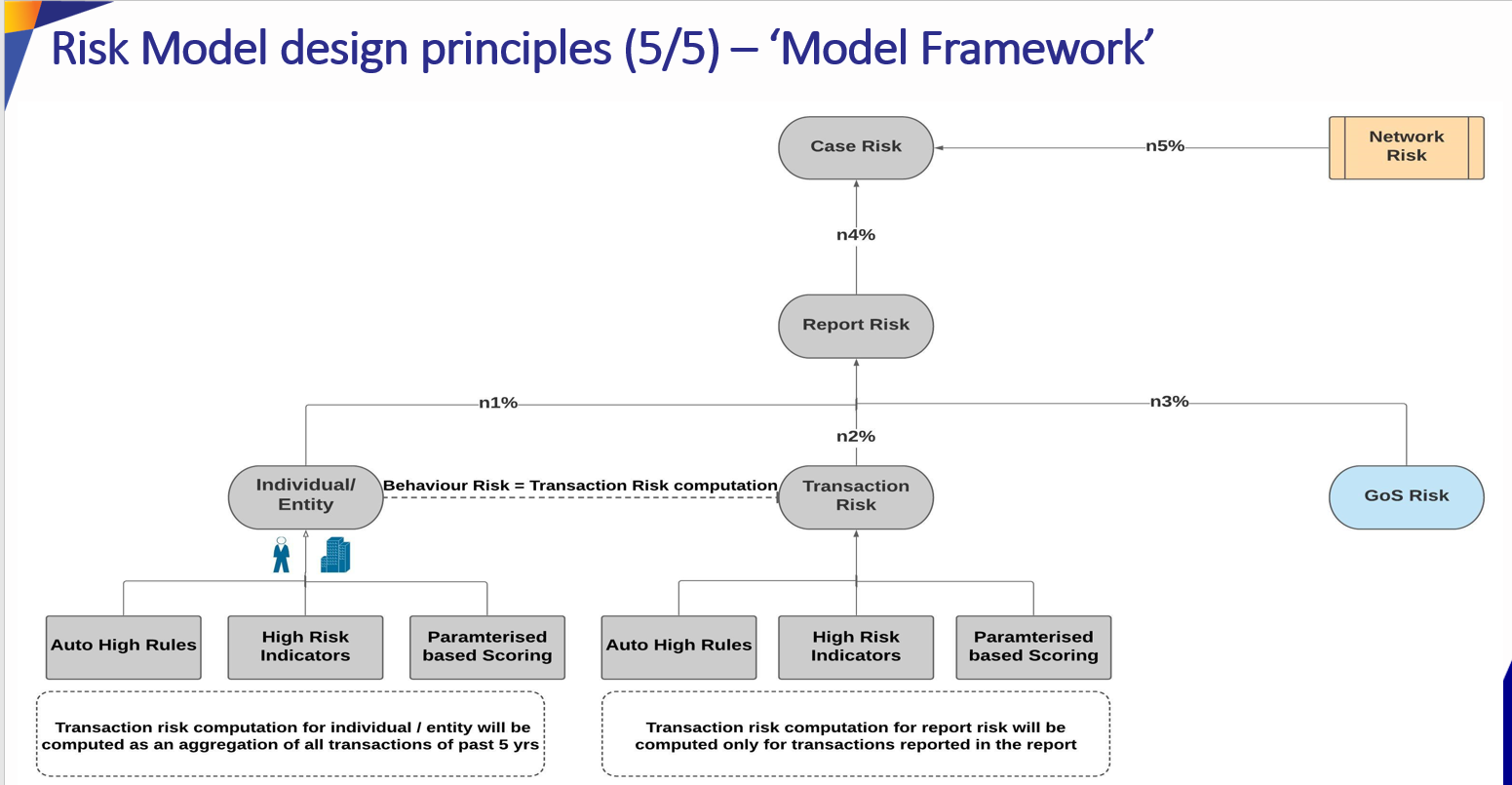
* Requirements Change won’t have any major impact on Design
* The rules are configurable. However, any modification/ addition of rule will be considered as enhancement as it will require additional development efforts.
* All the threshold values will be given by FIU.

# Risk Details

* **Location Risk** – This process involves rating Locations based on different risk factors, this process will be taken care by Machine Learning models and Dynamic Risk modelling
* **Entity** – Entities will be peer grouped based on Machine learning algorithms and will be scored applying Heuristic rule engine
* **Transaction Risk** – It includes calculation of transaction risk for all the reported transaction in any given report.
* **Report Risk** – This will be overall risk of report based on the transaction and entities involved in any given report.
* **Network Risk** – this will help in identifying the risk of all individuals/ Legal associated with the reported transaction
* **Case Risk** – This will be calculated considering all the above risk calculated and provides holistic view of risk involved in any given case.

The above risk will also be supported by 223 dynamic rules, which will run sequential to provide additional value to risk identification and entire journey of a case from being reported till investigation

Framework of the Risk Model design which is also included in Director presentation is as below:



# Solution Architecture

## ML models Workflow

The below diagram explains the working components of ML model along with risk rule engine in production environment.

