Tookitaki

Anti-Money Laundering Suite Name Screening

User Guide

www.tookitaki.ai

Legal Statement

The content of this document is proprietary information owned solely by Tookitaki Holding Pte. Ltd. The information is only for the intended recipients whom are properly authorized employees or consultants to use Tookitaki products. No part of its content may be used for any other purpose, disclosed to any person or firm or reproduced by any means, electronic or mechanical, without the express prior written permission of Tookitaki. The text, graphics and examples included herein are for the purpose of illustration and reference only. The specifications on which they are based are subject to change without notice. No legal or accounting advice is provided hereunder, and any discussion of regulatory compliance is purely illustrative.

The product described in this document is furnished under a license or access is provided as a service. The product may be used only in accordance with the terms of that license contract. Certain features and functionality described in this document may be features not included in a base product license and may only be accessed and used by licensees expressly licensed to use such features and functionality. Licensees should refer to their respective agreements to determine whether they have the right to use a particular feature or functionality.

Corporate and individual names and data used in examples herein are fictitious unless otherwise noted. Information in this document is subject to change without notice. Tookitaki reserves the right to revise or withdraw this document or any part thereof, including, without limitation, the elimination or modification of any product functionality, at any time.

Tookitaki, the Tookitaki logo are registered trademarks. All other marks are trademarks of their respective owners. Other company, brand, product service names are trademarks or registered trademarks of their respective holders.

Document Number: AMLS-NS-UG-B.1
May 6, 2020

About

Tookitaki is providing enterprise software solutions that create sustainable compliance programs for the financial services industry. We are innovating the regulatory compliance space by moving beyond rules-based applications and introducing software solutions to maximize efficiency and reduce risks in the compliance processes. Our mission is to provide machine learning-powered regulatory compliance solutions that are auditable, scalable and actionable. Incorporated in November 2014 in Singapore, the company is currently led by a core team with cumulative 150-years' experience in finance, AI, Big Data Analytics and financial crime. We are backed by institutional investors Jungle Ventures, Enterprise Singapore (Singapore Government) and Illuminate Financial and have a presence in the US, Singapore and India.

The uniqueness and robustness of our innovation in the field of machine learning have been acknowledged locally and worldwide. We received accreditation@SGD in 2017; a rare-to-achieve recognition initiated by the Singapore government for SMEs to get green channel access to government projects and partner with local banks in their key initiatives. In addition, we were selected as a Technology Pioneer (2019 cohort) by the World Economic Forum, recognizing our ability to shape the industry and the region in new and exciting ways. Also, in 2016, Tookitaki won the first prize in the MAS award (SME category) for our innovation and proving our platform through a live deployment. Our solutions have been part of a number of accelerators such as the ING Fintech Village (2019 cohort), Queen City Fintech Incubator (class 7), Catalyst: SocGen Startup Accelerator (2016), Voyager: Nomura Fintech Accelerator (2017) and HSBC-T-Hub Accelerator 2030 (2018).

In the regulatory compliance space, we currently focus on anti-money laundering and reconciliation management. Our two products Anti-Money Laundering Suite (AMLS) and Reconciliation Suite (RS) cater to anti-money laundering and reconciliation, respectively. Both these solutions use cutting-edge artificial intelligence/machine learning (AI/ML) algorithms and are built on top of our proprietary Tookitaki Decision Support System (TDSS). It uses cutting-edge cluster computing infrastructure and distributed data parallel architecture design and provides scalable AI/ML infrastructure for building enterprise-grade software suites. It facilitates the creation of end-to-end AI/ML workflows that are auditable and self-evolving. End-to-end security across the stack, easier and extensible integration options, and well-defined REST interfaces make TDSS an enterprise-ready machine learning platform for building intelligent and sustainable compliance programs.

Tookitaki AMLS is an end-to-end transaction monitoring and screening solution, which improves operational efficiency, mitigates the risk of money laundering and reduces the cost of compliance. Rapid growth in wire activity is making it difficult to monitor and detect suspicious transactions and names effectively and efficiently, while regulators are enforcing a higher level of scrutiny in the money laundering space. Today, banks mostly rely on manual efforts for their anti-money laundering compliance but adding more human resources is not a viable long-term solution. Rules-based applications that produce 95% false alerts (system alerts that are found to be non-risk cases after investigation), creating huge backlogs and massive ageing of alerts, cannot be sustainable either. With legacy systems, the costs of managing the process to read, analyze, and implement the changes in operations will only increase over time. It is noteworthy that banks currently handle around 150 anti-money laundering alerts on a daily basis. This is expected to surge due to the growing complexity of compliance regulations. More than 300 million pages of regulatory documents will be published by 2020.

Our AMLS solution came into existence following a lengthy and comprehensive study about the anti-money laundering compliance workflows within the industry. During our study, we have identified certain systemic issues within existing rules-based workflows that lead to a significantly large number of false alerts, missed true suspicious cases and limited auditability and scalability of workflows. The major challenges with current anti-money laundering systems are: 1) criminal organizations employee professional money laundering cells that do not operate within the confines of static, predefined, overly-broad transactional actions, 2) unidimensional detection logic (grouping based on transactional behavior) fail to successfully identify money laundering operations, 3) no robust explanation available for alerts and inability to scale with new policies.

Tookitaki AMLS marks a paradigm shift from existing systems and is based on three new revolutionary concepts: 1) new 'suspicious' case detection with advanced machine learning algorithms, 2) smart typology repository to update detection scenarios automatically and 3) intelligent triaging of alerts for faster and efficient alerts disposition. With AMLS, banks can drastically improve overall functioning of their AML/CFT programs by increasing effectiveness and efficiency.

AMLS is built with a design philosophy of providing maximum detection coverage and high alerts yield (low false alerts), while being fully transparent and scalable. Instead of being a black box (a major reason why machine learning solutions are not accepted by regulators), the solution explains the activities of machine learning models through dependency rules. This provides AML investigators with tools to explain the complex models in business terms.

During a pilot project with a major bank, the solution could reduce false alerts by 40% for transaction monitoring with a 5% uplift in STRs. For names and sanctions screening, the solution could reduce false alerts by 50-60% across individuals and corporates. Apart from the mentioned quantifiable benefits, AMLS is beneficial in a number of ways. It provides detailed explanations for every alert to ensure better regulatory compliance. The solution also gives actionable analytics for an integrated view across key components of transaction monitoring and name screening. In addition, from a system interoperability perspective, AMLS complements and integrates seamlessly with existing anti-money laundering systems for faster time-to-market.

Reconciliation management is Tookitaki's another area of expertise within regulatory compliance. Today, reconciliation processes in FIs have become more challenging due to a combination of factors like increasing process complexity, high transaction volumes, a large number of data sources, ever-changing regulatory requirements and the emergence of new asset types and structured deals. Regulations such as Basel III and MiFID II mandate increased depth and breadth in the reconciliation process. Manual processing and rules-based solutions are poised to become costly and prone to numerous errors in the long run. The problems with current reconciliation processes across the globe (rules-based or robotic process automation (RPA) solutions) are: 1) need for granular predefined rules or exact steps to reconcile transactions, 2) inability to reconcile complex transactions, resulting in exceptions or breaks, 3) exceptions form 10-15% of total transaction data but take 70-80% of an analyst's time in investigation and resolution, 4) exceptions handling and complex matching require human judgement beyond a basic set of rules, making the process costly and time & resource-intensive, and 5) some exceptions remain unresolved due to their complex nature or insufficient information causing aged breaks. This increases the risk of regulatory non-compliance. Most solutions in the market support simple matching but they provide minimal support on complex matching and exceptions handling.

In order to address this, Tookitaki has introduced its RS, which is an end-to-end AI/ML-powered matching and substantiation solution, which improves quality and efficiency in the investigation process and reduces the cost of reconciliation. RS comprises of two modules: 1) RS-Matching to solve complex matching cases and false breaks and 2) RS-Substantiation to solve 'true' breaks or exceptions. Our RS solution is highly efficient and offers automated, accurate matching and exceptions handling, along with adjustment amount recommendation and audit trails. The solution has a proven ability to offer 95% accuracy in break resolution and a 50% reduction in investigation time with dramatic improvement in match rates.

RS is built with a design philosophy of providing robust coverage and scale in automated break resolution across various reconciliation areas in the banking and financial services industry. It complements and seamlessly integrates with existing reconciliation systems. It can also independently work as a full-fledged competent solution with guaranteed improvement in operational efficiency across its modules. The matching module uses AI techniques (primarily game theory algorithms) to handle processing errors and system limitations to produce matches and automatic reconciliation. The substantiation module uses ML techniques (primarily multi-classifier algorithms) to accurately detect break type and recommend adjustment amount required to reconcile. In addition, it is able to detect outliers/anomalies and present an explanatory trail, helping in faster detection of 'new' break types. Apart from break type detection, the module can recommend the difference in transactions and support quick, accurate reconciliation. Such automation can save lot of manual calculation and ensures faster exceptions handling. It also provides a thorough audit trail that can explain cause of break in detail, helping banks with actionable steps. Besides, it provides explanation across model outcomes. NOSTRO-VOSTRO, FX, Stocks, Bonds, Derivatives, Front Office-Back Office, GL, Cash, Due From-Due To are some of the reconciliation areas that we currently address.

Contents

List Of Figures

List Of Tables

About This Guide

This document provides information about Tookitaki Anti-Money Laundering Suite (AMLS) Name Screening (NS) Module. It is made up of the Analytics and Investigative work flows consisting of the Tookitaki Data Science Studio (TDSS) and AMLS.

Detailed descriptions and procedures of the user interface and configuration are provided to assist the user understand AMLS NS to its optimum capabilities.

Who Should Read This Guide

This user guide is designed for authorized personnels such as Consultants, Analysts, Data Scientists.

Related Documents

- Anti-Money Laundering Suite – Installation Guide
- Anti-Money Laundering Suite – Configuration Guide
- Anti-Money Laundering Suite – Data Ingestion Guide
- Anti-Money Laundering Suite Transaction Monitoring - User Guide.
- Anti-Money Laundering Suite – Administrator Guide
- Anti-Money Laundering Suite – Release Notes
- Anti-Money Laundering Suite – Technical Notes

Conventions

The following conventions are used in this document:

Typographic Conventions

Appearance	Descriptions
Warning:	Information requiring very special attention
Important:	Information requiring special attention
Caution:	Caution
Notes:	Notes
`file.extension`	File Names
`directory`	Directory Names
`command`	Command to be typed
word	Emphasized word
`paragraph`	Code Example
bold	Navigation instructions. Example: Click **File** > **New** > **Project**

Symbolic Conventions

Appearance	Descriptions
>	Indicates menu item selections in a graphic user interface. e.g. File > New > Project
Footnote[1], Tablefootnote[2]	Numbers indicating reference to page or table footnotes.
🧭	Navigational tips on how to access the pages shown.

For Further Help

You can find information on how to contact your Tookitaki representative by clicking **Contact Us** at the Tookitaki web site, www.tookitaki.ai.

Technical Support

Telephone: +65 6250 2620

Email: support@tookitaki.com

If You Find an Error

Tookitaki makes every effort to prevent errors in its documentation. However, if you discover any errors or inaccuracies in this document, please inform your Tookitaki representative. Please quote the document number found at the bottom of the legal notice on the inside front cover.

Overview

Anti-Money Laundering Suite

Tookitaki Anti-Money Laundering Suite (AMLS) is an end-to-end machine learning-powered Name Screening solution for financial institutions. It improves operational efficiency, mitigates the risk or money laundering and reduces cost of compliance. AMLS comprises of two units:

- AMLS Analytics Module
- AMLS Name Screening (NS) Application Module

AMLS Name Screening

Name screening is the process of customer names in a financial institution are compared with a Watchlist which contains names of people previously known to have involved in Money Laundering, Terrorist activities, Human Trafficking, politically known people, high officials of a country and other suspicious activities to know if the customer in the financial institution is related to or actually the person in the Watchlist.

The current name screening system in financial institutions is using predefined rules-based process to raise hits if there is a match between customer name and Watchlist. This process is only considering certain conditions for matching without considering any other entities. Hence, many false alerts are raised.

AMLS NS overcomes the difficulties of these false alerts and provides a secondary screening solution which reduces the false alerts by about 40-50 percent with less mis-classification rate. The most important is determining if an alert is true or false.

AMLS NS uses the alerts data after investigation, the Watchlist data, primary screening system output, first name synonyms and last name synonyms. From the features present in the respective data AMLS NS develops new primary and secondary features based on the importance. These features include:

- Primary Features: Age Difference, Place of Birth, Nationality.
- Secondary Features: Linked Names, Occupation, Nationality.

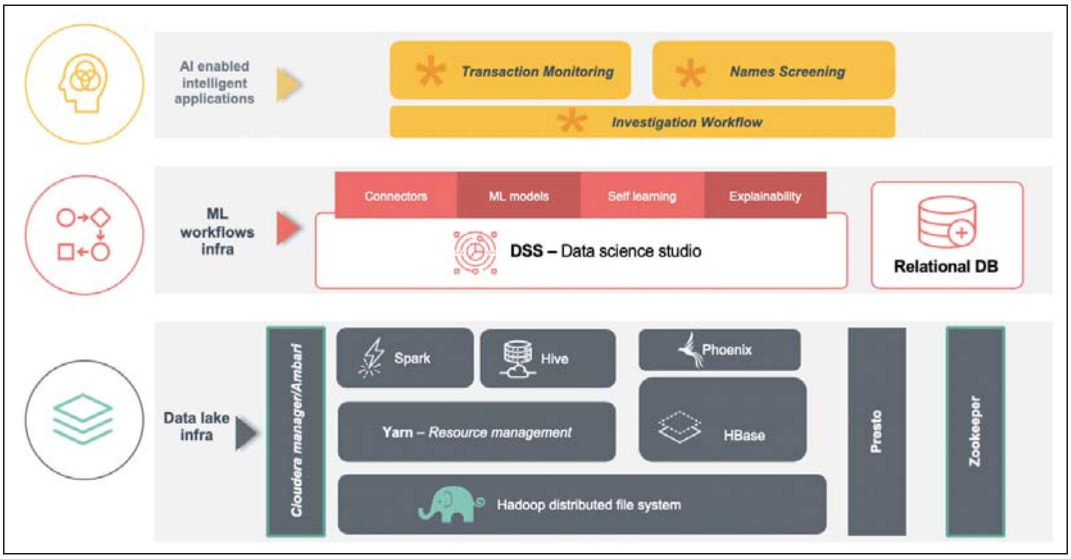
With these required features, a model is developed which gives a prediction whether an alert is true, false or there is insufficient data.

Architecture

Technology Stack

The foundation of the AMLS technology stack is the Hadoop Big Data cluster. The central layer is the Data Science Studio (DSS) platform, which handles analytics using machine learning. The top layer is the application layer containing AMLS-specific business logic and UI of Transaction Monitoring (TM) or Names Screening (NS).

Figure 1 - Technology Stack



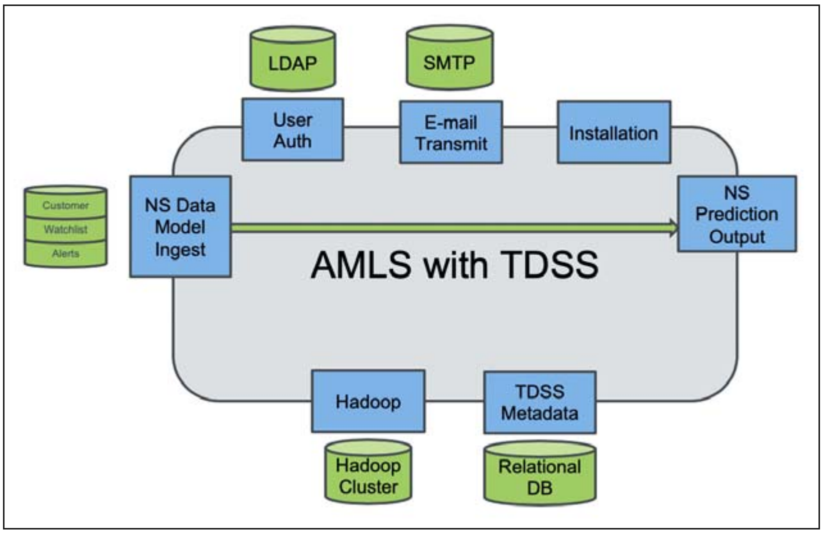
The Technology Stack details are listed in the table below.

Table 1 - Technology Stack

Component	**Details**
HDFS	Provide distributed storage capabilities.
YARN	Provide scheduling and resource management capabilities.
Spark	Provide data parallel distributed compute capabilities across the cluster.
Hive	Provide SQL interface for data stored on HDFS.
HBase	Provide low latency sub second querying capabilities on top of HDFS.
Apache Phoenix	Provide SQL interface on top of HBase NoSQL store.
Zookeeper	Provide coordination among various Hadoop services. (dependency by HBase).
Cloudera Manager	Provide easier configuration and infra management for all Hadoop services.
Relational DB	Store application metadata from DSS, TM and NS.
Data Science Studio (DSS)	Provide end to end Machine Learning infrastructure services for application running on top.
Name Screening	Provide NS services and Web UI capabilities.

Integration with Client Environment

Figure 2 - Integration with Client Environment

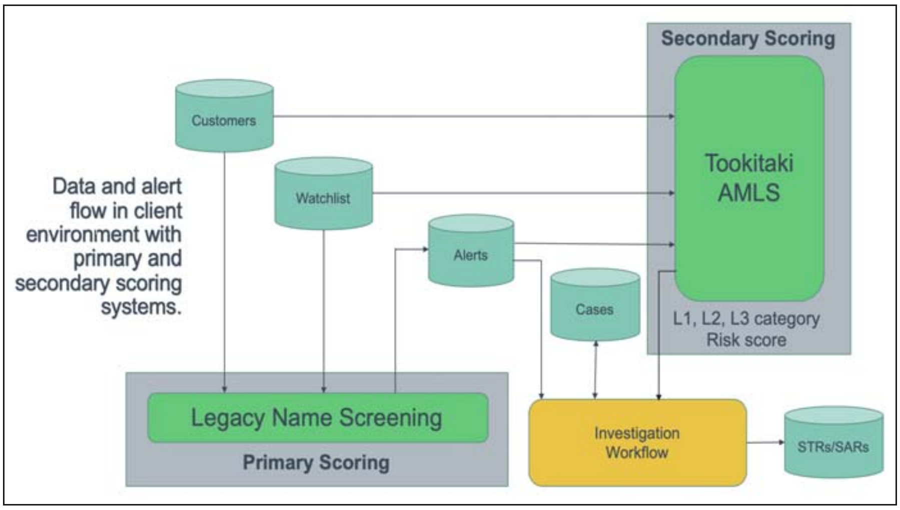


AMLS components integrate with the client components (in green). Key points of integration are:

- Ingesting name screening data:
  - Customers
  - Watchlists
  - Alerts
- Connection to corporate Active Directory for user authentication via LDAP.
- Connection to corporate e-mail system for communicating via e-mail.
- Connection to Hadoop cluster.
- Connection to SQL database where application data is stored.
- Integration with name screening prediction output (risk ranked name screening alerts).
- Integration with name screening output (ranked name screening alerts).

