

**Financial Crimes Alerts Insight (FCAI)**

Version 1.0.0

**Analytics Overview**

|  |  |  |
| --- | --- | --- |
| Version | Date | Author(s)/Reviewer(s) |
| 1.0 | 6/13/2018 | Christina Ma, Senior Data Scientist |

# About this Document

This document is meant to provide an understanding of the analytic components in the FCAI product. The focus will be on the technical walkthrough of the functions performed in each of the analytic components and how they can be potentially configured to satisfy different customer and/or performance requirement. This document will cover the required input for each analytic and the desired output that can be used to achieve the objectives. In addition, this document will provide necessary verification available for the generation of the analytic results.

# Table of Contents

[1. About this Document 2](#_Toc466273229)

[2. Table of Contents 3](#_Toc466273230)

[3. Overview 3](#_Toc466273231)

3.1 [Data Requirement 3](#_Toc466273232)

3.2 [Analytic Framework 4](#_Toc466273233)

[4. Data Exploration 5](#_Toc466273235)

[5. Feature Engineering…………………………………………………………………………………](#_Toc466273237)

[6. Peer Group Analysis Module…………………………………………………………………….](#_Toc466273237)

[7. Profile Generation Module 11](#_Toc466273237)

**8. Supervised Machine Learning Module………………………………………………………**

**9. Country Risk Module………………………………………………………………………………..**

**10. Risk Score Ensemble Module**

Overview

The IBM FCAI product is designed to provide alert triage and enrichment solution that utilizes cognitive technology to automate the process, accelerate disposition of the alert, and improve decision-making. The solution is intended to be an additional component that fits between the institution’s existing transaction monitoring and case management systems.

## Data Requirements

The default data requirements are specified through the following 14 input files, categorized under 4 main source types, namely, the Transaction Monitoring System, Case Management System, Internal Source, and External Source. The detailed data fields and mapping for each input files are included in the Data\_Requirements\_FCAI.xlsx

* ***TM System***: TXN; Alert\_TXN Mapping
* ***Case Management System***: Alert; (Alert\_TXN Mapping)
  + ***Internal Source***: Individual Customer; Business Customer; Cust\_Account\_Mapping; Account; Address; Address\_Cust\_Mapping; Phone; Phone\_Cust\_Mapping; Email; Email\_Cust\_Mapping
  + ***External Source***: Reference; Country\_Risk\_Table

The main fields for each input files consist of the following information,

* + Field Name
  + Definition
  + Data Type (string; numerical; timestamp; Boolean)
  + Mandatory/Optional (M/O)
  + Category (ID; who; when; what; why)
  + Example
  + Mantas Mapping (to 50+ Mantas tables)
  + Prime Mapping (to 10+ Prime tables)
  + Norkom Mapping (to 8 Norkom tables)
  + List Screening Mapping
  + Comments

While not all data fields are required, each one that is not available may result in some effect in the availability of analytics that can be executed.

## 3.2 Analytic Framework

The FCAI Analytics is built upon layered structure with configurable risk-based ensemble models, as shown in Figure 1. Based on the structure, different analytic flow (catalog) can be build. One example of the analytic flow is shown in Figure 2.



Figure 1 FCAI Layered Analytic Structure



Figure 2 A Typical FCAI Analytic Flow

# Data Exploration

The objectives of Data Exploration module include the following,

* + Provide Data Quality Report for Individual Input Files
  + Triangulate Across Various Input Data Sources
  + Prepare Base Analytic Data Input
  + Generate Corresponding Data Quality Insights

The detailed functions performed in the Data Exploration module are

* 1. Identify and Assign Data Type (Categorical; Numerical; Integer; Logical)
  2. Summarize Data Attributes (number of levels; distribution; counts; NA counts)
  3. Filter Input Variables (invariant; missing values; outlier)
  4. Exclude Incomplete Records
  5. Check Common Rules
  6. Determine Default Value
  7. Extract Additional Fields from Text Analytics
  8. Generate Charts for Better Understanding of Data Characteristics
  9. Pre-Process Key Components
     + Date
     + Name
     + Address

# Feature Engineering

The objectives of Feature Engineering module is to evaluate potential quantitative/qualitative attributes and generate derived measurements/statistics that can then be used as input for various analytics components. There are 4 main components related to Feature Engineering module.

* Base Feature: attributes/characteristic from input files
* Measurement/Statistics: variables that provide relevant information that can be used to evaluate potential risks
* Aggregator: focus of variable which defines the aggregation key for derived statistics. Examples of aggregator are customer, account, country, business type, tender type, originator/beneficiary, etc.
* Flag: Rules and (Pre)conditions based on unstructured data

The FCAI product provides a flexible way to configure additional features with several predefined templates.

# Peer Group Module

The purpose of peer group is to create homogeneous clusters such that objects within the same clusters share certain similar behavior and/or activity patterns. This information would help in separating population data in the analytic to be performed later. Examples of peer group may include

* customer type
* industry codes
* income/revenue
* transaction date/frequency/types/volume/amounts
* account type
* product type
* geographical information
* channels
* customer risk ratings

# Profile Generation Module

The objective of this module is to generate risk assessment based on individual features through selected peer groups such that right thresholds are being applied. Figure 3 illustrates the effect of the peer group on the risk assignment.