All the data will be coming from FINGate relevant tables through Fosfor in the production environment, Fosfor will check the new data availability as EOD jobs using Fosfor scheduling feature. All the new data will be processed in below mention sequence (A standard process in mention below for all the reports):

* Data Fetch using Fosfor from Fingate 2.0 & FINCore 2.0
* Risk Rule engine will trigger
* All the rules will run in sequence including Auto high-risk rules
* ML model will run to predict the risk of transaction, report, case, network & entity

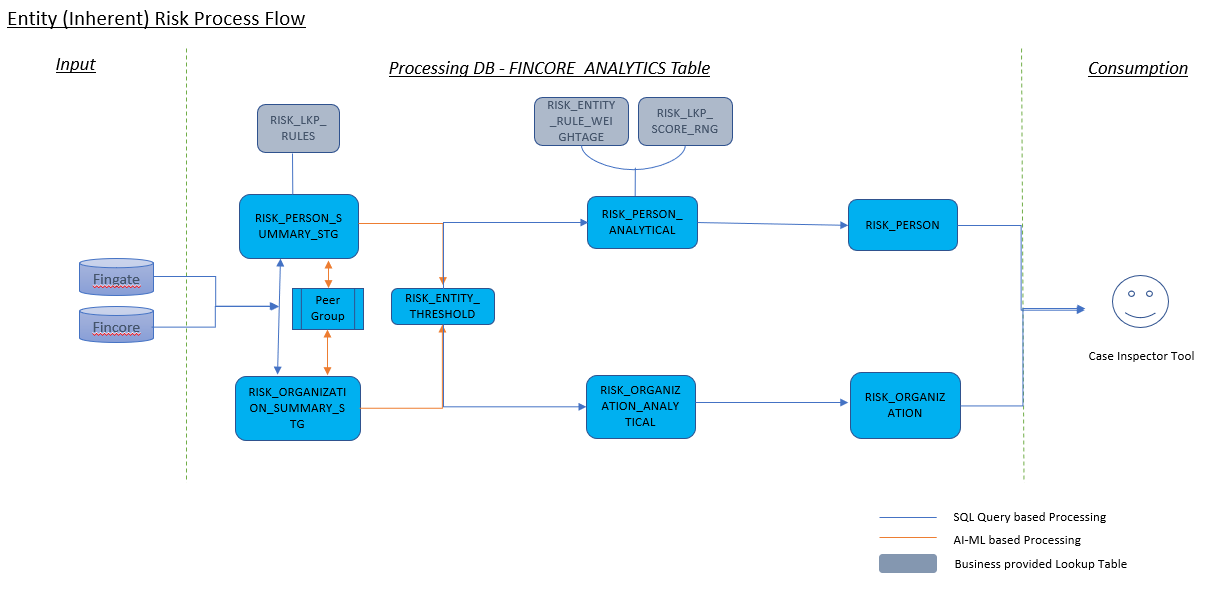
A picture containing text, clock, screenshot

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated

## Entity (Inherent) Risk Data Flow

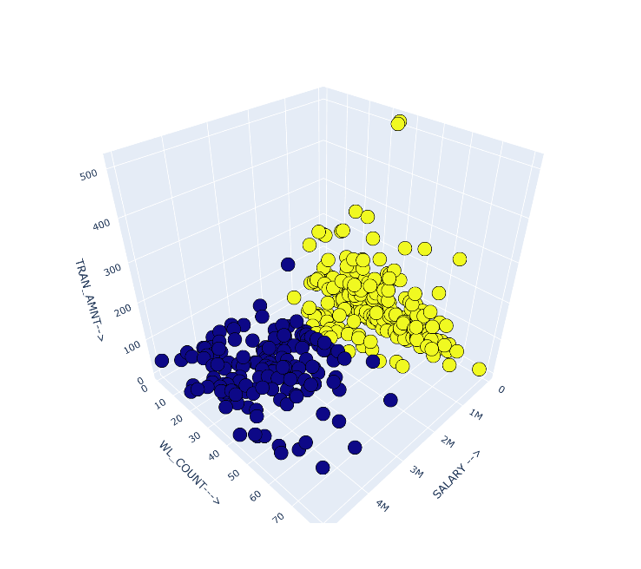


# Peer Grouping

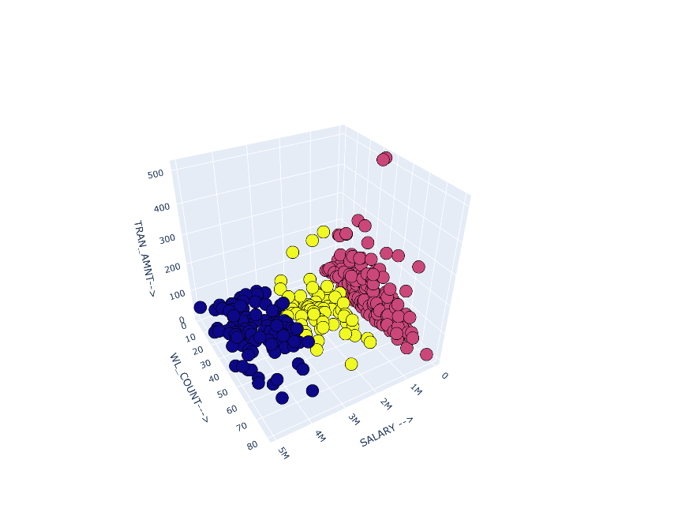
Diagram

Description automatically generated

* The custom algorithm to identify the significant variables for Peer Grouping will have the following steps:
* Converting the variable is numeric and scaled, for example using StandardScaler () and its fit\_transform() method
* Choose maximum available data of variables you want to retain (maxvars), the minimum and maximum number of clusters (kmin and kmax) and create an empty list: selected variables.
* Loop from kmin to kmax. Then, using every variable in turn, record the silhouette value for every combination of variable and number of clusters (from kmin to kmax), using K-Means.
* Choose the variable giving the maximum silhouette value, add it to selected variables and remove it from the list of variables to test.
* Repeat the process in 2 and 3 by using the selected variables list and adding to it each remaining variable in turn, until some stop criterion has been reached (in this case the number of variables to keep, maxvars).