Data & Alert Flow and Application Integration

Figure 3 - Data & Alert Flow and Application Integration

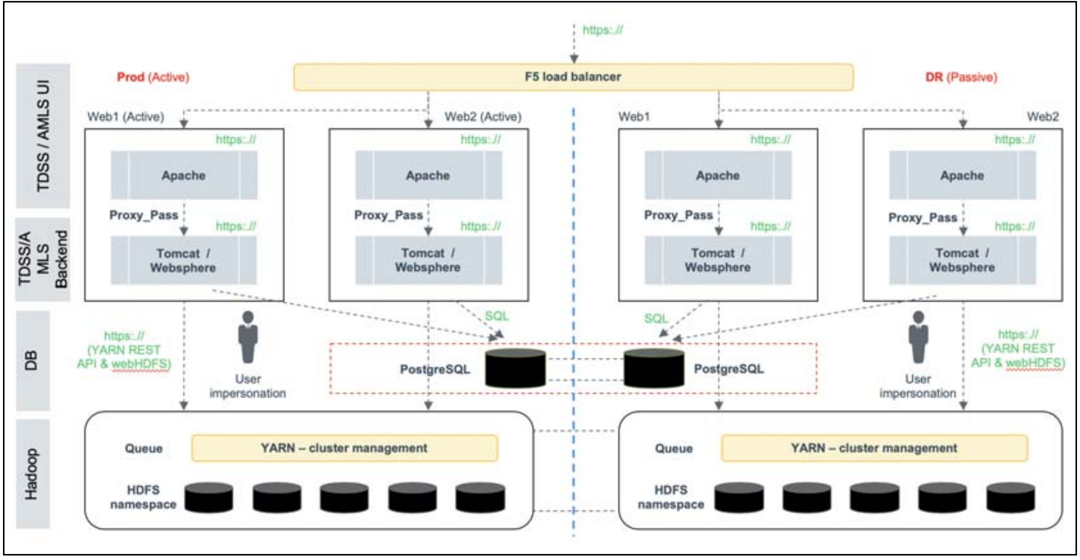


The data consisting of the customers and watchlists flows through the primary scoring in the legacy name screening system. This process generates the alerts.

These alerts as well as the data, then flow through the secondary scoring process in AMLS which generates their risk scores and prioritizes them into three categories, L1, L2 and L3, with L3 being the highest risk level.

Deployment Architecture

Figure 4 - Deployment Architecture



- User web requests entry through the load balancer, which decides which Apache Httpd server will service the request.
- The Apache Httpd server typically hands the request to the Apache Tomcat server, which integrates with the AMLS and TDSS Java application.
- The TDSS application accesses the database server to fetch/store meta data.
- The TDSS application accesses the Hadoop cluster (any number of worker nodes) to store data (Big Data) and perform large scale tasks

- Https protocol runs over various tcp ports, not just tcp/443.
- There must be no firewall between these servers (application server, database server, Hadoop servers).
- For Disaster Recovery (DR), the load balancer switches from the active left cluster to the passive right side.

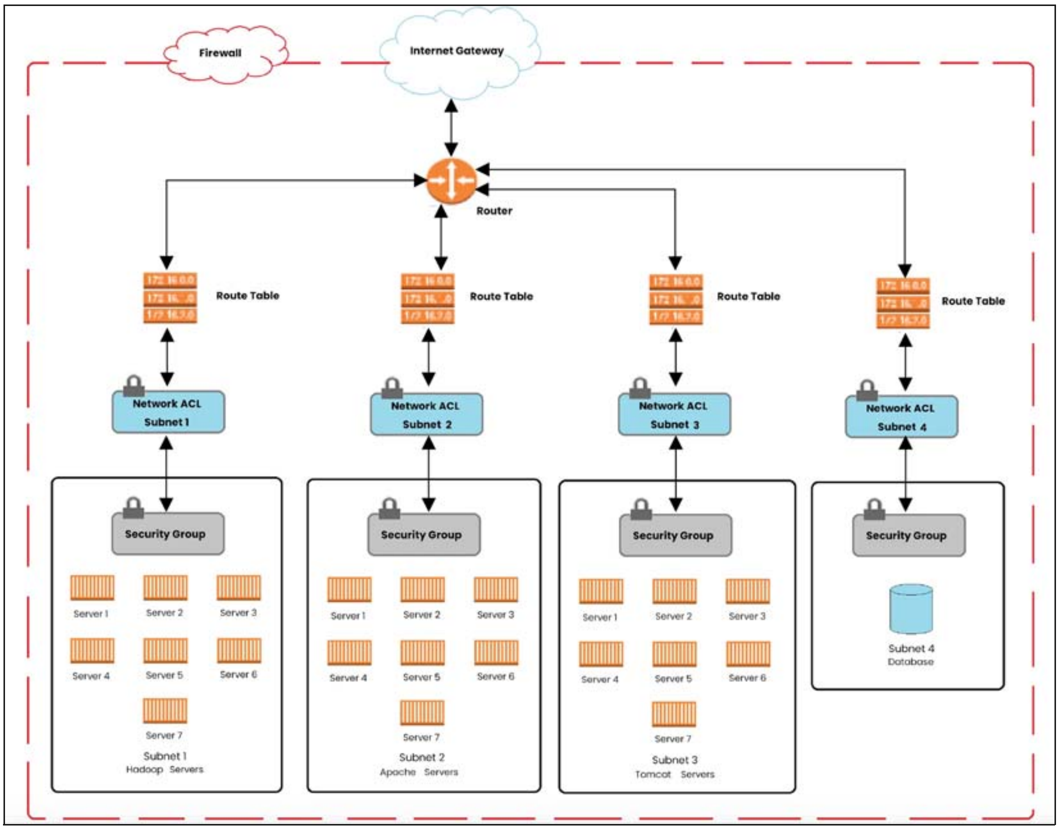
Security

- Communication from client to load balancer and web servers is done using SSL/TLS.
- Connection to database is also established using SSL/TLS.
- Data at REST in database is achieved by encrypting the file system where data files, access logs and audits are stored.
- Authentication to database is done using Active Directory (AD) account principals and credentials.
- Data at REST on HDFS is achieved by setting up encryption zone. Data is encrypted on client/driver side, so data in transit and data at REST are encrypted.
- Authentication at Hadoop layer is done using Kerberos and propagating the delegation tokens thereof.
- Key management to be performed using Hadoop KMS.

Network

The application servers and database are in the same network. Requests from the user's browser need to pass through the Firewall to communicate with the database and application servers. Specific ports for the servers are to be open for communication. For internal communication among the servers, all ports are to be open to pass the Firewall. Refer to “Server Ports” below.

Figure 5 - Network



Server Ports

The default WebUI ports for these components must be open for communication.

Table 2 - Server Ports

Component	Service	Default Port	Configuration
Hadoop	DataNode	50010	dfs.datanode.address
		1004	dfs.datanode.address
		50075	dfs.datanode.http.address
		50475	dfs.datanode.https.address
		1006	dfs.datanode.http.address
		50020	dfs.datanode.ipc.address
	NameNode	8020	fs.default.name or fs.defaultFS
		8022	dfs.namenode.servicepc-address
		50070	dfs.http.address
		50470	dfs.https.address
Hadoop YARN	ResourceManager	8032	yarn.resourcemanager.address
		8033	yarn.resourcemanager.admin.address
		8050	yarn.resourcemanager.webapp.address
	NodeManager	8042	yarn.nodemanager.webapp.address
		8044	yarn.nodemanager.webapp.https.address
	JobHistory Server	19888	mapreduce.jobhistory.webapp.address
		19890	mapreduce.jobhistory.webapp.https.address
Hive	Metastore	9083	hive metastore port
	HiveServer2	10000	hive. server2. thrift.port
	HiveServer2 Web UI	10002	hive. server2. webui.port in hive-site.xml
	WebHCat Server	50111	templeton.port
Tomcat		8080 / variable	User defined.
Apache		80/443	http/https
Database	MySQL	3306	TCP
	PostgreSQL	5432	TCP
	MariaDB	3306	TCP
TDSS	-	7080	UI port
AMLS	-	7090	UI port

AMLS Analytics Unit

AMLS Analytics module also called Tookitaki Data Science Studio (TDSS) is the analytics engine inside AMLS that allows advanced analytics professionals to work with data and setup Tookitaki AI based analytics work flow for Name Screening (NS) use case. It drives the end-to-end machine learning-based work flows that are interpretable, self-evolving and easily integrated.

The AMLS TDSS engine provides the following functionalities and tasks that can be done for the NS use case:

- NS Alert Prioritization (Supervised Pipeline)
  - Connectors for the database
  - Training & Validation
  - View Executions
  - View Model Performance
  - View Model Features Contribution
  - Model Evolution
- Prediction Pipeline for NS Alert Prioritization
  - Predict New Data
  - Feature Contribution and Prediction Explainability
- Pipeline Execution
  - Pipeline Setup and Execution Steps

Name Screening Application Module

Alerts Investigation View

The Alerts Investigation View lists all alerts extracted from the existing NS system with the secondary scoring results. The alerts are prioritized into bucket levels of L1, L2 and L3. It also enables the user to navigate to the alert details page for further investigation purposes.

Refer to “Alerts Investigation View” on page 52.

Risk View

The Risk view provides detailed insights into the statistics of alerts, hits and break down of alerts by segments. It displays values for both the current period and the previous period as defined by the selected preset date or the custom date range.

The statistics are displayed on widgets categorized as:

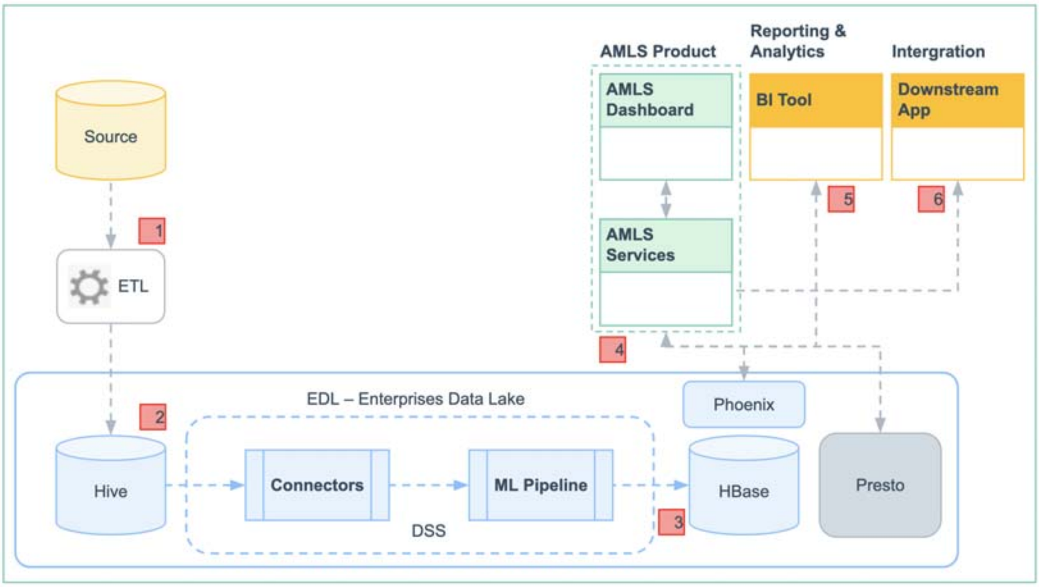
- Alerts & Hits
- Hits Distribution
- Business Segments & Parameters
- Customer View for All Segments

Refer to “Risk View” on page 73

Name Screening Application Flow

Integration Flow

Figure 6 - Integration Flow



1 Initial ETL to be performed by the clients to feed the masked data to the product.

2 DSS to use Hive/HDFS connector to pull in the required data as appropriate for the pipeline.

3 ML pipeline would generate the final predictions and other model metadata to be stored in HBase for low latency querying.

4 Application services to consume the required data from HBase through sql like interface provided by Phoenix.

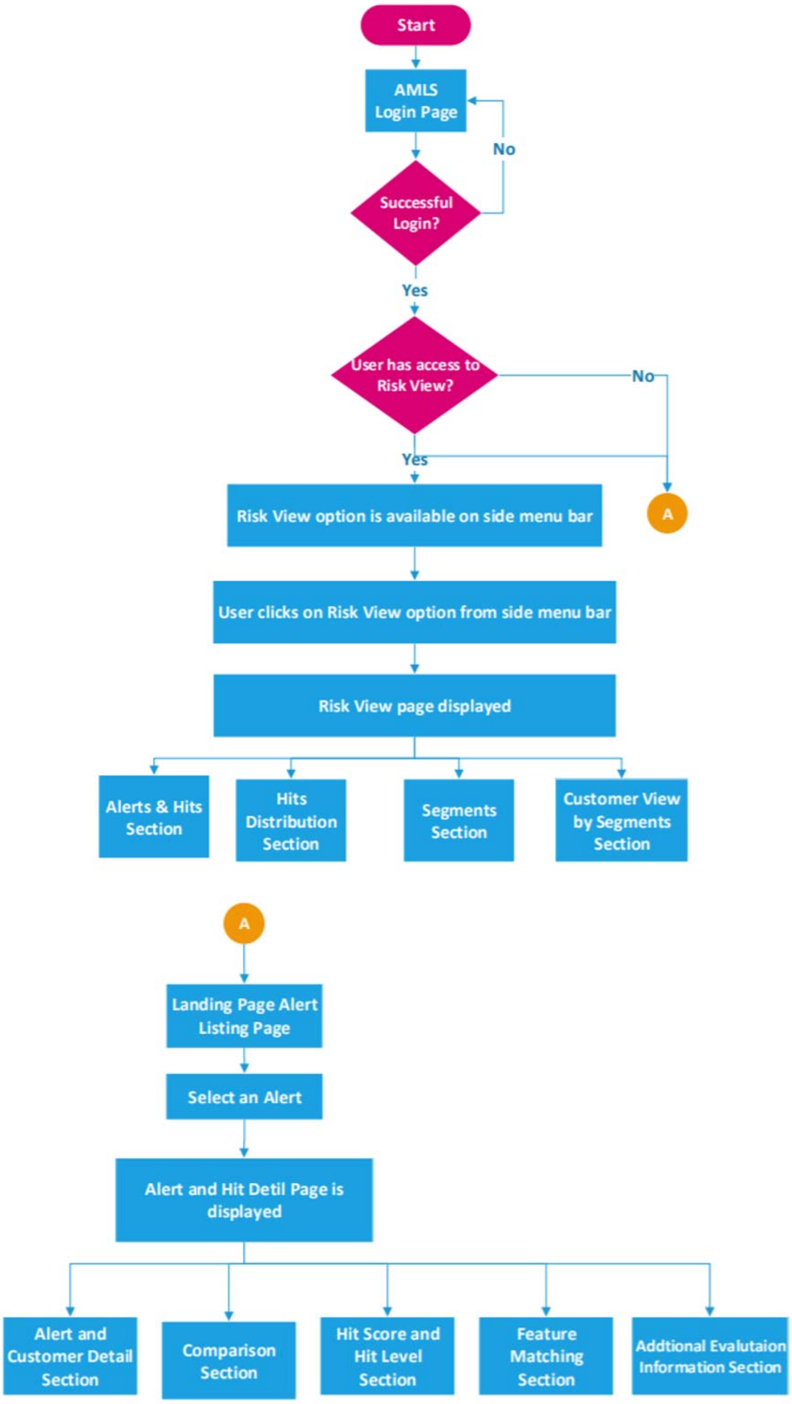
5 Existing BI Tools can directly fetch the required data from HBase and HDFS through Apache Phoenix (SQL interface) and PrestoDB.

6 Other downstream applications can use the rest services directly as appropriate (for application data) or can connect through Phoenix to acquire the raw data.

Application Process Flow

The end-to-end work flow is represented below, which shows how AMLS and existing rules systems for NS co-exist and used by analysts/users.

Figure 7 - AMLS NS Work Flow


Start
AMLS Login Page
No
Successful Login?
Yes
User has access to Risk View?
No
Yes
Risk View option is available on side menu bar
A
User clicks on Risk View option from side menu bar
Risk View page displayed
Alerts & Hits Section
Hits Distribution Section
Segments Section
Customer View by Segments Section
A
Landing Page Alert Listing Page
Select an Alert
Alert and Hit Detil Page is displayed
Alert and Customer Detail Section
Comparison Section
Hit Score and Hit Level Section
Feature Matching Section
Addtional Evalutaion Information Section

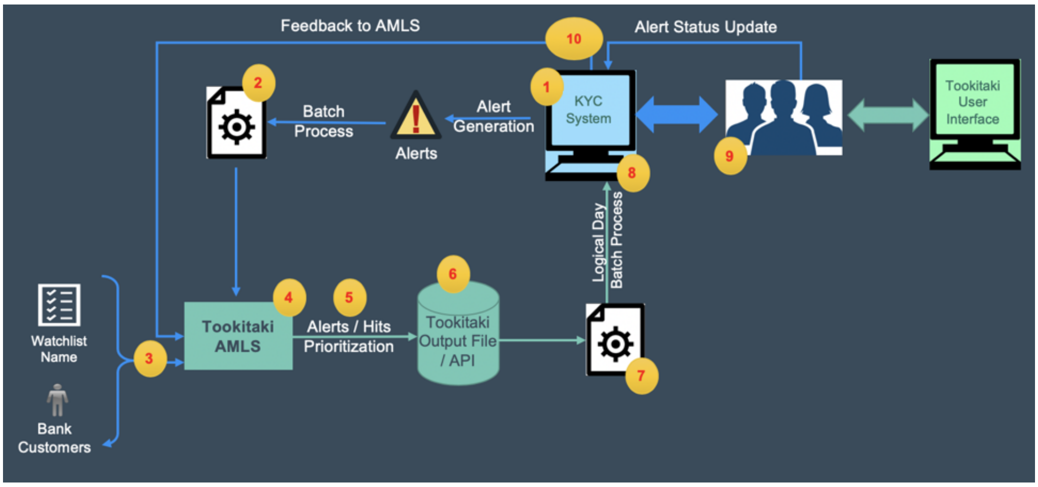
Name Screening Process Flow

The NS module performs the following primary function:

- Scoring of current name screening system alerts to prioritize the alerts based on their level of risk. The system generates both a risk score and categorizes the alert into 3 categories L1, L2, L3

The process flow for NS module is as shown in the figure below.

Figure 8 - NS Process Flow



1 The Bank's existing AML KYC system generates Name Screening alerts and hits and sent to AMLS for prioritization.

2 The Bank's designated file transfer process provides the daily alerts and hits data for AMLS application ingestion and prioritization.

3 Daily changes on customer and associated information, changes to related customer and account information, changes to Watchlist One data and complete feed of Bank internal watch list are ingest into AMLS via a separate Tookitaki ETL process for AMLS application ingestion & prioritization.

4 The AMLS application executes the supervised prediction pipeline to prioritise the alerts and /hits.

5 The alerts and /hits are prioritized into L1, L2 & L3 buckets with alerts score and corresponding explainability link. This will be a logical day batch process and typically would be available the next logical day for analyst review.

6 AMLS analytics application generates the prediction file in delimited format and the Banks's designated file transfer process ingests into the KYC system.