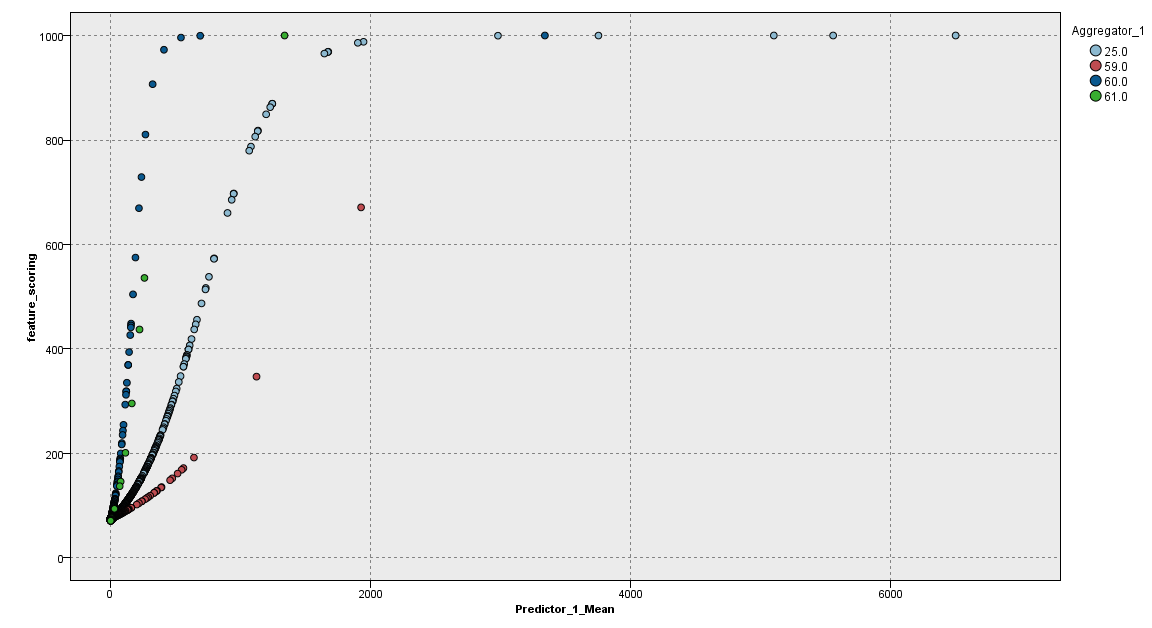


Figure 3. Profile Risk Score based on Different Peer Group

The detailed scoring process is demonstrated through the following tables.

1. Obtain quantitative feature (derived measurement/statistics) for the specified aggregator (e.g. Customer)



1. Calculate General Interestingness Index (GII) which provides importance weights for each feature, based on the distribution of each feature



1. Calculate Peer Group Score



1. Calculate Profile Subscore



# 8.Supervised Machine Learning Module

The FCAI Supervised Machine Learning module is applied to utilize modern cognitive machine learning capability to predict risk level for specified aggregator based on similar historical behavior and disposition. Three major steps are performed,

1. Prepare Input Data
   * Obtain/create features for the selected aggregator
   * Generate target variable (disposition) for the selected aggregator
2. Generate Scoring Model
   * Split input data into training and testing sets
   * Generate model using training data set
3. Test Model Prediction
   * Predict disposition for the testing data set
   * Evaluate performance for the testing data set

Current benchmark results based on Prime historical data has shown promising accuracy. Two settings have been performed using default threshold of 0.5 for identifying SAR and nonSAR,

1. Scenario #1 : Random Forest with Undersampling



This result shows the model identify 85% of the SAR with 12% of non-productive alert by reviewing top 12.9% of the cases.

1. Scenario #2 : Random Forest with Undersampling and Peer Group Clustering



This result shows the model identify 90% of the SAR with 12% of non-productive alert by reviewing top 12.8% of the cases.

# 9.Country Risk Module

The determination of country specific risk level is depending upon the external country risk file which specifies the risk score for each country. The FCAI product would score each transactions based on the customer country as well as counterparty country using the risk score from the country risk file. The detailed scoring for country risk module is illustrated in the following tables.

1. Find customer country and counterparty country



1. Match country risk file



1. Calculate country risk subscore



# 10.Overall Risk Score Ensemble Module

The overall risk score is a function of various subscores. The default ensemble module uses weighted average based on the results from different analytic components. It can be adjusted based on feedback from analyst. Other available methods for ensemble include majority vote and meta-model. The majority vote will use the risk assessment base on the agreement from the majority of the analytic components. The meta-model will enable additional model training using information from subscores.

For each alert, a set of alerted transactions are extracted. Each transaction would be assigned a risk score based on the corresponding aggregator (e.g. the account), and the transaction countries involved. The overall risk score is then calculated across all the alerted transactions.