**Sample Output with 2 clusters**



**Sample Output With 3 clusters**

After the feature selection the elbow method will be used to decide the number of cluster since it is the best method statistically to identify the k numbers

## Peer Grouping

1. Data will be extracted from Risk\_Entity\_Summary\_Stg stored in Fincore\_analytics.
2. Clustering will be performed on both entities type i.e., Organization and individual.
3. .
4. Each entity will be assigned a peer group based on ML clustering.
5. For each cluster under each profession, threshold will be calculated for CTR count, STR count, WL count etc.
6. Threshold values will be stored in threshold table.

## Algorithm Used

1. Peer grouping is implemented using k-means clustering algorithm.
2. Number of clusters formed for each profession will be predicted by the K-means algorithm based on the metric called silhouette score.
3. Number of clusters can also be predefined for each profession.

After the complete Run, the algorithm will form different clusters based on Profession, Income Range & Transaction amount. \

All the thresholds will be stored in SQL table RISK\_LKP\_ENTITY\_THRESHOLD to refer if the previous reported entities are reported again.

Assumption of Peer Grouping:

* Assumption of Peer Grouping:
  + we will have to go by income range
  + if Income range not available then consider credit amount
  + if credit amount is not available then consider debit
* peer group having less than 30 (<30) members will be assigned to “others cluster”

# Location Risk

## Available data:

* Transactional Data for STR’s mentioned in above Case ID
* CTR and NTR Enriched data w.r.t. Case ID
* GoS Activity
* KYC data related to accounts reported in STR’s
* Account number mapping with Location.

## Assumptions:

* Location considered for RISK is for the account reported in given reports.

**Rules Implemented to Assign Location Risk:**

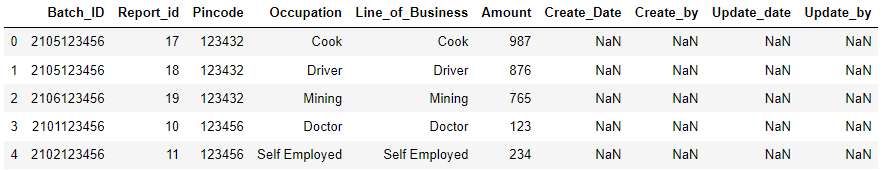
| RULES | Description | Location Marked |
| --- | --- | --- |
|  |  |  |
| Rule 1 | List of High-Risk locations provided by the FIU will be flagged as Auto High-Risk Locations. |  |
| Rule 2 | If an entity’s activity falls under P1 category, that location is flagged as high-risk location. |  |
| Rule 3 | If FIU has raised an alert for that entity, that location is flagged as high risk. |  |
| Rule 4 | If an entity’s activity falls under P2 category and its occupation falls under P1 category (High Risk Occupations), that location is high risk. |  |
| Rule 5 | When there is significant (TBD) percentage increase in STR Count between previous quarter and current quarter |  |
| Rule 6 | When there is significant (TBD) percentage increase in CTR count between previous quarter and current quarter |  |
| Rule 7 | When there is significant (TBD) percentage increase in CTR amount between previous quarter and current quarter |  |

## Calculation matrix:

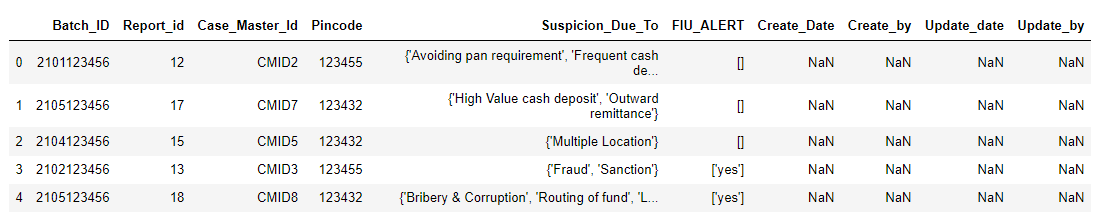
|  |  |  |
| --- | --- | --- |
| **RISK VALUE MATRIX** | | |
|
|  | ***\*\*Rules to be applied SEQUENTIALLY*** |  |
|  | **LOCATION RISK RULES** | **RISK\_VALUES** |
| 1 | LOCATION IN AUTO HIGH LOCATION LIST | 10 |
| 2 | ACTIVITY\_TAG IS P1 | 9 |
| 3 | FIU ALERT GENERATED FOR LOCATION | 8 |
| 4 | ACTIVITY\_TAG IS P2 & OCCUPATION/NOB LISTED ARE P1 | 7 |
| 5 | % INCREASE IN STR COUNT > 100% | MAX(MAX(8,EXISTING RISK+1),10) |
| 6 | % INCREASE IN CTR COUNT > 150% | MAX(MAX(7,EXISTING RISK+1),10) |
| 7 | % INCREASE IN CTR AMOUNT > 150% | MAX(MAX(7,EXISTING RISK+1),10) |
|  |  |  |
|  |  |  |
|  |  |  |