7 Same as 6.

8 The alerts and hits bucket, score and explainability link is available in KYC system, to supplement the alerts and hits investigation process.

9 Same as 8.

10 The updated outcome of the investigations from the KYC system is then sent back to Tookitaki AMLS application for model learning.

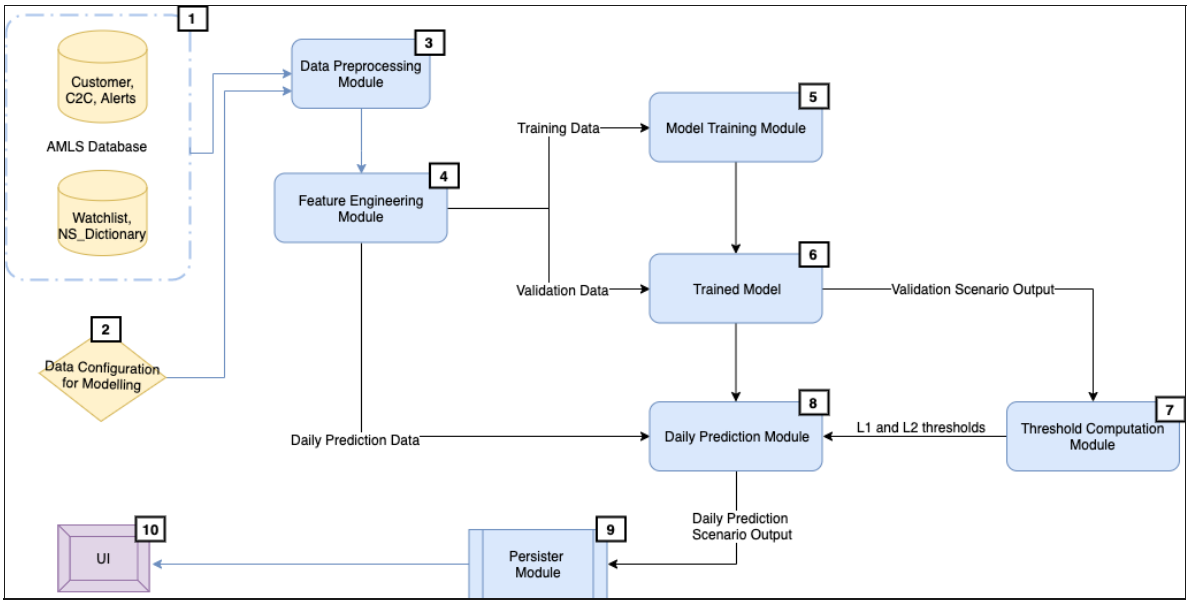
Machine Learning & Advanced Analytics

This chapter describes the machine learning and advanced analytics process used and features available in AMLS.

Supervised Models

Supervised pipelines process the existing alerts along with the associated information to prioritize them into L1, L2 and L3 buckets.

Figure 9 - Supervised Pipeline



1 Data from AMLS Databases
Data consisting of Customer, Alerts, C2C, Watchlist, NS Dictionary and labeled data (true hits, false hits, insufficient hits) is present in the target schema (AMLS), after completion of the ETL process.

2 **Data Configuration for Modeling**
Labeled categories which belong to true hit, false hit and Insufficient Information are pre-configured in the system for modeling purposes.

3 **Data Pre-processing Module**
The pre-processing module performs data availability checks and computes the necessary joins and grouping, to create a joined data frame suitable to be used in the feature engineering module. In the primary scoring alerts system, status is not captured at the hit level, hence the model is trained using only alerts with single-hits. And the alerts associated with multiple hits are filtered out in order to avoid introducing errors in data which can corrupt model training.

4 **Feature Engineering Module**
The feature creation framework generates features on the ingested raw data as required for modeling.

5 **Model Training Module**
Machine Learning Model is trained using the output of feature engineering module on training data. Suitable hyper-parameter configurations are computed based on grid search and cross validation, so that model performance is within the acceptable error bounds.

6 **Trained Model**
Trained model instance is created after model training step. This model instance is used for all prediction tasks, until the model is retrained with self-learning functionality of TDSS.

7 **Threshold Computation Module**
This module computes L1 and L2 probability threshold values for alert prioritization.

8 **Daily Prediction Module**
This module performs alert prioritization on the daily alerts data. It makes use of the computed L1 and L2 thresholds to perform hit-level alert prioritization. For alert-level prioritization, maximum priority across all hits in the alert is used.

9 **Persister Module**
This module helps in storing the prioritized alerts along with explainability in the NS output schema for rendering on the UI. Those alerts which are not prioritized are persisted with an associated error code.

10 **User Interface (UI)**
This module fetches the persisted data to display information of the predicted prioritized alerts and prediction-level explanations

Explainability

Tookitaki provides transparency around complex models and Interpretability of the model results and predictions through the Explainability functionality.

Model explainability provides transparency to model users such as data scientists, business decision makers, to understand the complex machine learning model in a human readable format.

Prediction Explainability enables data scientists to understand and validate the prediction done by the model. It also aids to understand the relevance of the features contributing to the prediction.

Machine learning models do not readily provide functionality to understand the results in an easy business-friendly terminologies. In AMLS, Business explainability is provided via a dedicated web application for the end users.

Model Explainability

The model explainability functionality is to help data scientists and investigators to understand the complex machine learning models in human readable formats. Model explainability also creates transparency around complex machine learning models which is otherwise not possible through out-of-box solutions. Model explainability is not used for any day-to-day decision making and hence, it has no bearing on model predictions.

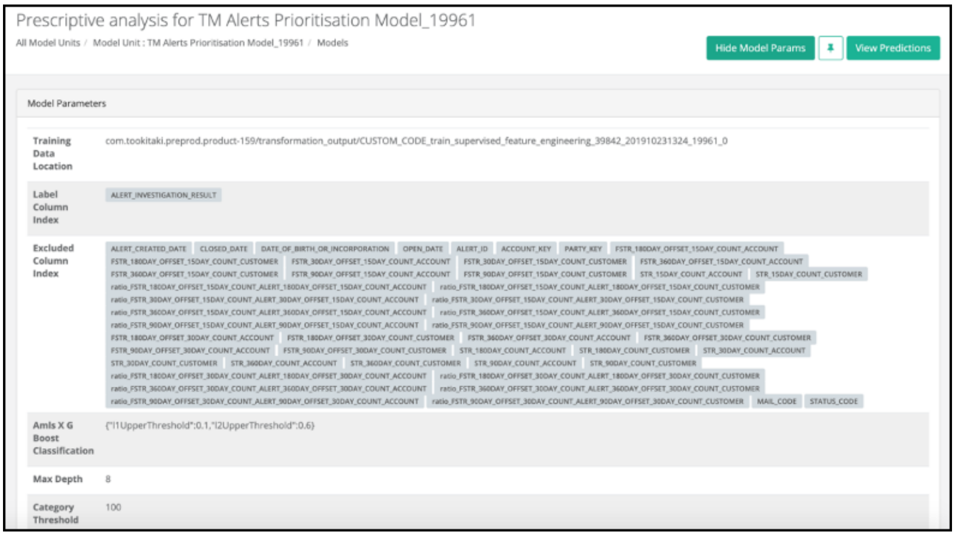
In order to explain the complex model for predictions, TDSS automatically creates an auxiliary model (decision tree). This auxiliary model takes training data input and the output of predictions model to be trained and creates the decision tree and decision rules that explains the Model prediction for the application users.

The following components in TDSS Information Panel are used.

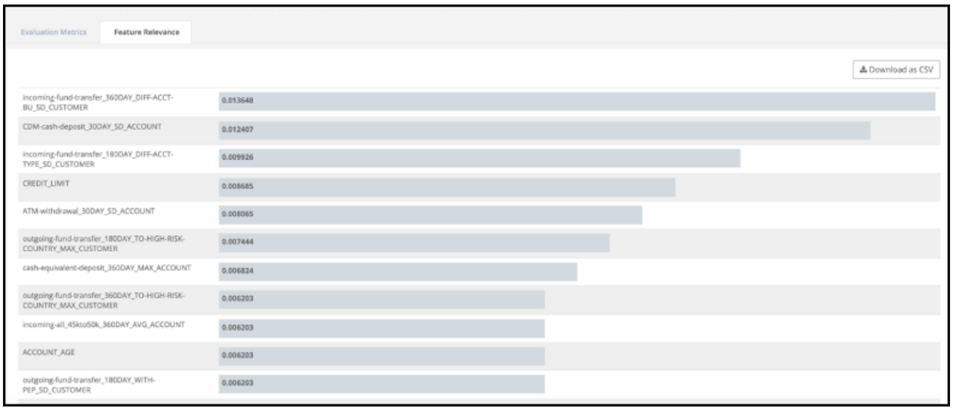
- Training parameters for model
- Label on which training is done
- Excluded column list
- Feature relevance scores for model
- Auxiliary Tree
  - This provides a simplistic approximation of the machine learning model used for predictions

– Since this tree is an approximation, this cannot be used for assessing the prediction accuracy of the model.

Figure 10 - Sample UI Model Output


Prescriptive analysis for TM Alerts Prioritisation Model_19961
All Model Units / Model Unit : TM Alerts Prioritisation Model_19961 / Models
Hide Model Params
View Predictions
Model Parameters

Training Data Location	com.tookitaki.preprod.product-159/transformation_output/CUSTOM_CODE_train_supervised_feature_engineering_39842_201910231324_19961_0
Label Column Index	ALERT_INVESTIGATION_RESULT
Excluded Column Index	ALERT_CREATED_DATE CLOSED_DATE DATE_OF_BIRTH_OR_INCORPORATION OPEN_DATE ALERT_ID ACCOUNT_KEY PARTY_KEY FSTR_180DAY_OFFSET_15DAY_COUNT_ACCOUNT FSTR_180DAY_OFFSET_15DAY_COUNT_CUSTOMER FSTR_30DAY_OFFSET_15DAY_COUNT_ACCOUNT FSTR_30DAY_OFFSET_15DAY_COUNT_CUSTOMER FSTR_360DAY_OFFSET_15DAY_COUNT_ACCOUNT FSTR_360DAY_OFFSET_15DAY_COUNT_CUSTOMER FSTR_90DAY_OFFSET_15DAY_COUNT_ACCOUNT FSTR_90DAY_OFFSET_15DAY_COUNT_CUSTOMER STR_15DAY_COUNT_ACCOUNT STR_15DAY_COUNT_CUSTOMER ratio_FSTR_180DAY_OFFSET_15DAY_COUNT_ALERT_180DAY_OFFSET_15DAY_COUNT_ACCOUNT ratio_FSTR_180DAY_OFFSET_15DAY_COUNT_ALERT_180DAY_OFFSET_15DAY_COUNT_CUSTOMER ratio_FSTR_30DAY_OFFSET_15DAY_COUNT_ALERT_30DAY_OFFSET_15DAY_COUNT_ACCOUNT ratio_FSTR_30DAY_OFFSET_15DAY_COUNT_ALERT_30DAY_OFFSET_15DAY_COUNT_CUSTOMER ratio_FSTR_360DAY_OFFSET_15DAY_COUNT_ALERT_360DAY_OFFSET_15DAY_COUNT_ACCOUNT ratio_FSTR_360DAY_OFFSET_15DAY_COUNT_ALERT_360DAY_OFFSET_15DAY_COUNT_CUSTOMER ratio_FSTR_90DAY_OFFSET_15DAY_COUNT_ALERT_90DAY_OFFSET_15DAY_COUNT_ACCOUNT ratio_FSTR_90DAY_OFFSET_15DAY_COUNT_ALERT_90DAY_OFFSET_15DAY_COUNT_CUSTOMER FSTR_180DAY_OFFSET_30DAY_COUNT_ACCOUNT FSTR_180DAY_OFFSET_30DAY_COUNT_CUSTOMER FSTR_360DAY_OFFSET_30DAY_COUNT_ACCOUNT FSTR_360DAY_OFFSET_30DAY_COUNT_CUSTOMER FSTR_90DAY_OFFSET_30DAY_COUNT_ACCOUNT FSTR_90DAY_OFFSET_30DAY_COUNT_CUSTOMER STR_180DAY_COUNT_ACCOUNT STR_180DAY_COUNT_CUSTOMER STR_30DAY_COUNT_ACCOUNT STR_30DAY_COUNT_CUSTOMER STR_360DAY_COUNT_ACCOUNT STR_360DAY_COUNT_CUSTOMER STR_90DAY_COUNT_ACCOUNT STR_90DAY_COUNT_CUSTOMER ratio_FSTR_180DAY_OFFSET_30DAY_COUNT_ALERT_180DAY_OFFSET_30DAY_COUNT_ACCOUNT ratio_FSTR_180DAY_OFFSET_30DAY_COUNT_ALERT_180DAY_OFFSET_30DAY_COUNT_CUSTOMER ratio_FSTR_360DAY_OFFSET_30DAY_COUNT_ALERT_360DAY_OFFSET_30DAY_COUNT_ACCOUNT ratio_FSTR_360DAY_OFFSET_30DAY_COUNT_ALERT_360DAY_OFFSET_30DAY_COUNT_CUSTOMER ratio_FSTR_90DAY_OFFSET_30DAY_COUNT_ALERT_90DAY_OFFSET_30DAY_COUNT_ACCOUNT ratio_FSTR_90DAY_OFFSET_30DAY_COUNT_ALERT_90DAY_OFFSET_30DAY_COUNT_CUSTOMER MAIL_CODE STATUS_CODE
Amls X G Boost Classification	{"l1UpperThreshold":0.1,"l2UpperThreshold":0.6}
Max Depth	8
Category Threshold	100



incoming-fund-transfer_360DAY_DIFF-ACCT-BU_SD_CUSTOMER	0.013648
CDM-cash-deposit_30DAY_SD_ACCOUNT	0.012407
incoming-fund-transfer_180DAY_DIFF-ACCT-TYPE_SD_CUSTOMER	0.009926
CREDIT_LIMIT	0.008685
ATM-withdrawal_30DAY_SD_ACCOUNT	0.008065
outgoing-fund-transfer_180DAY_TO-HIGH-RISK-COUNTRY_MAX_CUSTOMER	0.007444
cash-equivalent-deposit_360DAY_MAX_ACCOUNT	0.006824
outgoing-fund-transfer_360DAY_TO-HIGH-RISK-COUNTRY_MAX_CUSTOMER	0.006203
incoming-all_45kto50k_360DAY_AVG_ACCOUNT	0.006203
ACCOUNT_AGE	0.006203
outgoing-fund-transfer_180DAY_WITH-PEP_SD_CUSTOMER	0.006203

Prediction Explainability

Prediction explainability is important to understand the outcome of the model i.e. probabilities for different classes assigned to each data point.

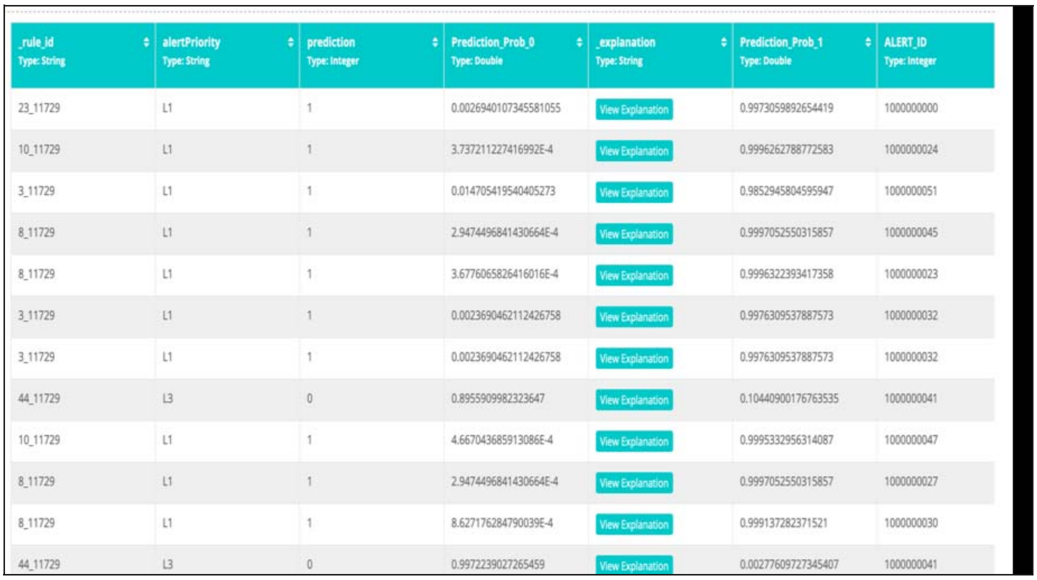
The components for Predictions Explainability are:

- Prediction path for XGBoost.
- Directional Impact Plots for the features.
- Feature matrix for particular prediction.

For each model run, all the predictions can be viewed in TDSS as shown in the screenshots below:

1 The prediction probability, alert priority assigned, the alert ID etc.

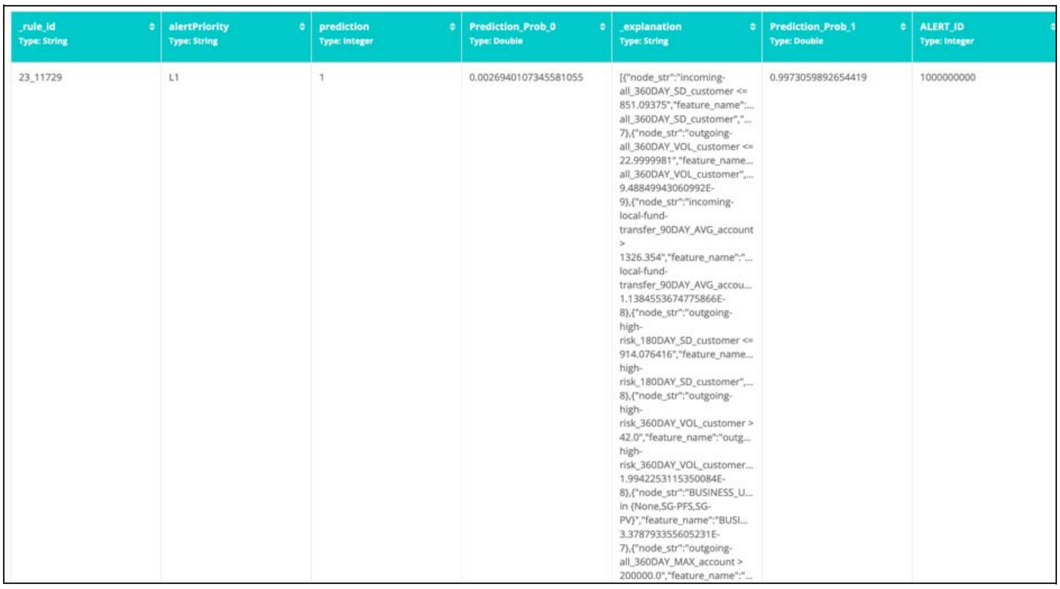
Figure 11 - Prediction Output (Alert Priority and probabilities)



_rule_id Type: String	alertPriority Type: String	prediction Type: Integer	Prediction_Prob_0 Type: Double	_explanation Type: String	Prediction_Prob_1 Type: Double	ALERT_ID Type: Integer
23_11729	L1	1	0.0026940107345581055	View Explanation	0.9973059892654419	1000000000
10_11729	L1	1	3.737211227416992E-4	View Explanation	0.9996262788772583	1000000024
3_11729	L1	1	0.014705419540405273	View Explanation	0.9852945804595947	1000000051
8_11729	L1	1	2.9474496841430664E-4	View Explanation	0.9997052550315857	1000000045
8_11729	L1	1	3.6776065826416016E-4	View Explanation	0.9996322393417358	1000000023
3_11729	L1	1	0.0023690462112426758	View Explanation	0.9976309537887573	1000000032
3_11729	L1	1	0.0023690462112426758	View Explanation	0.9976309537887573	1000000032
44_11729	L3	0	0.8955909982323647	View Explanation	0.10440900176763535	1000000041
10_11729	L1	1	4.6670436859130086E-4	View Explanation	0.9995332956314087	1000000047
8_11729	L1	1	2.9474496841430664E-4	View Explanation	0.9997052550315857	1000000027
8_11729	L1	1	8.627176284790039E-4	View Explanation	0.999137282371521	1000000030
44_11729	L3	0	0.9972239027265459	View Explanation	0.00277609727345407	1000000041

2 The XGBoost decision rules associated with each prediction are also shown.

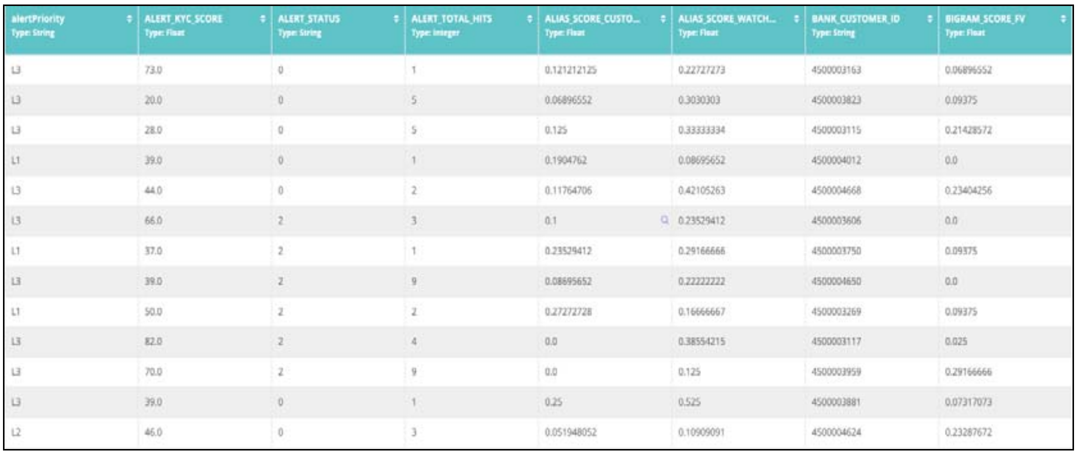
Figure 12 - Prediction Output (Decision rules of XGBoost)



_rule_id Type: String	alertPriority Type: String	prediction Type: Integer	Prediction_Prob_0 Type: Double	_explanation Type: String	Prediction_Prob_1 Type: Double	ALERT_ID Type: Integer
23_11729	L1	1	0.0026940107345581055	[{"node_str":"incoming-all_360DAY_SD_customer <= 851.09375","feature_name":... all_360DAY_SD_customer","... 7},{"node_str":"outgoing-all_360DAY_VOL_customer <= 22.9999981","feature_name... all_360DAY_VOL_customer",... 9.48849943060992E-9},{"node_str":"incoming-local-fund-transfer_90DAY_AVG_account > 1326.354","feature_name":"... local-fund-transfer_90DAY_AVG_accou... 1.1384553674775866E-8},{"node_str":"outgoing-high-risk_180DAY_SD_customer <= 914.076416","feature_name... high-risk_180DAY_SD_customer",... 8},{"node_str":"outgoing-high-risk_360DAY_VOL_customer > 42.0","feature_name":"outg... high-risk_360DAY_VOL_customer... 1.9942253115350084E-8},{"node_str":"BUSINESS_U... in (None,SG-PFS,SG-PV)","feature_name":"BUSI... 3.378793355605231E-7},{"node_str":"outgoing-all_360DAY_MAX_account > 200000.0","feature_name":"...	0.9973059892654419	1000000000

3 The feature matrix associated with the prediction.

Figure 13 - Prediction Output (Feature Matrix)



alertPriority Type: String	ALERT_KYC_SCORE Type: Float	ALERT_STATUS Type: String	ALERT_TOTAL_HITS Type: Integer	ALIAS_SCORE_CUSTO... Type: Float	ALIAS_SCORE_WATCH... Type: Float	BANK_CUSTOMER_ID Type: String	BIGRAM_SCORE_FV Type: Float
L3	73.0	0	1	0.121212125	0.22727273	4500003163	0.06896552
L3	20.0	0	5	0.06896552	0.3030303	4500003823	0.09375
L3	28.0	0	5	0.125	0.33333334	4500003115	0.21428572
L1	39.0	0	1	0.1904762	0.08695652	4500004012	0.0
L3	44.0	0	2	0.11764706	0.42105263	4500004668	0.23404256
L3	66.0	2	3	0.1	0.23529412	4500003606	0.0
L1	37.0	2	1	0.23529412	0.29166666	4500003750	0.09375
L3	39.0	2	9	0.08695652	0.22222222	4500004650	0.0
L1	50.0	2	2	0.27272728	0.16666667	4500003269	0.09375
L3	82.0	2	4	0.0	0.38554215	4500003117	0.025
L3	70.0	2	9	0.0	0.125	4500003959	0.29166666
L3	39.0	0	1	0.25	0.525	4500003881	0.07317073
L2	46.0	0	3	0.051948052	0.10909091	4500004624	0.23287672

The details of the attributes are defined in the table below.

Table 3 - Prediction Explainability

Field	**Value**	**Interpretation**
alertPriority	L1, L2, L3 bucket	The severity of alert
Prediction_prob_0	0 to 1	Model prediction for class 0
Explanation	Explanation generated by XGBoost	Path traversed by XGBoost

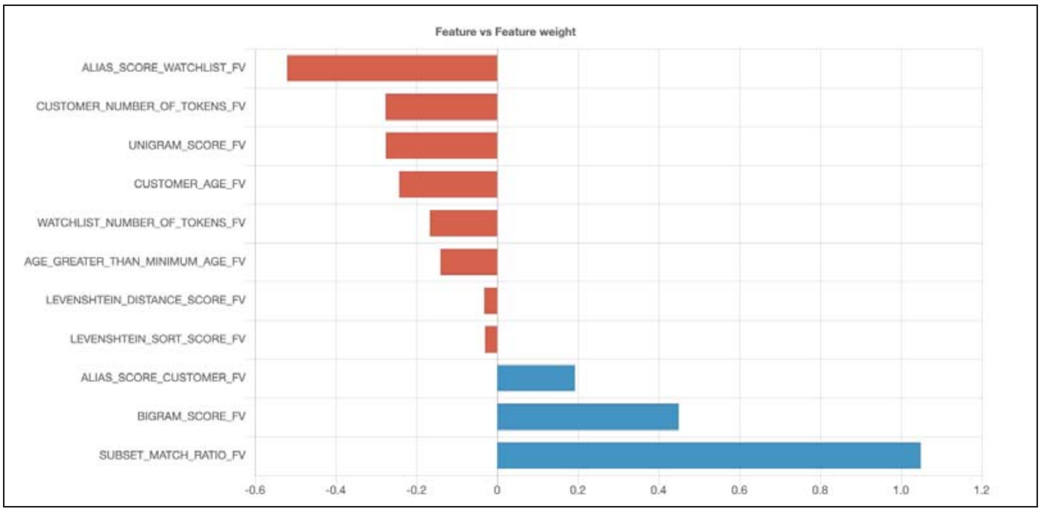
Table 3 - Prediction Explainability

Prediction_prob_1	0 to 1	Model prediction for class 1
Alert_ID	AlphaNumberical ID	Alert Id in AMLS system
Feature Matrix (multiple columns)	All features created for particular prediction with corresponding data types	The actual values of the features computed for the alert associated with the prediction

Directional Impact Plots

Directional impact plots enables the users to understand the positive and negative impact of the features on the predictions.

Figure 14 - Directional Impact Plots



Self Learning

The Self-Learning unit automatically ensures the Alert Prioritization Unit is updated and maintained with time. The unit enables automatic, continuous learning of models to avoid performance degradation and applying champion-challenger approach to learn from incremental data. Incremental data refers to new records and analyst/investigator feedback.

This self learning process is triggered at a defined/fixed frequency (monthly, quarterly, yearly) and event driven. The Champion Challenger framework is designed to work as described below.

1. Start with the best initial model and mark it as the Champion.
2. Build a Challenger model by extracting newer data, automatic selection of top features and automatic hyper-parameter selection.
3. Prepare the testing scenarios/validation dataset to compare both models.
4. Compare both current Champion and Challenger models based on the validation framework.
5. If the Challenger model improves by a set threshold, the Challenger is automatically promoted.

6. Record history and evolution of models and their performance for model audit purpose.

How It Works

Self-learning framework gets data points from three sources:

- Any new customer datasets.
- Investigator feedbacks while reviewing alerts.
- New unknown cases generated from unsupervised approach.

New Challenger Model

1. Create a new Challenger Model.
   - Periodic model retraining, self-learning, and calibration is fully automated.
2. Model review when challenger beats champion.
   - When better challenger found, client model management team needs to evaluate the challenger model against the champion in TDSS.
   - Review vital decision parameters – model precision, misclassification, etc.
3. Replacing a champion with new challenger.
   - **Make as Champion** button in TDSS.
   - No data clean up necessary; data kept for future auditing.

Champion-Challenger in Self-Learning

- Self-learning begins with an initial trained model – the current champion.
- Over time, more data is accumulated.
- Periodically, the model is trained with a new dataset and tested against this new challenger.
- If the challenger is not as good, it is ignored.
- If the challenger is better, it can become the new champion (automatically/ manually).
- Machine learning models degrade over time in a dynamic space such as AML. However, with self-learning, improved models take over when old model degradation occurs.

Names Screening Analytics Unit

The Analytics Unit is the core engine of AMLS where Models, Pipelines and the components are built for Alerts Prioritization and the respective predictions. The users are able to view details of the following:

- NS Alert Prioritization (Supervised Pipeline)
  - Connectors for the database
  - Training & Validation
  - View Executions
  - View Model Performance
  - View Model Features Contribution
  - Model Evolution
- Prediction Pipeline for NS Alert Prioritization
  - Predict New Data
  - Feature Contribution and Prediction Explainability
- Pipeline Execution
  - Pipeline Setup and Execution Steps

Note: TDSS is a read-only application. The functions featured in this document are subject to limitations.

NS Alerts Prioritization

Alerts Prioritization

In Names Screening Alerts Prioritization, the pipeline is to train supervised model for name screening. This is where the Champion-Challenger framework is used. If there is no initial model, the default best model is made the champion. If a champion is present, the latest challenger is built to compare its performance with the champion.

Supervised Pipelines

Pipelines are sequential operations on data which may include stages from data preparation, data transformation, model building to model validation, mentioned in an interactive way and which can be automated for similar operations in the future.

Each stage of pipeline can be analyzed and to view the input and output of the files. When self-learning has been configured, the model part of pipeline also mentions the evaluation metrics and feature relevance list of pipelines that can be viewed in the system.

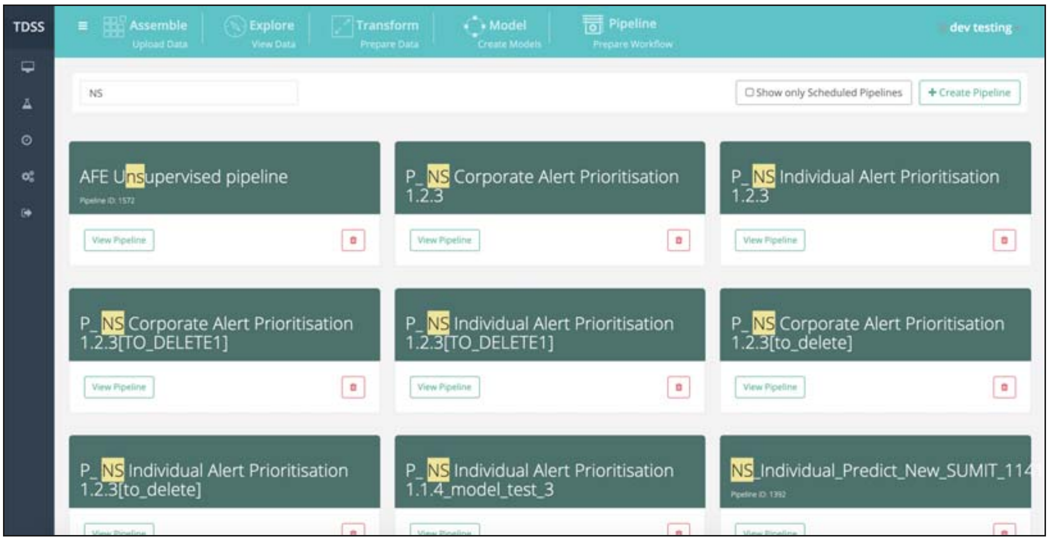
For names screening alerts prioritization, the supervised pipeline is used for both training/validation and prediction of individual or corporate accounts.

View Pipelines

In TDSS, goto **Dashboard** > **Pipeline** > **View Pipeline**.

The pipeline page displays the preset pipelines available for the application. The supervised and unsupervised pipelines have two sets of pipelines each, for the purpose of training/validation and prediction.

Figure 15 - View Pipeline



AFE Unsupervised pipeline
Pipeline ID: 1572

P_NS Corporate Alert Prioritisation 1.2.3

P_NS Individual Alert Prioritisation 1.2.3

P_NS Corporate Alert Prioritisation 1.2.3[TO_DELETE1]

P_NS Individual Alert Prioritisation 1.2.3[TO_DELETE1]

P_NS Corporate Alert Prioritisation 1.2.3[to_delete]

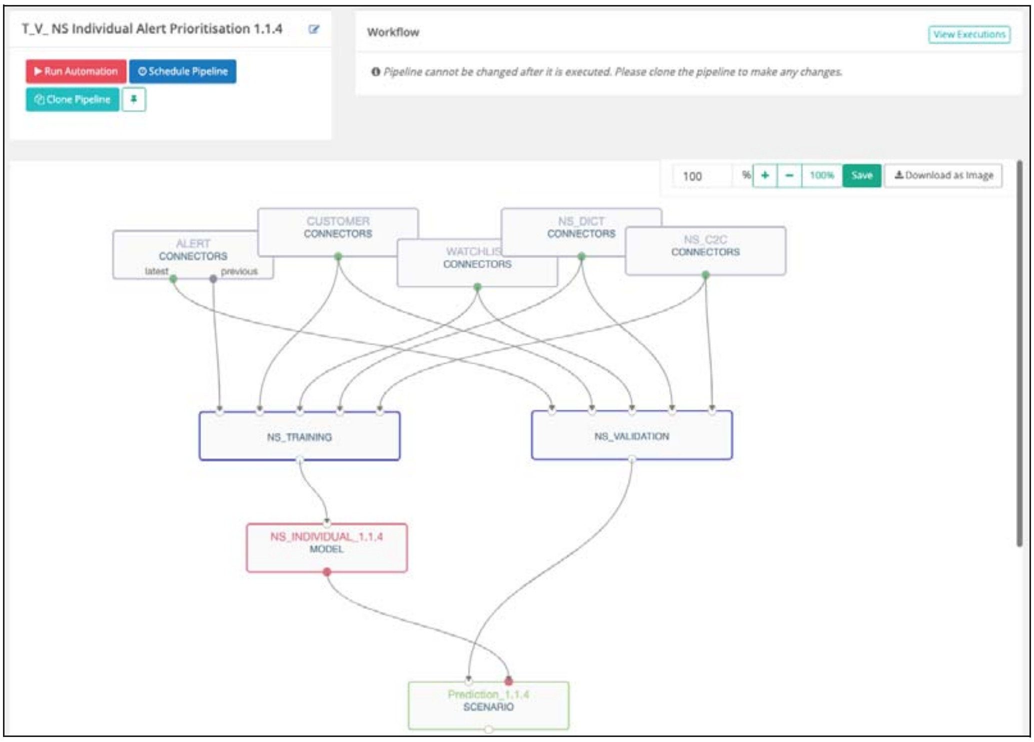
P_NS Individual Alert Prioritisation 1.2.3[to_delete]

P_NS Individual Alert Prioritisation 1.1.4_model_test_3

NS_Individual_Predict_New_SUMIT_114
Pipeline ID: 1392

Select the pipeline to view and the components of the pipeline are displayed as shown below.

Figure 16 - View Pipeline



T_V_ NS Individual Alert Prioritisation 1.1.4

Workflow

Pipeline cannot be changed after it is executed. Please clone the pipeline to make any changes.

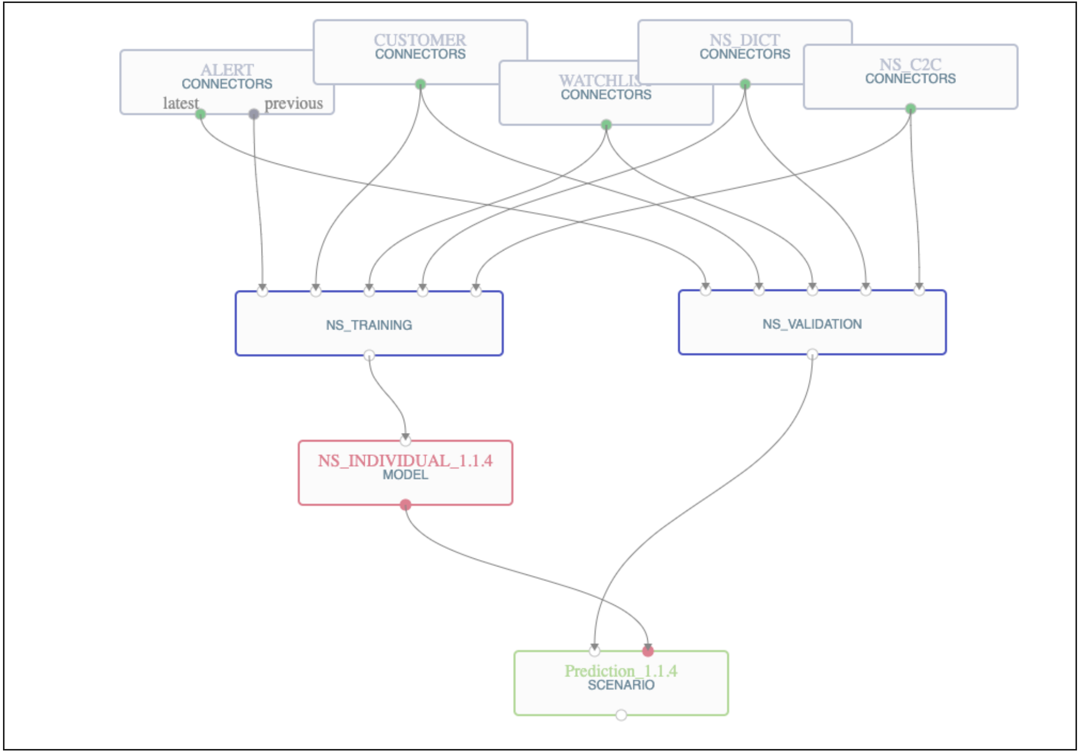
Connectors for Database

In TDSS, go to **Dashboard** > **Pipeline** > **View Pipeline**.