## Location Risk Flow

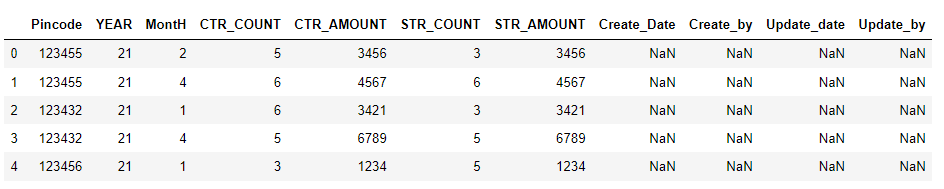
1. First table contains data for STRs raised.



1. Second table contains case related data along with suspicions and FIU alerts raised.

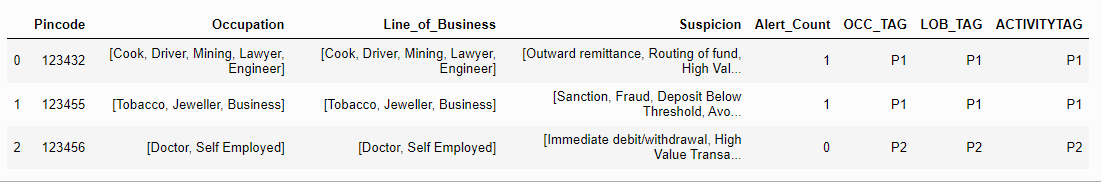


1. Third table contains monthly data for STRs and CTRs.

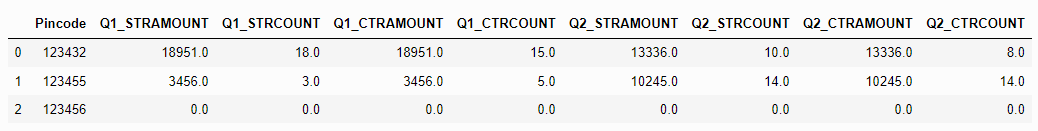


1. Grouping the data based on locations generates the following

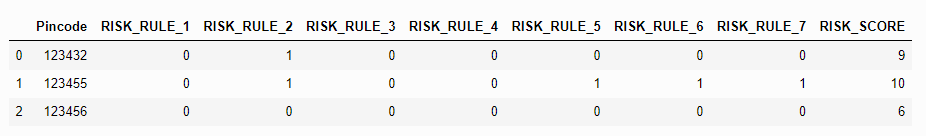
* All the Occupations and line of business reported on that location.
* All the suspicious activities reported on that location.
* Count of alert raised on that location
* OCC\_TAG and LOB\_TAG column contains tag ‘P1’ and ‘P2’ base on high risk occupations provided by the FIU and non high risk occupations respectively.
* ACTIVITYTAG contains tag ‘P1’, ‘P2’ or ‘P3’ based on suspicion raised on that location belongs to P1, P2 or P3 category. `



1. Calculating Quarterly data of CTRs and STRs.



1. Applying all the 7 rules sequentially yields the following result.



# Entity Risk:

Database used for data extraction: FincoreDB and Fincore\_Bridge\_DB. Below embedded Excel has all the Entity Rules listed along with Additional details:

Flow:

**Step 1:** Entity Summary table is to be created based on Risk rules.

Below data will be extracted from Fincore and Fingate database based in each rule.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Entity\_Id** | **Entity Name** | **Entity Type** | **Date\_Of\_Birth** | **PAN** | **Ckyc\_Number** | **Nationality** | **Type\_of\_Customer** |
| 1 | ABC | INDIVIDUAL | 01/01/1990 | DAS2DFFS2 | 123123124124 | INDIAN | SALARIED |
| 2 | XYZ | INDIVIDUAL | 03/02/1980 | ASDF1FASD | 124123124124 | INDIAN | SALARIED |
| 3 | MNO Ltd | ORGANIZATION | NULL | ASFAS1SD1 | 121224335353 | INDIAN | LLP |
| 4 | PRQ CO LTD | ORGANIZATION | NULL | AFSA1AFS1 | 123145353533 | INDIAN | LLC |
| 5 | APPLE | ORGANIZATION | NULL | ASFASWQ1 | 135353453355 | INDIAN | LLC |
| 6 | ORANGE | INDIVIDUAL | 05/06/1976 | AVHDV1JB2 | 343465465753 | INDIAN | SALARIED |
| 7 | PEAR | INDIVIDUAL | 09/09/2000 | ASDFA4654 | 235464786864 | INDIAN | SALARIED |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Client\_On\_Boarding\_date** | **Date\_Of\_Last\_KYC** | **Nbr\_of\_PAN\_Rule\_1.6** | **Nbr\_of\_DIN\_Rule\_1.6** | **Nbr\_of\_Accts\_hrsk\_dmstc\_lctn\_Rule\_2.1.4** | **Nbr\_of\_Addrs\_hrsk\_dmsc\_lctn\_Rule\_2.1.5** | **Peer Group** |
| 02/02/2020 | 02/02/2021 | 1 | NULL | 1 | 3 |  |
| 01/01/2012 | 01/01/2018 | 2 | NULL | 2 | 3 |  |
| 01/04/2018 | 01/04/2019 | 3 | 3 | 3 | 3 |  |
| 01/04/2019 | 01/04/2019 | 3 | 0 | 8 | 9 |  |
| 01/04/2017 | 01/04/2018 | 8 | 5 | 7 | 4 |  |
| 09/09/2001 | 09/09/2017 | 4 | NULL | 3 | 0 |  |
| 09/09/2021 | 09/09/2021 | 5 | NULL | 0 | 0 |  |

**Step 2:** ML Model will predict the peer groups for each entities in summary table

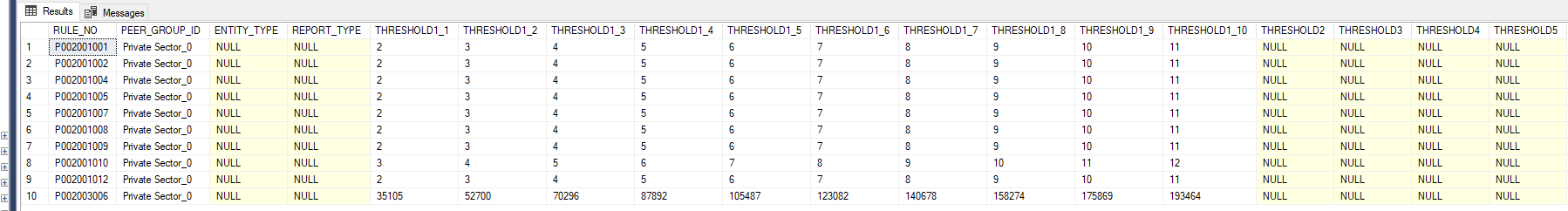
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Entity\_Id** | **Entity Name** | **Entity Type** | **Date\_Of\_Birth** | **PAN** | **Ckyc\_Number** | **Nationality** | **Type\_of\_Customer** |
| 1 | ABC | INDIVIDUAL | 01/01/1990 | DAS2DFFS2 | 123123124124 | INDIAN | SALARIED |
| 2 | XYZ | INDIVIDUAL | 03/02/1980 | ASDF1FASD | 124123124124 | INDIAN | SALARIED |
| 3 | MNO Ltd | ORGANIZATION | NULL | ASFAS1SD1 | 121224335353 | INDIAN | LLP |
| 4 | PRQ CO LTD | ORGANIZATION | NULL | AFSA1AFS1 | 123145353533 | INDIAN | LLC |
| 5 | APPLE | ORGANIZATION | NULL | ASFASWQ1 | 135353453355 | INDIAN | LLC |
| 6 | ORANGE | INDIVIDUAL | 05/06/1976 | AVHDV1JB2 | 343465465753 | INDIAN | SALARIED |
| 7 | PEAR | INDIVIDUAL | 09/09/2000 | ASDFA4654 | 235464786864 | INDIAN | SALARIED |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Client\_On\_Boarding\_date** | **Date\_Of\_Last\_KYC** | **Nbr\_of\_PAN\_Rule\_1.6** | **Nbr\_of\_DIN\_Rule\_1.6** | **Nbr\_of\_Accts\_hrsk\_dmstc\_lctn\_Rule\_2.1.4** | **Nbr\_of\_Addrs\_hrsk\_dmsc\_lctn\_Rule\_2.1.5** | **Peer Group** |
| 02/02/2020 | 02/02/2021 | 1 | NULL | 1 | 3 | P1 |
| 01/01/2012 | 01/01/2018 | 2 | NULL | 2 | 3 | P1 |
| 01/04/2018 | 01/04/2019 | 3 | 3 | 3 | 3 | P4 |
| 01/04/2019 | 01/04/2019 | 3 | 0 | 8 | 9 | P4 |
| 01/04/2017 | 01/04/2018 | 8 | 5 | 7 | 4 | P5 |
| 09/09/2001 | 09/09/2017 | 4 | NULL | 3 | 0 | P1 |
| 09/09/2021 | 09/09/2021 | 5 | NULL | 0 | 0 | P2 |

**Step 2:** Create and update of threshold table for each rule based on peer groups assigned using ML algorithms.

Threshold will be based on statistical boundaries (IQR) of each parameter for a given Peer group.

**Step 3:** Create and update of entity analytical table based on comparison of entity summary values with threshold table based on Rule\_ID and Peer\_Group.



**Step 4:** Calculation of Inherent Risk using Rule weightage table and analytical table.

Rule weightage table will be based on FIU input.