Alerts Prioritization Supervised Pipeline

The connectors for importing individual accounts data from the source used in the supervised training and validation pipeline is shown below.

Figure 17 - Individual Account Connectors (Training & Validation)



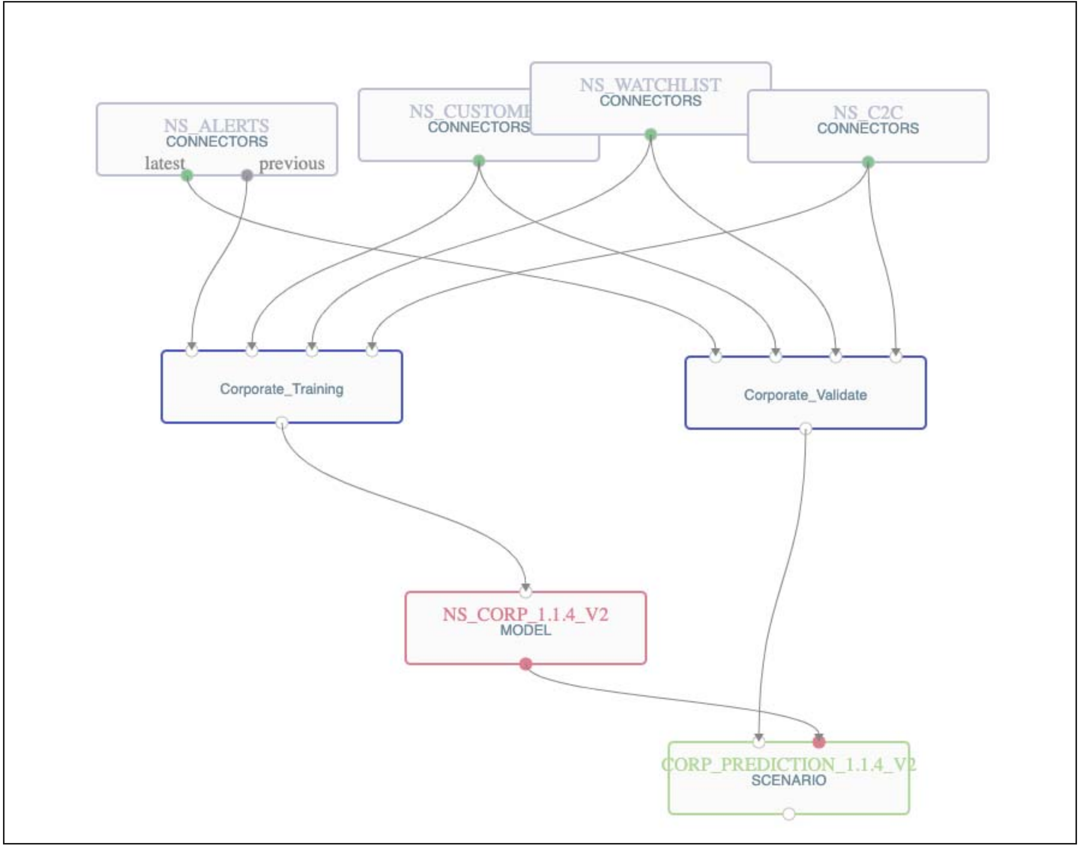
Each connector represents the functionalities used in the supervised pipeline for training and validation as described in the table below.

Table 4 - Individual Account Connectors (Training & Validation)

Connector	**Description**
Alerts	All daily unseen alerts for prioritization.
Customer	All active customer data for training.
Watchlist	All active watchlist data for training.
NS Dictionary	Name variations.
C2C	All related customers data from customer_to_customer table.

The connectors for importing corporate accounts data from the source used in the supervised training and validation pipeline is shown below.

Figure 18 - Corporate Account Connectors (Training & Validation)



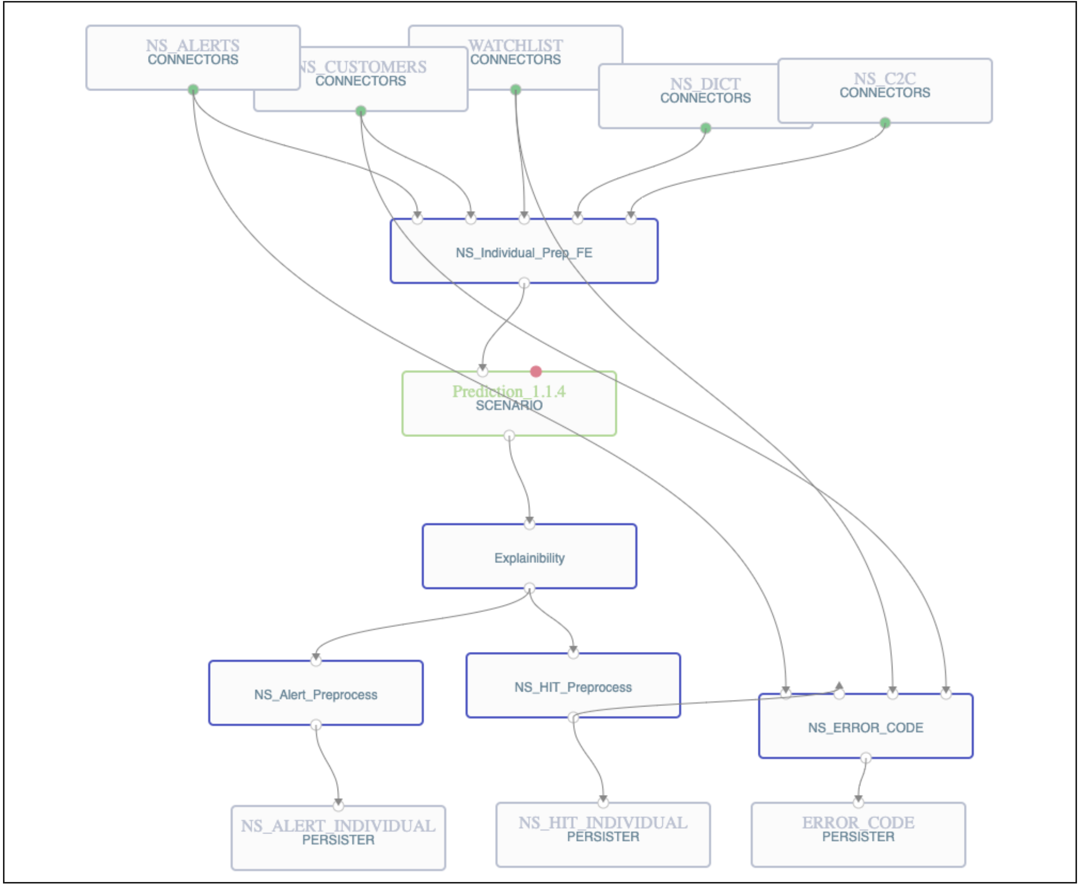
Each connector represents the functionalities used in the supervised pipeline for training and validation as described in the table below.

Table 5 - Corporate Account Connectors (Training & Validation)

Connector	**Description**
Alerts	All daily unseen alerts for prioritization.
Customer	All active customer data for training.
Watchlist	All active watchlist data for training.
C2C	All related customers data from customer_to_customer table.

The connectors for importing individual account data from the source used in the supervised prediction pipeline is shown below.

Figure 19 - Individual Account Connectors (Prediction)



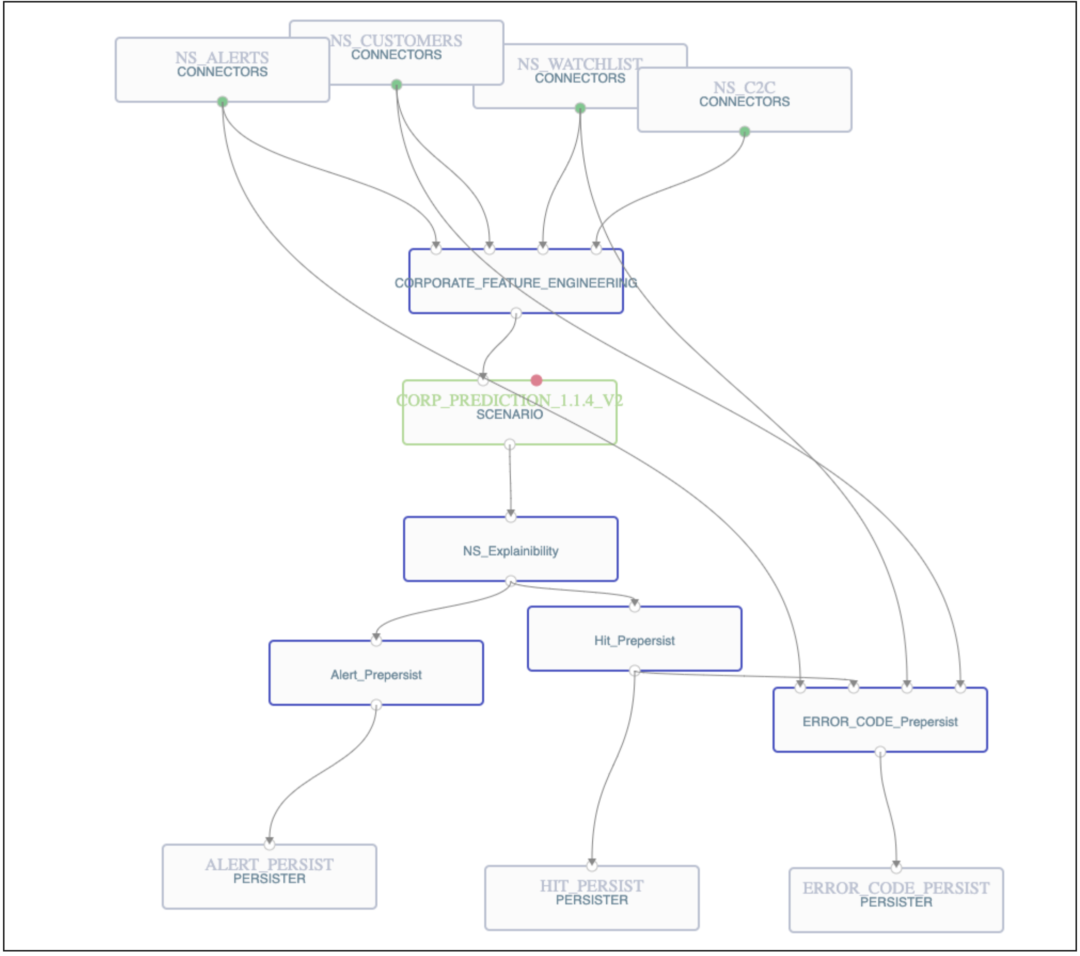
Each connector represents the functionalities used in the supervised pipeline for prediction as described in the table below

Table 6 - Individual Account Connectors (Prediction)

Connector	Description
Alerts	All daily unseen alerts for prioritization.
Customers	All active customer data for prediction.
Watchlist	All active watchlist data for prediction.
NS Dictionary	Name variations.
C2C	All related customers data from customer_to_customer table.

The connectors for importing corporate account data from the source used in the supervised prediction pipeline is shown below.

Figure 20 - Corporate Account Connectors (Prediction)



Each connector represents the functionalities used in the supervised pipeline for prediction as described in the table below

Table 7 - Corporate Account Connectors (Prediction)

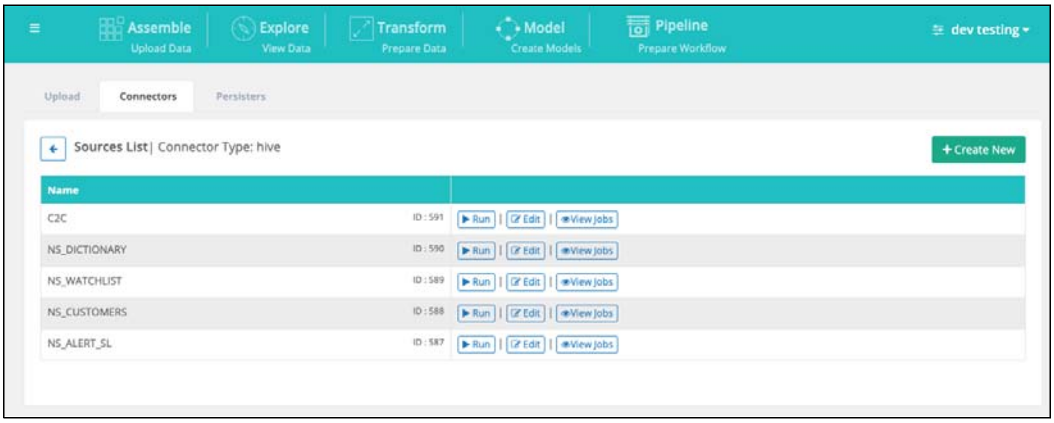
Connector	**Description**
Alerts	All daily unseen alerts for prioritization.
Customers	All active customer data for prediction.
Watchlist	All active watchlist data for prediction.
C2C	All related customers data from customer_to_customer table.

Setting Up Connectors

In TDSS, go to **Dashboard** > **Assemble** > **Connectors**.

Under the Connectors tab, the list of connectors used are displayed. As an example, the NS_ALERTS_SL connector is used here. The user is able to view the details of the self-learning connector by clicking **Edit** on the NS_ALERTS_SL connector.

Figure 21 - Setting Up Connectors

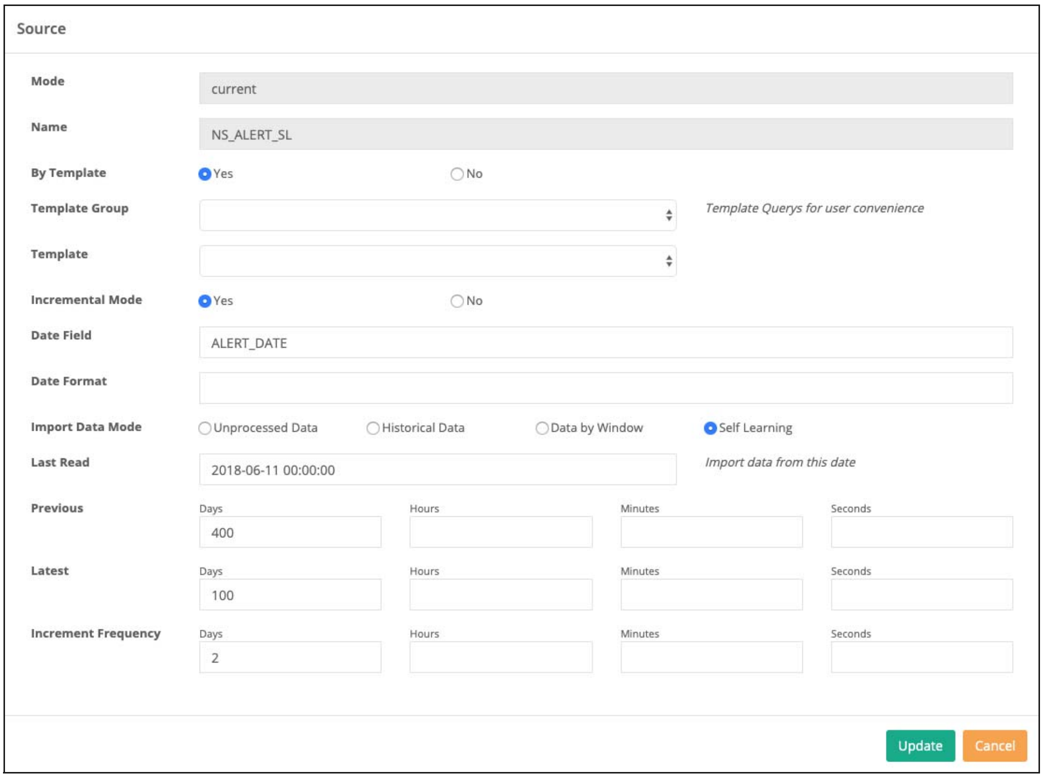


Sources List | Connector Type: hive

Name		
C2C	ID : 591	▶ Run | Edit | View Jobs
NS_DICTIONARY	ID : 590	▶ Run | Edit | View Jobs
NS_WATCHLIST	ID : 589	▶ Run | Edit | View Jobs
NS_CUSTOMERS	ID : 588	▶ Run | Edit | View Jobs
NS_ALERT_SL	ID : 587	▶ Run | Edit | View Jobs

The pop up screen as shown below is displayed. The user is able to view or edit the details such as Last Read, Previous, Latest and Increment Frequency of data import.

Figure 22 - Edit Connectors



Source

Mode	current			
Name	NS_ALERT_SL			
By Template	● Yes	○ No		
Template Group			Template Querys for user convenience	
Template				
Incremental Mode	● Yes	○ No		
Date Field	ALERT_DATE			
Date Format				
Import Data Mode	○ Unprocessed Data	○ Historical Data	○ Data by Window	● Self Learning
Last Read	2018-06-11 00:00:00		Import data from this date	
Previous	Days 400	Hours	Minutes	Seconds
Latest	Days 100	Hours	Minutes	Seconds
Increment Frequency	Days 2	Hours	Minutes	Seconds

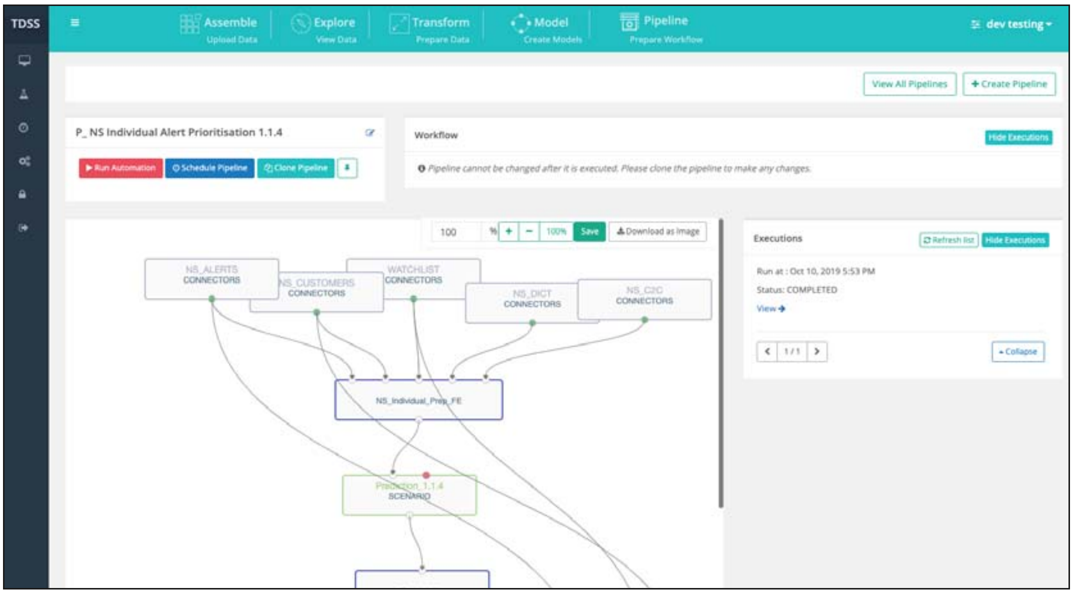
View Executions

In TDSS, goto **Pipeline** > **View Pipeline**> **View Executions**.

In a pipeline, user is able to view the executions of the pipeline and its components. This shows whether the pipeline has been executed successfully or there are errors in certain parts of the pipeline. The status of the executions is indicated in the side bar that appears when user clicks **View Executions**.

Successful or unsuccessful executions displays the **Completed** or **Failed** status respectively. Executions that are currently running shows the **Running** status.

Figure 23 - View Executions



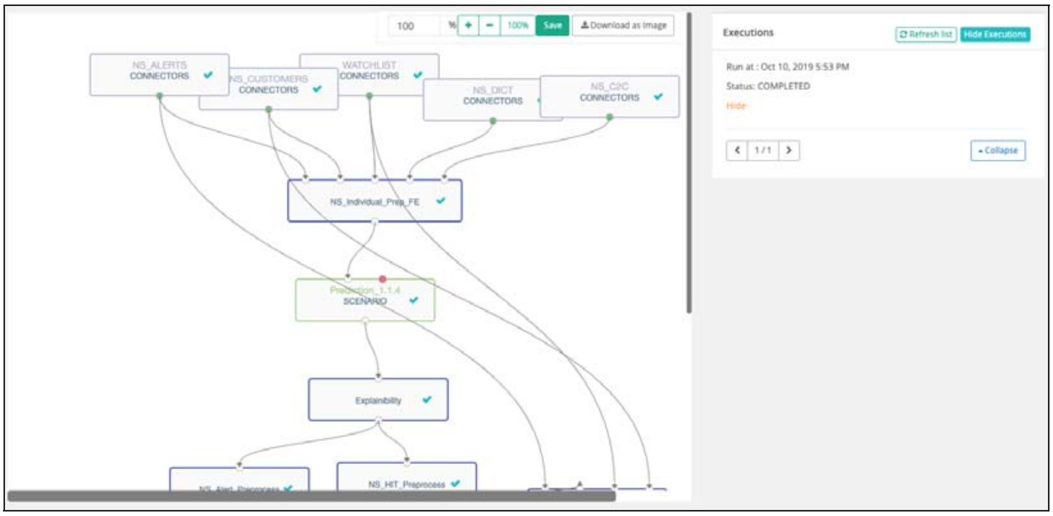
Pipeline cannot be changed after it is executed. Please clone the pipeline to make any changes.

Run at : Oct 10, 2019 5:53 PM

Status: COMPLETED

Click **View** to see the statuses of the pipeline components and view the errors contributing to failed executions. Refer to “Pipeline Component Status”.

Figure 24 - Execution Status



Run at : Oct 10, 2019 5:53 PM

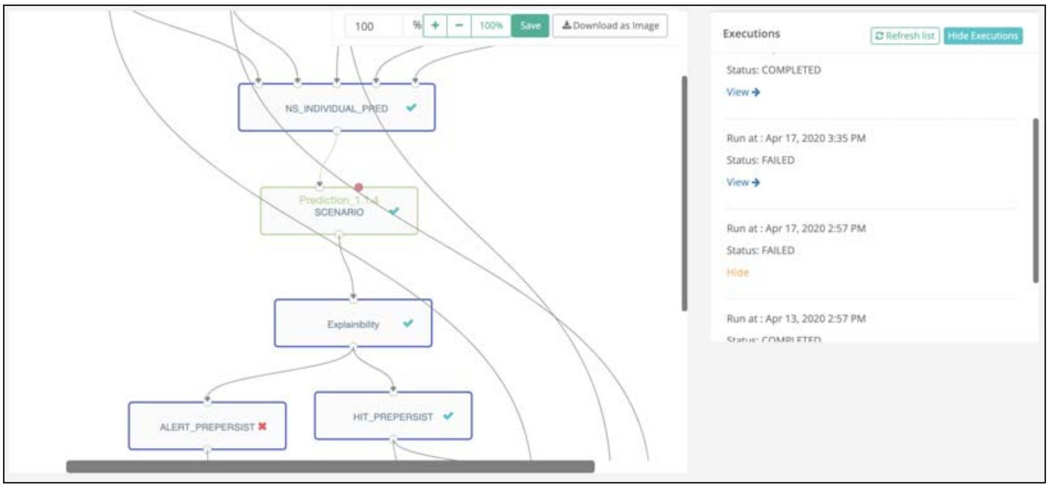
Status: COMPLETED

Pipeline Component Status

User is able to view each of the pipeline component status by clicking the specific component to view.

In all executions, components with no errors are indicated with a ✓. When user views a failed execution, the component with the error is indicated with an ✖ as shown in the example below.

Figure 25 - Pipeline Component Status



Status: COMPLETED
View →

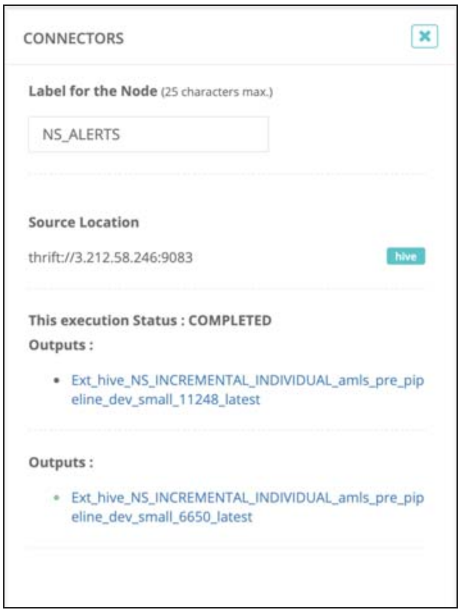
Run at : Apr 17, 2020 3:35 PM
Status: FAILED
View →

Run at : Apr 17, 2020 2:57 PM
Status: FAILED
Hide

Run at : Apr 13, 2020 2:57 PM
Status: COMPLETED

Click the component with the ✓ to view more details on the side bar shown in the example below.

Figure 26 - Component Details



Label for the Node (25 characters max.)

NS_ALERTS

Source Location

thrift://3.212.58.246:9083 hive

This execution Status : COMPLETED

Outputs :

- Ext_hive_NS_INCREMENTAL_INDIVIDUAL_amls_pre_pipeline_dev_small_11248_latest

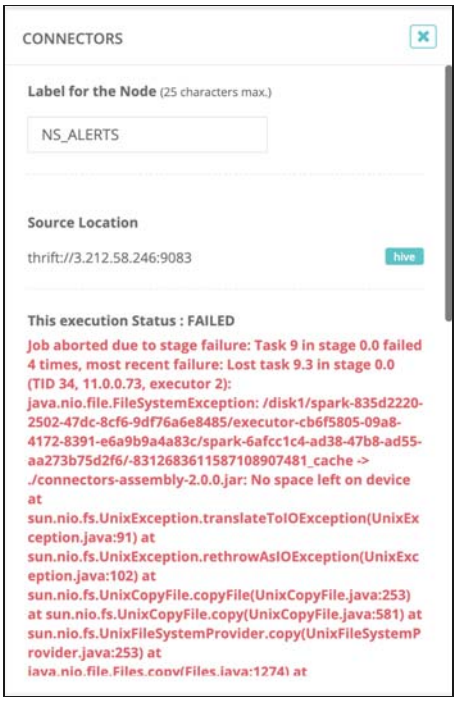
Outputs :

- Ext_hive_NS_INCREMENTAL_INDIVIDUAL_amls_pre_pipeline_dev_small_6650_latest

Click the output links provided to be redirected to view the output details.

Click the component with the ✖ to view the details of the error on the side bar as shown in the example below.

Figure 27 - Component Error Details



Label for the Node (25 characters max.)

NS_ALERTS

Source Location

thrift://3.212.58.246:9083

This execution Status : FAILED

```
Job aborted due to stage failure: Task 9 in stage 0.0 failed
4 times, most recent failure: Lost task 9.3 in stage 0.0
(TID 34, 11.0.0.73, executor 2):
java.nio.file.FileSystemException: /disk1/spark-835d2220-
2502-47dc-8cf6-9df76a6e8485/executor-cb6f5805-09a8-
4172-8391-e6a9b9a4a83c/spark-6afcc1c4-ad38-47b8-ad55-
aa273b75d2f6/-8312683611587108907481_cache ->
./connectors-assembly-2.0.0.jar: No space left on device
at
sun.nio.fs.UnixException.translateToIOException(UnixEx
ception.java:91) at
sun.nio.fs.UnixException.rethrowAsIOException(UnixExc
eption.java:102) at
sun.nio.fs.UnixCopyFile.copyFile(UnixCopyFile.java:253)
at sun.nio.fs.UnixCopyFile.copy(UnixCopyFile.java:581) at
sun.nio.fs.UnixFileSystemProvider.copy(UnixFileSystemP
rovider.java:253) at
java.nio.file.Files.copy(Files.java:1274) at
```

User can view the errors of the execution status and rectify accordingly, if needed.

View Model Performance

In TDSS, go to **Dashboard** > **Model Units** > **Models**.

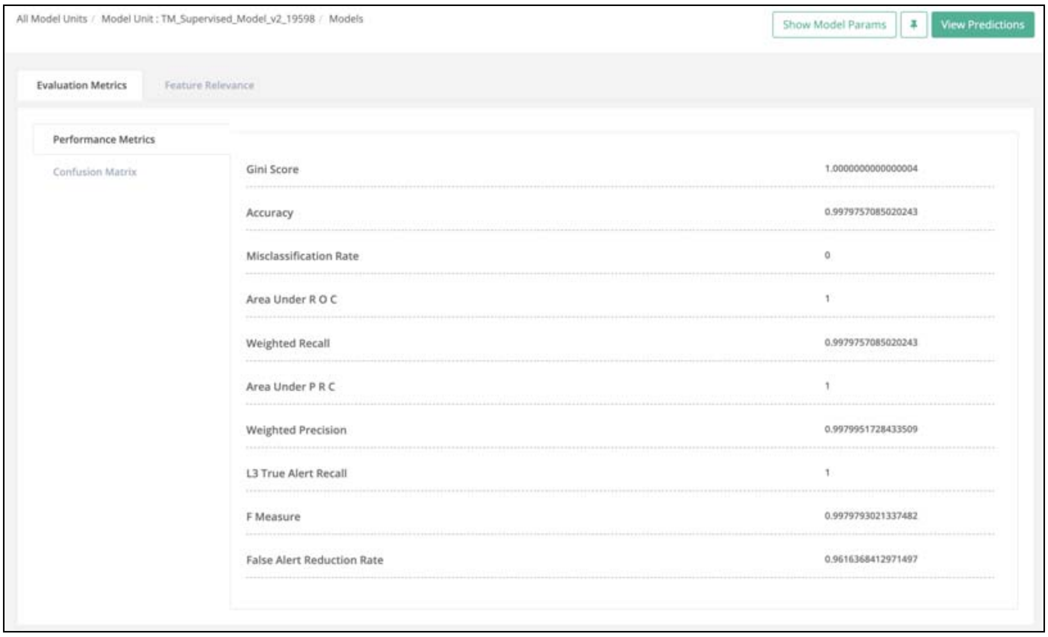
Evaluation Metrics

In View Model, under the Evaluation Metrics tab, user is able to view the details of its Performance Metrics and Confusion Matrix as shown below.

Performance Metrics

The Performance Metrics show the calculations pertaining to the model. The list of calculations vary depending on the model details.

Figure 28 - Performance Metrics

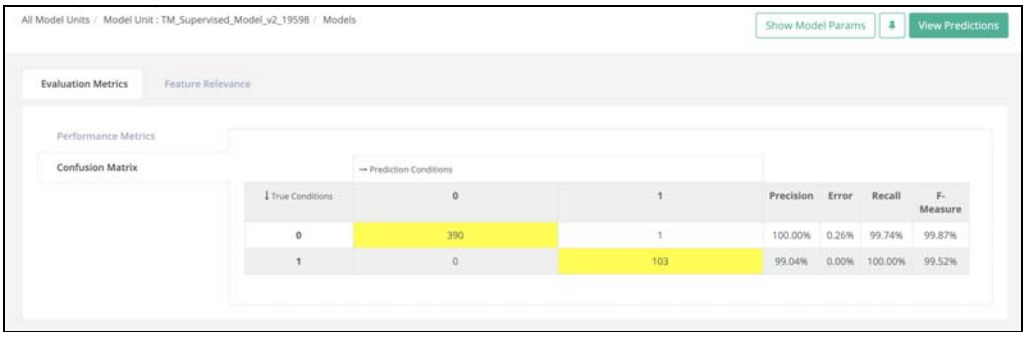


Gini Score	1.0000000000000004
Accuracy	0.9979757085020243
Misclassification Rate	0
Area Under R O C	1
Weighted Recall	0.9979757085020243
Area Under P R C	1
Weighted Precision	0.9979951728433509
L3 True Alert Recall	1
F Measure	0.9979793021337482
False Alert Reduction Rate	0.9616368412971497

Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or classifier) on a set of data for which the true values are known.

Figure 29 - Confusion Matrix



	→ Prediction Conditions					
↓ True Conditions	0	1	Precision	Error	Recall	F-Measure
0	390	1	100.00%	0.26%	99.74%	99.87%
1	0	103	99.04%	0.00%	100.00%	99.52%

The boxes that are highlighted yellow show the true results (true negatives 0 and true positives 1) and the other two boxes show where predictions are not correct (false negative 0 and false positive 1).

View Model Features Contribution

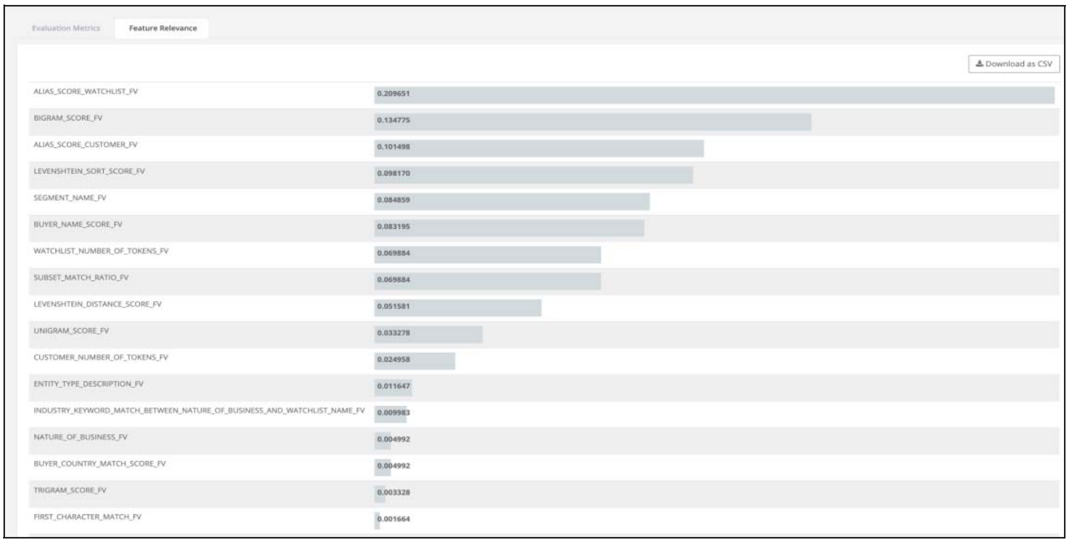
In TDSS, go to **Dashboard** > **Models** > **View Models**.

Feature Relevance

Under the Feature Relevance tab, it shows the relative importance of top features in the built model with relevance score. This is specific to only certain models: Random Forest, XGboost and stacked model with combiner model. This information is not specific to any particular prediction outcome, but to the whole model.

The features vary depending on the model.

Figure 30 – Feature Relevance

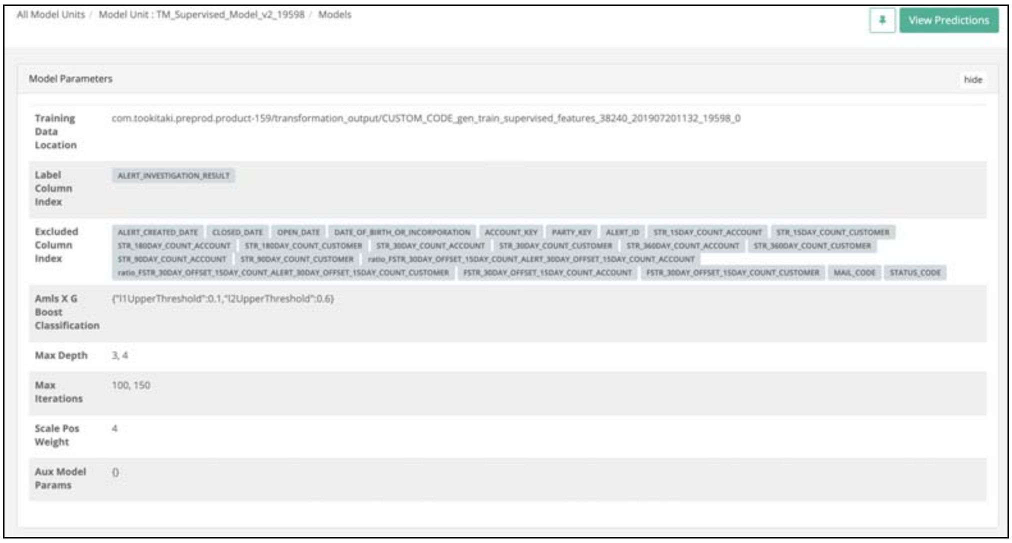


ALIAS_SCORE_WATCHLIST_FV	0.209651
BIGRAM_SCORE_FV	0.134775
ALIAS_SCORE_CUSTOMER_FV	0.101498
LEVENSHTEIN_SORT_SCORE_FV	0.098170
SEGMENT_NAME_FV	0.084859
BUYER_NAME_SCORE_FV	0.083195
WATCHLIST_NUMBER_OF_TOKENS_FV	0.069884
SUBSET_MATCH_RATIO_FV	0.069884
LEVENSHTEIN_DISTANCE_SCORE_FV	0.051581
UNIGRAM_SCORE_FV	0.033278
CUSTOMER_NUMBER_OF_TOKENS_FV	0.024958
ENTITY_TYPE_DESCRIPTION_FV	0.011647
INDUSTRY_KEYWORD_MATCH_BETWEEN_NATURE_OF_BUSINESS_AND_WATCHLIST_NAME_FV	0.009983
NATURE_OF_BUSINESS_FV	0.004992
BUYER_COUNTRY_MATCH_SCORE_FV	0.004992
TRIGRAM_SCORE_FV	0.003328
FIRST_CHARACTER_MATCH_FV	0.001664

Model Parameters

On the top right, user is able to click **Show Model Params** to view the model parameters as shown below. The parameters displayed vary depending on the model.