## Risk Calculation:

The risk rules will be categorised into 4 bands – 1, 2, 3 and 4. A band-wise weightage will be configured as below initially. The logic for categorizing rules into bands are as per the risk priority (Earlier, we had given Risk Weightage on scaled of 1 to 10 for each Risk Rule)

|  |  |  |
| --- | --- | --- |
| Band | Band Weightage | Risk Priority (Weightage provided earlier) |
| 1 | 0.5 | 7 to 9 |
| 2 | 0.3 | 5 to 6 |
| 3 | 0.15 | 3 to 4 |
| 4 | 0.05 | 1 to 2 |

Lets assume that the following rules had been hit for an entity.

|  |  |  |  |
| --- | --- | --- | --- |
| **RULE\_DESC** | **RULE\_NO** | **PRIORITY\_NO** | **RISK** |
| Ageing of Reports Filed (STR, CTR, CBWTR, CCR, NTR) <age> years | P003001007 | 1 | 7 |
| Number of mobile Numbers: <value> | P002001008 | 2 | 4 |
| Number of addresses: <value> | P002001012 | 2 | 1 |
| Number of bank accounts : <value> | P002001001 | 2 | 8 |
| Number of Directorships: <value> | P002001015 | 2 | 1 |
| Number of email Ids : <value> | P002001009 | 3 | 6 |
| Number of names/aliases reported across different REs : <value> | P002001007 | 2 | 1 |
| Type of accounts held : <value> | P003001004 | 3 | 5 |
| No of Low Risk NTRs filed : <value> | P002001034 | 4 | 9 |

Weighted average of the square of unit risk score would be taken for each band or Priority no.

The final risk score is calculated as sum(weighted\_average\*band\_weightage)/sum(band\_weightage)

Sample Risk Scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Priority Number** | **Sum of Square of Unit Risk Score** | **Sum of Unit Risk Score** | Weighted average | Band Weightage | Application of Band Weightage |
| 1 | 49 | 7 | 7 | 0.5 | 3.5 |
| 2 | (16+1+64+1+1) = 83 | 15 | 5.5333 | 0.3 | 1.66 |
| 3 | (36+25) = 61 | 11 | 5.5454 | 0.15 | 0.83181 |
| 4 | 9 | 9 | 9 | 0.05 | 0.45 |
| **Final Risk** |  |  |  |  | **6.44** |

# Transaction Risk

Database Used: Fincore\_db and Fincore\_bridge\_db

Flow:

1. FINCORE\_BRIDGE\_DB Transactional Tables table will be used for data extraction.

2. Following columns are extracted to implement transaction risk:

1.Transaction type.

2.Transaction Amount.

3.Transaction Date.

4.Location of Transaction.

5.Transaction Id.

6.Account Id

7.Entity Id

8.BatchId

9.ReportId

10.ReportType

3. Transaction risk will be calculated at Transaction Level but stored at Report Level.

4. We are considering Entities for below Account Relationship Type for an Account

- Account Holder

- Proprietor

- Life Assured

- Beneficial Owner

5. Aggregate the data based on batchid, reportid, AccountId and EntityId to find the transaction volume and transaction value for the report.

6. PEER\_GROUP\_ID would be assigned based on ENTITY\_ID for each Transaction. Thresholds will be calculated for each PEER\_GROUP\_ID and RULE\_NO.

7. Transaction risk will be calculated by using the weighted average of all transaction risk rules.

Formula used for weighted\_average:

weighted\_average = (xi\*wi)/sum(wi)

# Report Risk

1. Part A of Report Risk is calculated as Weighted Average of Individual Risk, Organization Risk, GOS Risk and Transaction Risk
2. Part B of Report Risk is calculated as Weighted Average of Report Risk Rules. There are 29 Risk Rules which are calculated at Report Level
3. PEER\_GROUP\_ID is assigned based on ENTITY\_ID reported for each Account of the Report. Thresholds would be calculated for each PEER\_GROUP\_ID, RULE\_NO and REPORT\_TYPE.
4. Finally, Report Risk is calculated as Weighted Average of Part A and Part B calculated in Step 1 and 2.

# Gos Risk

GOS Risk is the composition of the three Tags:

1.TAG1 (suspicious due to)

2.TAG2 (Source of Alert)

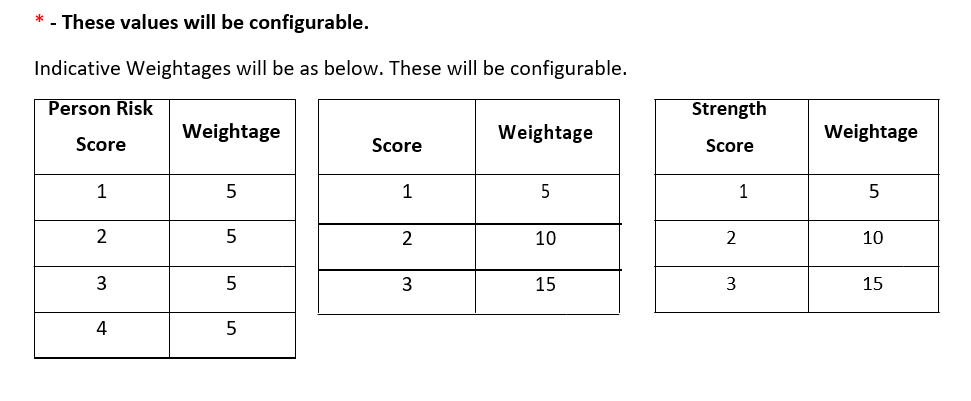
3.TAG4 (Offense suspected)

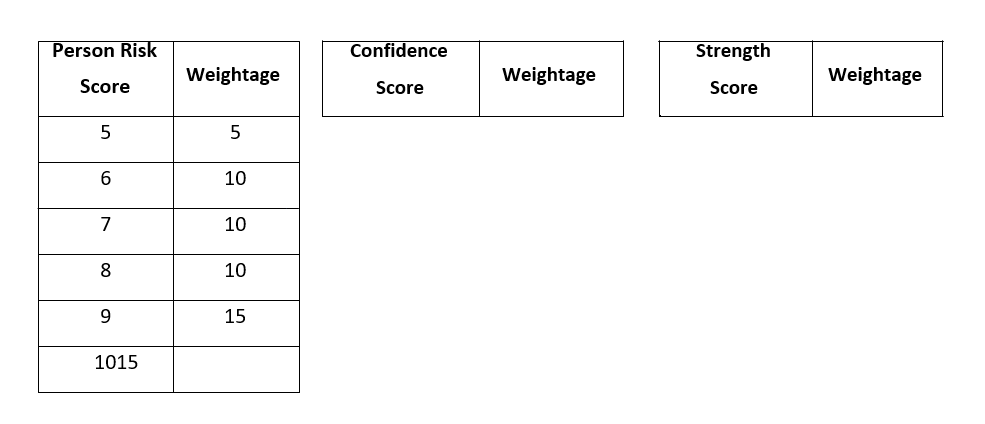
The GOS filled will have their associated TAGs (TAG1, TAG2, TAG3 & TAG4). Every TAG will have their risk score in metadata. The GOS RISK will be calculated on the bases of TAG1, TAG2 and TAG4 as the weighted average of the GOS Risk.

# Network Risk

The network risk for each reported person will be computed as below:

1. The 3 variables for computing are the linked person's risk score, the strength and the confidence of the relation with the linked person
2. Auto High Rules - If the count of persons with risk score of 9**\*** and above is more than 0, then the entire network risk will be high.
3. Auto High Rules - If the count of persons with a risk score of 9**\*** and above is 0, but the percentage of persons with a risk score between 5**\*** and 9**\*** is more than 90%**\*** of the total count of linked persons, then the entire network risk will be high
4. If the above auto-high rules are not met, the following calculation will be applied
5. Weighted average of the Person risk score of all linked persons will be computed = A
6. Weighted average of the confidence score of all linked persons will be computed = B
7. Weighted average of the strength score of all linked persons will be computed = C
8. The product of values A\*B\*C will be computed. This will range between 1 and 90
9. The values in the previous step will be converted into a scale of 1 to 10
10. Each rule should have a parameter to make the rule Active / Inactive



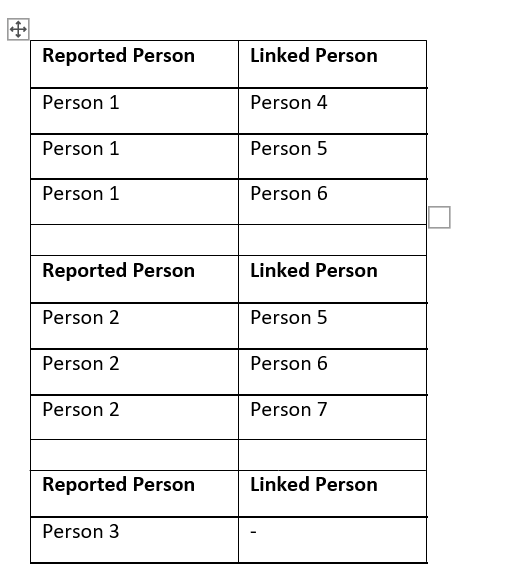


**Sample Scenario:**

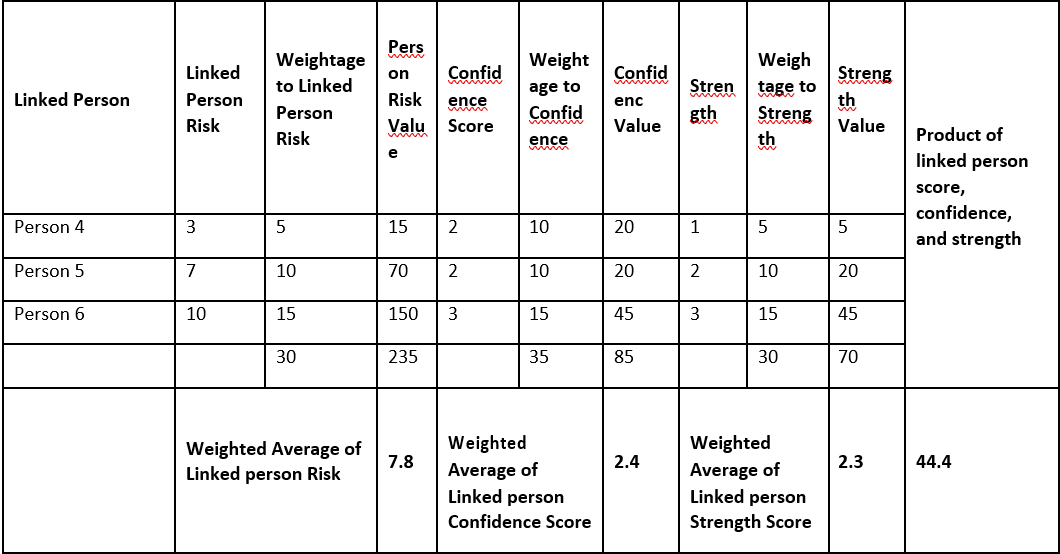
Person 1, Person 2, Person 3 is reported by RE