Figure 31 - Model Parameters



All Model Units / Model Unit : TM_Supervised_Model_v2_19598 / Models
View Predictions
Model Parameters
hide

Training Data Location	com.tookitaki.preprod.product-159/transformation_output/CUSTOM_CODE_gen_train_supervised_features_38240_201907201132_19598_0
Label Column Index	ALERT_INVESTIGATION_RESULT
Excluded Column Index	ALERT_CREATED_DATE CLOSED_DATE OPEN_DATE DATE_OF_BIRTH_OR_INCORPORATION ACCOUNT_KEY PARTY_KEY ALERT_ID STR_15DAY_COUNT_ACCOUNT STR_15DAY_COUNT_CUSTOMER STR_180DAY_COUNT_ACCOUNT STR_180DAY_COUNT_CUSTOMER STR_30DAY_COUNT_ACCOUNT STR_30DAY_COUNT_CUSTOMER STR_360DAY_COUNT_ACCOUNT STR_360DAY_COUNT_CUSTOMER STR_90DAY_COUNT_ACCOUNT STR_90DAY_COUNT_CUSTOMER ratio_FSTR_30DAY_OFFSET_15DAY_COUNT_ALERT_30DAY_OFFSET_15DAY_COUNT_ACCOUNT ratio_FSTR_30DAY_OFFSET_15DAY_COUNT_ALERT_30DAY_OFFSET_15DAY_COUNT_CUSTOMER FSTR_30DAY_OFFSET_15DAY_COUNT_ACCOUNT FSTR_30DAY_OFFSET_15DAY_COUNT_CUSTOMER MAIL_CODE STATUS_CODE
Amls X G Boost Classification	{"l1UpperThreshold":0.1,"l2UpperThreshold":0.6}
Max Depth	3, 4
Max Iterations	100, 150
Scale Pos Weight	4
Aux Model Params	{}

Self-Learning

In TDSS self-learning is its ability to make models learn on their own iteratively. As new data is periodically added, the behavior predicted by the model might become obsolete and require updating. This is ensured by self learning feature which uses challenger and champion method to find better model at each iteration.

At each iteration, new challenger model is built with new data (including the old data) whose performance is compared with the current champion model. It replaces the champion model in case the difference of performance crosses the threshold set by user. The performance metric for comparing challenger and champion model and threshold value for the swapping is set by user. The system can carry out Model Evolution if it is automated for an interval in Self-Learning Parameters mentioned below.

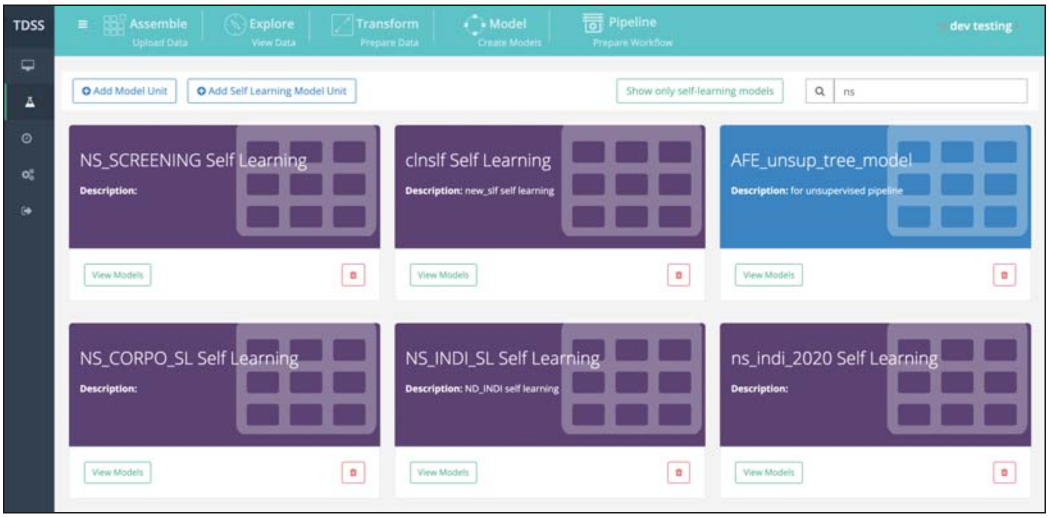
For details on building of Self-Learning units, refer to “Add Self-Learning Model” on page 44.

Model Evolution

In TDSS, go to **Dashboard** > **Models**.

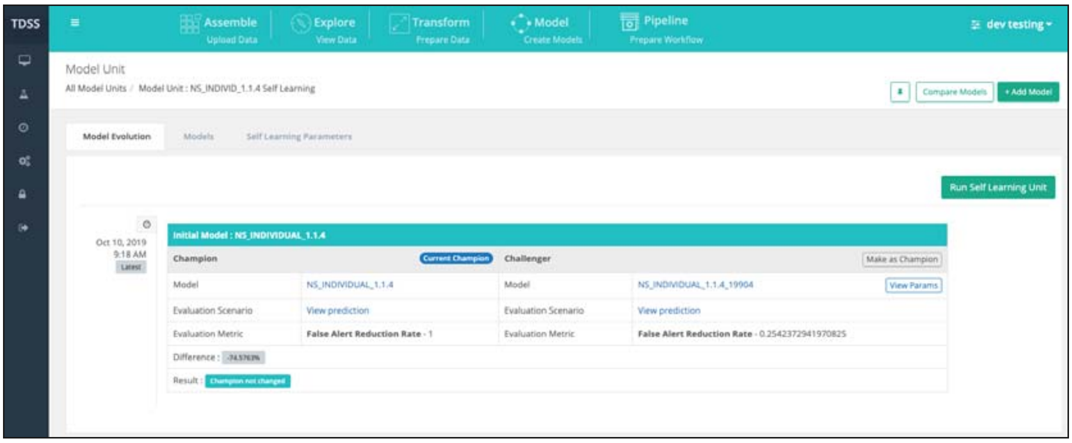
The self-learning models are displayed as shown below. The models in purple indicate self-learning models. Scroll and select the desired self-learning model by clicking **View Model**.

Figure 32 - Self-Learning Model


TDSS
Assemble
Upload Data
Explore
View Data
Transform
Prepare Data
Model
Create Models
Pipeline
Prepare Workflow
dev testing
Add Model Unit
Add Self Learning Model Unit
Show only self-learning models
ns
NS_SCREENING Self Learning
Description:
clnslf Self Learning
Description: new_slf self learning
AFE_unsup_tree_model
Description: for unsupervised pipeline
NS_CORPO_SL Self Learning
Description:
NS_INDI_SL Self Learning
Description: ND_INDI self learning
ns_indi_2020 Self Learning
Description:
View Models

In Model Evolution tab, it displays the Champion-Challenger outcome. It shows the current champion and challenger. User can view the predictions, matrix and the result.

Figure 33 - Model Evolution



Model Unit

All Model Units / Model Unit : NS_INDIVID_1.1.4 Self Learning

Oct 10, 2019
9:18 AM

Initial Model : NS_INDIVIDUAL_1.1.4

Champion		Challenger	
Model	NS_INDIVIDUAL_1.1.4	Model	NS_INDIVIDUAL_1.1.4_19904
Evaluation Scenario	View prediction	Evaluation Scenario	View prediction
Evaluation Metric	**False Alert Reduction Rate** - 1	Evaluation Metric	**False Alert Reduction Rate** - 0.2542372941970825

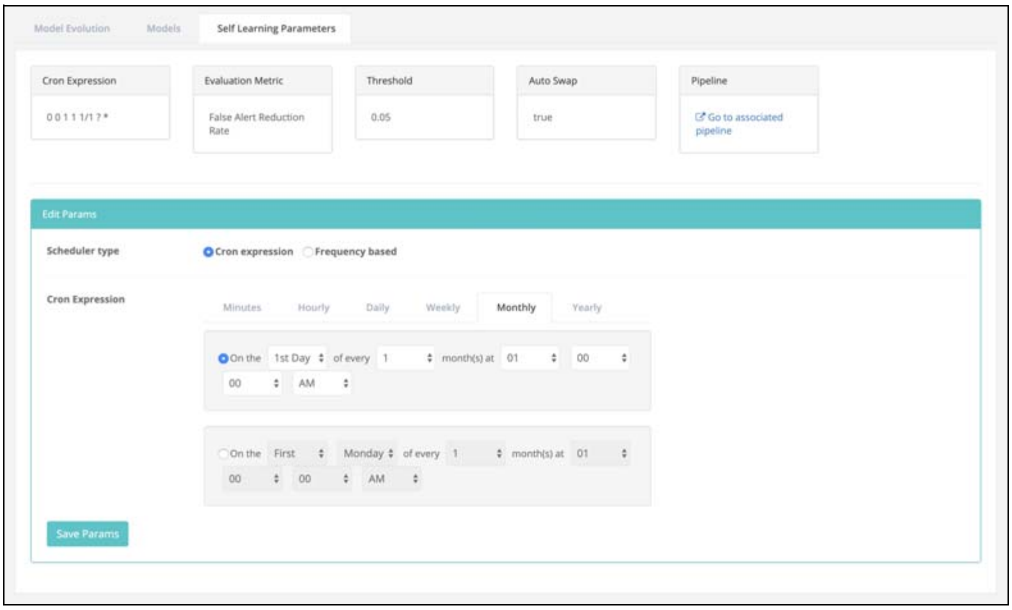
Difference : -74.5763%

Result : Champion not changed

Self Learning Parameters

In the Self Learning Parameters tab, it allows the user to review the self-learning parameters. The boxes on the top are the current parameters used and the bottom portion allows users to edit the parameters accordingly.

Figure 34 - Self Learning Parameters



Cron Expression: 0 0 1 1 1/1 ? *

Evaluation Metric: False Alert Reduction Rate

Threshold: 0.05

Auto Swap: true

Pipeline: Go to associated pipeline

Scheduler type: ● Cron expression ○ Frequency based

Cron Expression:

● On the 1st Day of every 1 month(s) at 01 00 00 AM

○ On the First Monday of every 1 month(s) at 01 00 00 AM

The parameters in the box are:

Table 8 - Self Learning Parameters

Item	**Description**
Frequency	Frequency of a challenger created and tested against the champion (current model scenario).
Cron Expression	Input of the Cron expression.
Start Time	The date and time when processing of the model last began.
Evaluation Metric	The attribute used to determine the champion.
Threshold	The minimum value that a challenger must score better to replace the current champion.
Auto Swap	Auto swap between the champion and challenger. Can be set to True or False depending on whether the client wants automatic or manual switching between champion and challenger.
Pipeline	Option to view the associated pipeline.

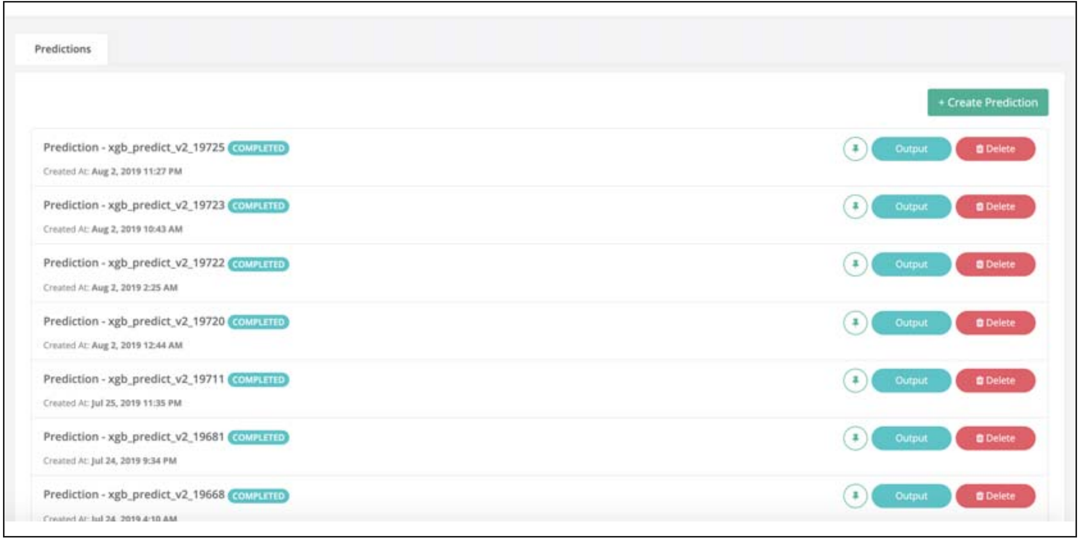
Prediction Pipeline for Names Screening

Prediction Historical Record

In TDSS, go to **Dashboard** > **Model Unit** > **Models** > **View Predictions**.

The user is able to view the list of prediction history for the specific model chosen. The user is also able to view the output of each prediction or delete it as shown in the example below.

Figure 35 – Predictions History



Prediction - xgb_predict_v2_19725 COMPLETED
Created At: Aug 2, 2019 11:27 PM

Prediction - xgb_predict_v2_19723 COMPLETED
Created At: Aug 2, 2019 10:43 AM

Prediction - xgb_predict_v2_19722 COMPLETED
Created At: Aug 2, 2019 2:25 AM

Prediction - xgb_predict_v2_19720 COMPLETED
Created At: Aug 2, 2019 12:44 AM

Prediction - xgb_predict_v2_19711 COMPLETED
Created At: Jul 25, 2019 11:35 PM

Prediction - xgb_predict_v2_19681 COMPLETED
Created At: Jul 24, 2019 9:34 PM

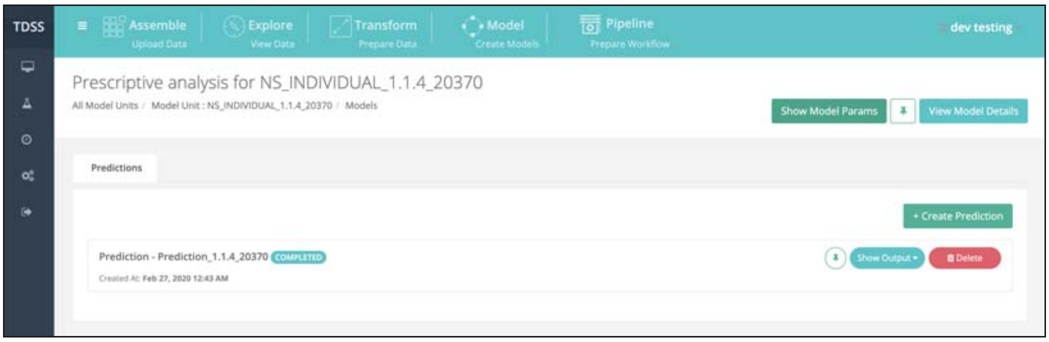
Prediction - xgb_predict_v2_19668 COMPLETED
Created At: Jul 24, 2019 4:10 AM

Feature Contribution and Prediction Explainability

In TDSS, go to **Dashboard** > **Model Unit** > **Models** > **View Predictions**.

From the list of predictions, user can click **Output** to view the prediction output and its explainability.

Figure 36 – Predictions



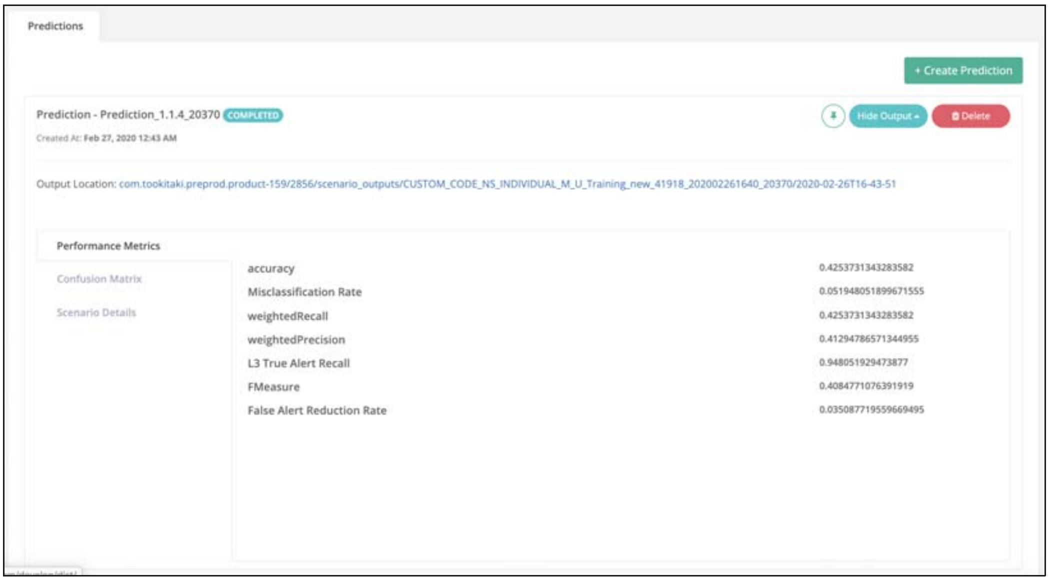
Prescriptive analysis for NS_INDIVIDUAL_1.1.4_20370

All Model Units / Model Unit : NS_INDIVIDUAL_1.1.4_20370 / Models

Prediction - Prediction_1.1.4_20370 COMPLETED
Created At: Feb 27, 2020 12:43 AM

Click **Show Output** to view the prediction output details such as the Output Location, Performance Metrics and Confusion Matrix similarly shown above.

Figure 37 - Output



Prediction - Prediction_1.1.4_20370

Created At: Feb 27, 2020 12:43 AM

Output Location: com.tookitaki.preprod.product-159/2856/scenario_outputs/CUSTOM_CODE_NS_INDIVIDUAL_M_U_Training_new_41918_202002261640_20370/2020-02-26T16-43-51

accuracy	0.4253731343283582
Misclassification Rate	0.051948051899671555
weightedRecall	0.4253731343283582
weightedPrecision	0.41294786571344955
L3 True Alert Recall	0.948051929473877
FMeasure	0.4084771076391919
False Alert Reduction Rate	0.035087719559669495

Output Location

The Output Location link can be provided in different components such as models, predictions, pipeline components, transformation and more.