Person 4, Person 5, Person 6, Person 7 is linked persons

Networks in the case are as below:



Network Risk for the Sample Scenario:



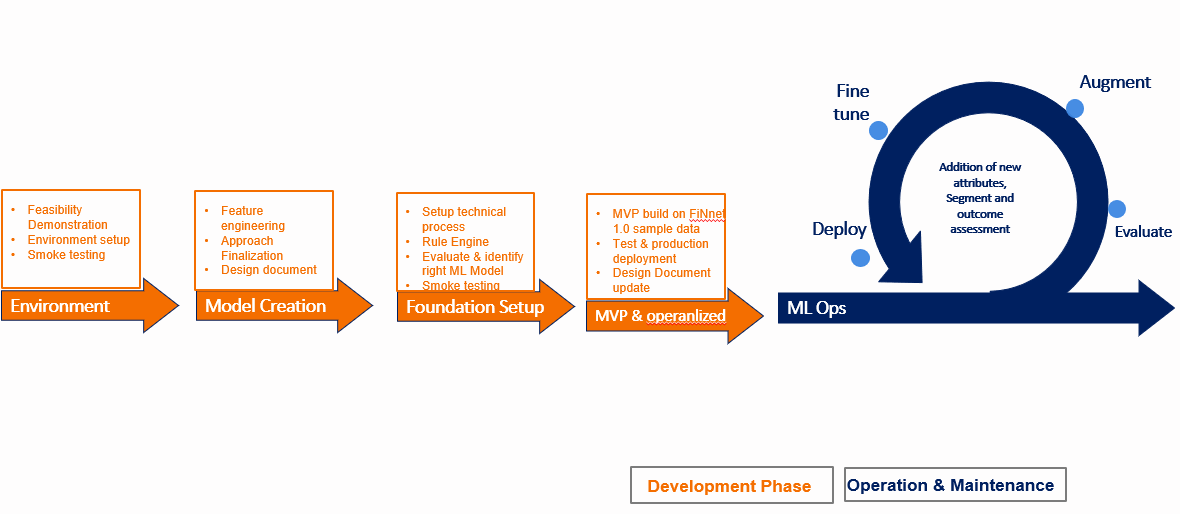
# Case Risk

1. Case Risk is calculated as Weighted Average of Report Risk and Network Risk
2. Based on Case Risk Value, Case is assigned a Priority of P1 or P2.
   1. P2 Priority is assigned for a Case if CASE\_RISK\_SCORE = 0
   2. P1 Priority is assigned for a Case if CASE\_RISK\_SCORE >=1

Assignment of P1 and P2 is configurable and will be changed later once FIU confirms.

# Implementation

The below mention diagram represents the implementation of risk rule-based engine and ML based models. The diagram contains high-level fundamentals which will be a part of this entire exercise from building to deployment.



# Implementation Plan

## Location Risk

Location risk will be the first risk parameter which will be captured to address the geographical risk. Location risk will be done on “City/PIN Code” level as per the business requirement. Static rules along with ML approach will be leveraged to assign risk to any Location.

## Transaction Risk

Steps involved in identifying Transaction Risk Rule are:

* Step 1: Creation of “Transaction Summary Staging Table” which will be all Transactional data for all Reports
* Step 2: Entities will be assigned PEER\_GROUP\_ID. We will leverage the PEER\_GROUP\_ID assigned while doing Entity Risk.
* Step 3: Creation of “Transaction Risk Rule Table” which will have Grouped up data for each Rule for each Report, PEER\_GROUP\_ID.
* Step 4: Creation of “Peer Thresholds Table” based on data profiling of each peer
* Step 5: Calculation of Transaction Risk will be done using “Transaction Risk Rule Table”

## Entity Risk

Entity Risk is to be calculated based on the static Risk rules which is helped by Machine Learning to determine rules thresholds and their weightages.

Steps involved in identifying risk rule are:

* Step 1: Creation of “Entity Summary Table” which will be entity based aggregate table which will be aligned with Risk rules.
* Step 2: Entities will be sub-grouped based on “Profession” (in case of Individual) or “Line of Business” (in case of Legal Entities).
* Step 3: Peer Groups will be created among each sub-group based on unsupervised clustering methodologies.
* Step 4: Creation of “Peer Thresholds Table” based on data profiling of each peer groups individually.
* Step 5: Creation of “Entity Analytical Table” which will store binary score (Y/N) against each risk rule based on comparison between “Entity Summary Table” and ““Peer Thresholds Table”.
* Step 5: Creation of “Rule Weightage Table” which will assign each weightage to each rule.
* Step 6: Calculation of Entity Risk using “Entity Analytical Table” and “Rule Weightage Table”.

# Conclusion

This entire exercise is very complex and extensive which creates multiple ways for business to investigate reported transactions with ease. This exercise will produce meaningful results which will help in reducing the time and energy of the current process.