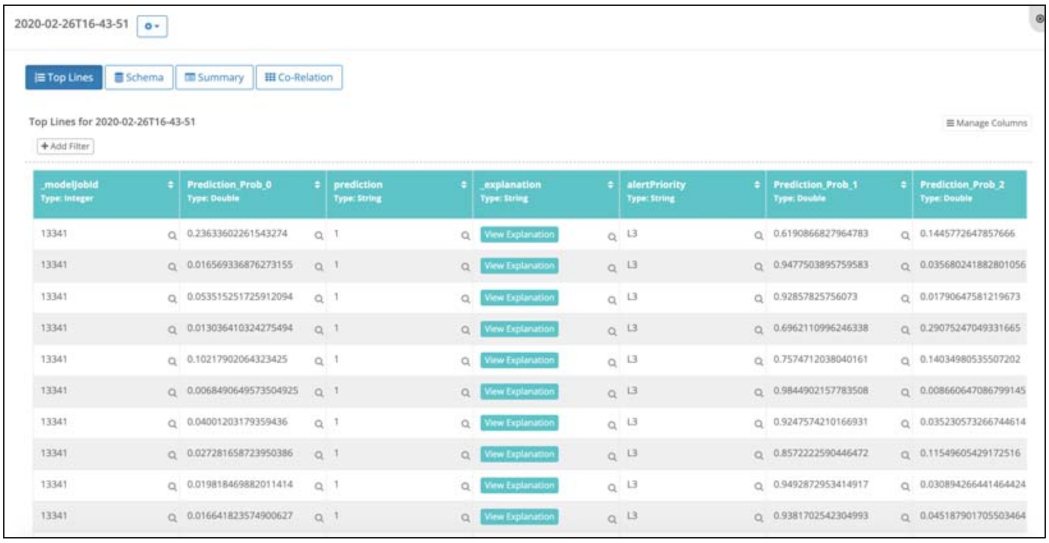
From the Output Location link provided in the prediction output, user is redirected to view the output and feature contributions according to the tabs available:

- Top Lines
- Schema

Top Lines

User can analyze the output in detail and also view explanation in the Top Lines. The Top Lines page displays the features and their contributions to the prediction. User can also scroll down to the bottom to view the Explanation and Feature Details.

Figure 38 - Top Lines



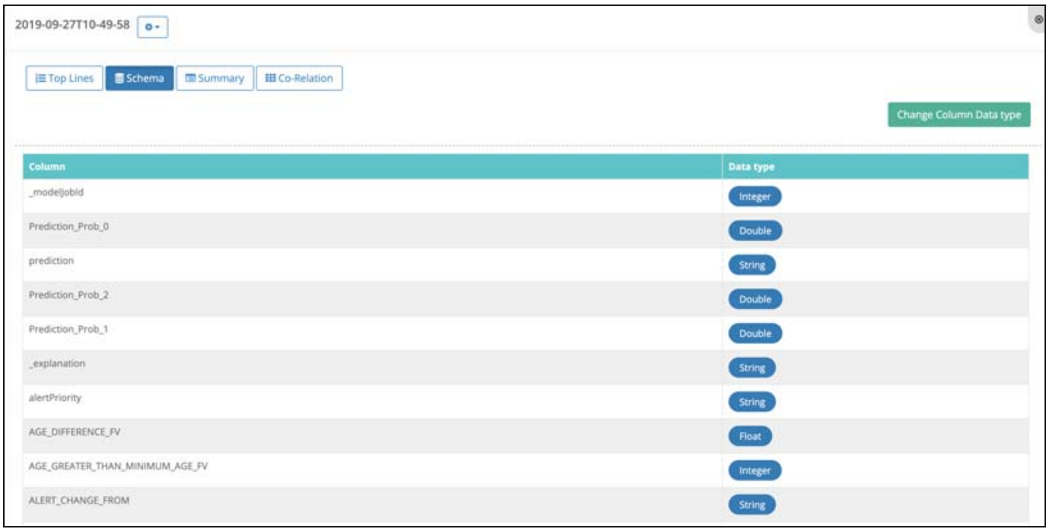
Top Lines for 2020-02-26T16-43-51

_modeljobId Type: Integer	Prediction_Prob_0 Type: Double	prediction Type: String	_explanation Type: String	alertPriority Type: String	Prediction_Prob_1 Type: Double	Prediction_Prob_2 Type: Double
13341	0.23633602261543274	1	View Explanation	L3	0.6190866827964783	0.1445772647857666
13341	0.016569336876273155	1	View Explanation	L3	0.9477503895759583	0.035680241882801056
13341	0.053515251725912094	1	View Explanation	L3	0.92857825756073	0.017906475812 19673
13341	0.013036410324275494	1	View Explanation	L3	0.6962110996246338	0.29075247049331665
13341	0.10217902064323425	1	View Explanation	L3	0.7574712038040161	0.14034980535507202
13341	0.0068490649573504925	1	View Explanation	L3	0.9844902157783508	0.008660647086799145
13341	0.04001203179359436	1	View Explanation	L3	0.9247574210166931	0.035230573266744614
13341	0.027281658723950386	1	View Explanation	L3	0.8572222590446472	0.11549605429172516
13341	0.019818469882011414	1	View Explanation	L3	0.9492872953414917	0.030894266441464424
13341	0.016641823574900627	1	View Explanation	L3	0.9381702542304993	0.045187901705503464

Schema

The Schema section contains column names and the data type of each of the columns.

Figure 39 - Schema

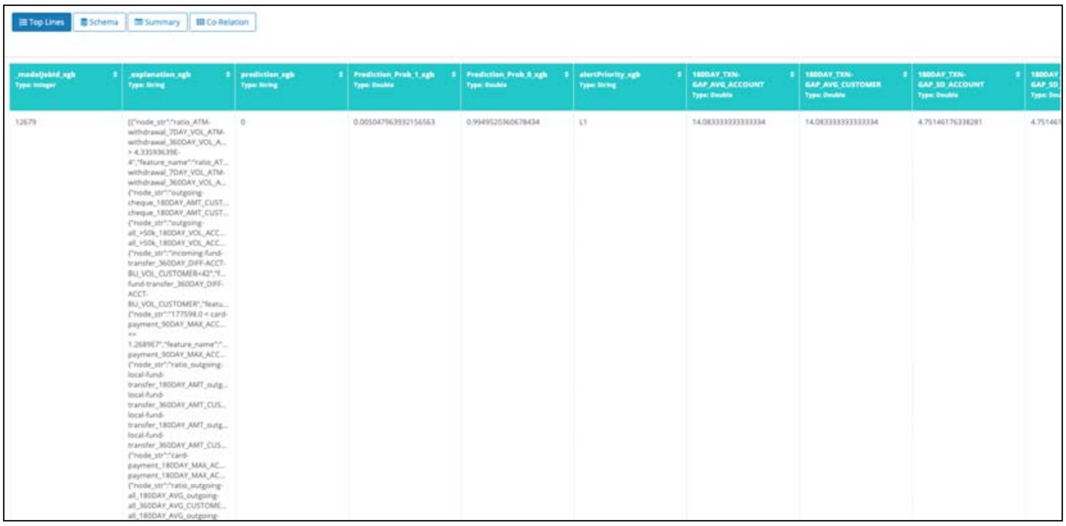


Column	Data type
_modeljobId	Integer
Prediction_Prob_0	Double
prediction	String
Prediction_Prob_2	Double
Prediction_Prob_1	Double
_explanation	String
alertPriority	String
AGE_DIFFERENCE_FV	Float
AGE_GREATER_THAN_MINIMUM_AGE_FV	Integer
ALERT_CHANGE_FROM	String

Explainability

From the Top Lines page, user can click **View Explanation** to view explanation in detail as shown below.

Figure 40 - View Explanation

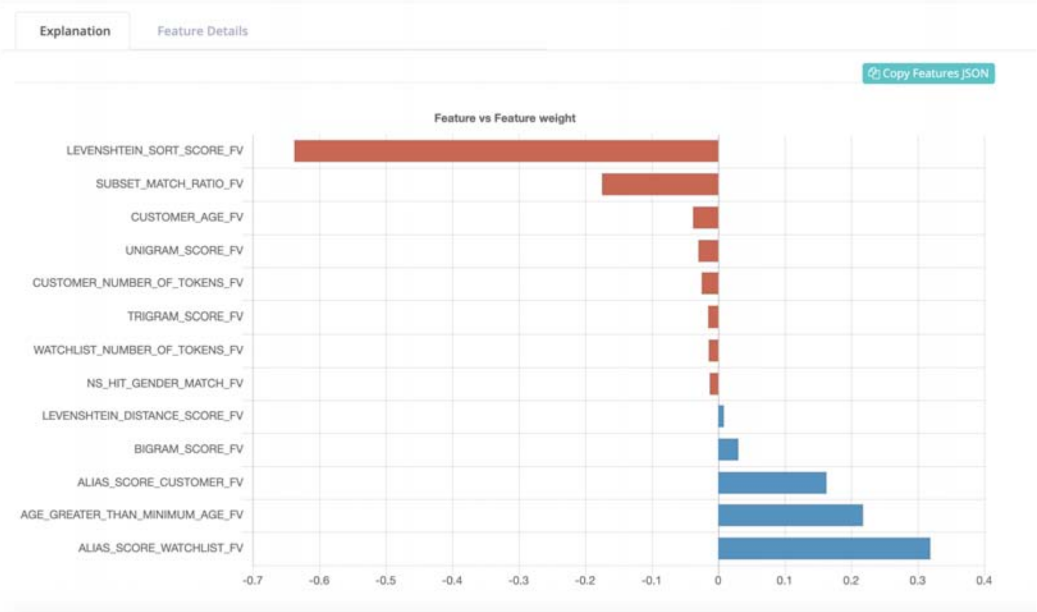


_modeljobid_xgb Type: Integer	_explanation_xgb Type: String	prediction_xgb Type: String	Prediction_Prob_1_xgb Type: Double	Prediction_Prob_0_xgb Type: Double	alertPriority_xgb Type: String	180DAY_TXN-GAP_AVG_ACCOUNT Type: Double	180DAY_TXN-GAP_AVG_CUSTOMER Type: Double	180DAY_TXN-GAP_SD_ACCOUNT Type: Double	180DAY_ GAP_SD_ Type: Dou
12679	[{"node_str":"ratio_ATM-withdrawal_7DAY_VOL_ATM-withdrawal_360DAY_VOL_A... > 4.33593639E-4","feature_name":"ratio_AT... withdrawal_7DAY_VOL_ATM-withdrawal_360DAY_VOL_A... {"node_str":"outgoing-cheque_180DAY_AMT_CUST... cheque_180DAY_AMT_CUST... {"node_str":"outgoing-all_>50k_180DAY_VOL_ACC... all_>50k_180DAY_VOL_ACC... {"node_str":"incoming-fund-transfer_360DAY_DIFF-ACCT-BU_VOL_CUSTOMER<42","f... fund-transfer_360DAY_DIFF-ACCT-BU_VOL_CUSTOMER","featu... {"node_str":"177598.0 < card-payment_90DAY_MAX_ACC... <= 1.2689E7","feature_name":"... payment_90DAY_MAX_ACC... {"node_str":"ratio_outgoing-local-fund-transfer_180DAY_AMT_outg... local-fund-transfer_360DAY_AMT_CUS... local-fund-transfer_180DAY_AMT_outg... local-fund-transfer_360DAY_AMT_CUS... {"node_str":"card-payment_180DAY_MAX_AC... payment_180DAY_MAX_AC... {"node_str":"ratio_outgoing-all_180DAY_AVG_outgoing-all_360DAY_AVG_CUSTOME... all_180DAY_AVG_outgoing-	0	0.00504796393215656 3	0.994952036067843 4	L1	14.083333333333334	14.083333333333334	4.75146176338281	4.751461

Scroll to the bottom to view the explanation based on the features versus the feature weight as shown below. This is specific to only certain models: linear and logistic regression.

This information is not specific to any particular prediction outcome, but to the whole model.

Figure 41 - Explanation



NS Pipeline Setup and Execution

Prerequisites

Perform Data Validation

1 Connect to the system using Putty.

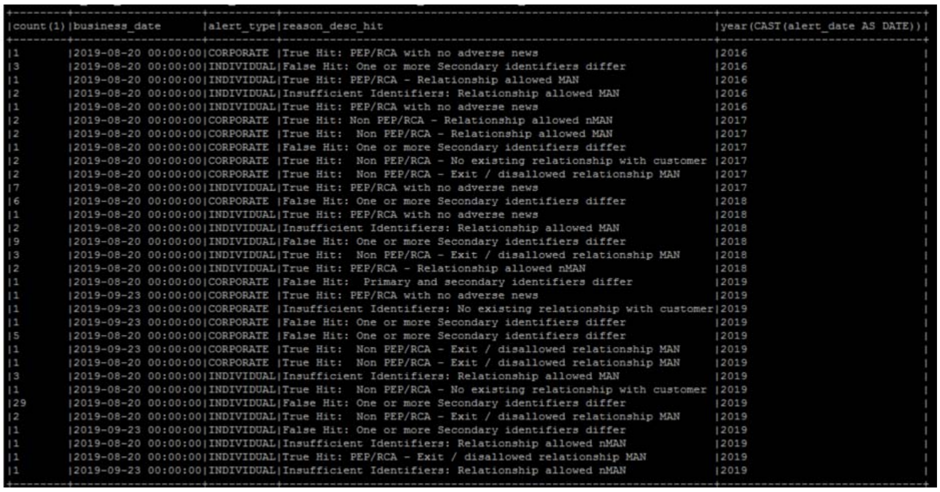
2 Check the following pre-pipeline tables exist with suitable data:

- NS_ALERTS
- NS_DICT
- NS_CUSTOMER
- NS_C2C
- NS_WATCHLIST

3 Run the ns_data_validation script to check if data is suitable for model and save the output in a file for observation.

4 Critical checks to perform in the validation script output:

- Observe the training and validation period and ensure minimum 10 records are present in the Training set (2.5 years) and minimum 1 record is present in the validation set (3 months) for each category (True/ False/ Insufficient) in NS_ALERTS table.
- Ensure that the training set is chosen such that it has at least one record from each of the categories: True/False/Insufficient referring to the labels from the 'reason_desc_hit' field. This check is required to be performed for both the alert types - INDIVIDUAL and CORPORATE.



```
+--------+-------------------+----------+----------------------------------------------------------------+------------------------------+
|count(1)|business_date      |alert_type|reason_desc_hit                                                 |year(CAST(alert_date AS DATE))|
+--------+-------------------+----------+----------------------------------------------------------------+------------------------------+
|1       |2019-08-20 00:00:00|CORPORATE |True Hit: PEP/RCA with no adverse news                          |2016                          |
|3       |2019-08-20 00:00:00|INDIVIDUAL|False Hit: One or more Secondary identifiers differ             |2016                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA - Relationship allowed MAN                    |2016                          |
|2       |2019-08-20 00:00:00|INDIVIDUAL|Insufficient Identifiers: Relationship allowed MAN              |2016                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA with no adverse news                          |2016                          |
|2       |2019-08-20 00:00:00|CORPORATE |True Hit: Non PEP/RCA - Relationship allowed nMAN               |2017                          |
|2       |2019-08-20 00:00:00|CORPORATE |True Hit:  Non PEP/RCA - Relationship allowed MAN               |2017                          |
|1       |2019-08-20 00:00:00|CORPORATE |False Hit: One or more Secondary identifiers differ             |2017                          |
|2       |2019-08-20 00:00:00|CORPORATE |True Hit:  Non PEP/RCA - No existing relationship with customer |2017                          |
|2       |2019-08-20 00:00:00|CORPORATE |True Hit:  Non PEP/RCA - Exit / disallowed relationship MAN     |2017                          |
|7       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA with no adverse news                          |2017                          |
|6       |2019-08-20 00:00:00|CORPORATE |False Hit: One or more Secondary identifiers differ             |2018                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA with no adverse news                          |2018                          |
|2       |2019-08-20 00:00:00|INDIVIDUAL|Insufficient Identifiers: Relationship allowed MAN              |2018                          |
|9       |2019-08-20 00:00:00|INDIVIDUAL|False Hit: One or more Secondary identifiers differ             |2018                          |
|3       |2019-08-20 00:00:00|INDIVIDUAL|True Hit:  Non PEP/RCA - Exit / disallowed relationship MAN     |2018                          |
|2       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA - Relationship allowed nMAN                   |2018                          |
|1       |2019-08-20 00:00:00|CORPORATE |False Hit:  Primary and secondary identifiers differ            |2019                          |
|1       |2019-09-23 00:00:00|CORPORATE |True Hit: PEP/RCA with no adverse news                          |2019                          |
|1       |2019-09-23 00:00:00|CORPORATE |Insufficient Identifiers: No existing relationship with customer|2019                          |
|1       |2019-09-23 00:00:00|CORPORATE |False Hit: One or more Secondary identifiers differ             |2019                          |
|5       |2019-08-20 00:00:00|CORPORATE |False Hit: One or more Secondary identifiers differ             |2019                          |
|1       |2019-09-23 00:00:00|CORPORATE |True Hit:  Non PEP/RCA - Exit / disallowed relationship MAN     |2019                          |
|1       |2019-08-20 00:00:00|CORPORATE |True Hit:  Non PEP/RCA - Exit / disallowed relationship MAN     |2019                          |
|3       |2019-08-20 00:00:00|INDIVIDUAL|Insufficient Identifiers: Relationship allowed MAN              |2019                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|True Hit:  Non PEP/RCA - No existing relationship with customer |2019                          |
|29      |2019-08-20 00:00:00|INDIVIDUAL|False Hit: One or more Secondary identifiers differ             |2019                          |
|2       |2019-08-20 00:00:00|INDIVIDUAL|True Hit:  Non PEP/RCA - Exit / disallowed relationship MAN     |2019                          |
|1       |2019-09-23 00:00:00|INDIVIDUAL|False Hit: One or more Secondary identifiers differ             |2019                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|Insufficient Identifiers: Relationship allowed nMAN             |2019                          |
|1       |2019-08-20 00:00:00|INDIVIDUAL|True Hit: PEP/RCA - Exit / disallowed relationship MAN          |2019                          |
|1       |2019-09-23 00:00:00|INDIVIDUAL|Insufficient Identifiers: Relationship allowed nMAN             |2019                          |
+--------+-------------------+----------+----------------------------------------------------------------+------------------------------+
```

NS Pipelines

The following is the list of NS pipelines:

1. T_V_ NS Individual Alert Prioritization
2. T_V_ NS Corporate Alert Prioritization
3. P_ NS Individual Alert Prioritization
4. P_ NS Corporate Alert Prioritization
5. NS DELTA

Check post-pipeline HBASE tables:

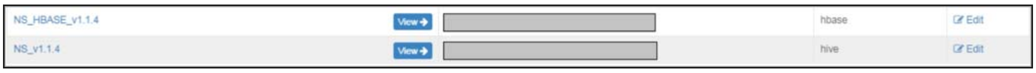
1. Connect to HBASE shell.
2. Ensure that all below post-pipeline tables are present:
   - NS_HITS
   - NS_CLUSTER_FEATURES
   - NS_ERROR_CODES
   - NS_MISSING_ALERT
   - NS_ADDITIONAL_FEATURES

Persister and Connector Settings

Note: Below step is already automated. If any discrepancies are found, below steps need to be followed.

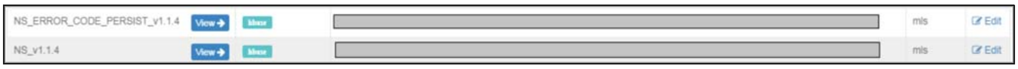
Once new pipelines and connectors are in place, the default factory settings have to be changed.

1. Go to connectors - change the database name and the destination URI appropriately.



NS_HBASE_v1.1.4	View →	[redacted]	hbase	Edit
NS_v1.1.4	View →	[redacted]	hive	Edit

2. Go to persisters - change the database name and the destination URI appropriately.



NS_ERROR_CODE_PERSIST_v1.1.4	View →	hbase	[redacted]	mls	Edit
NS_v1.1.4	View →	hbase	[redacted]	mls	Edit

Get details about database name, source URI, destination URI from the implementation engineer.

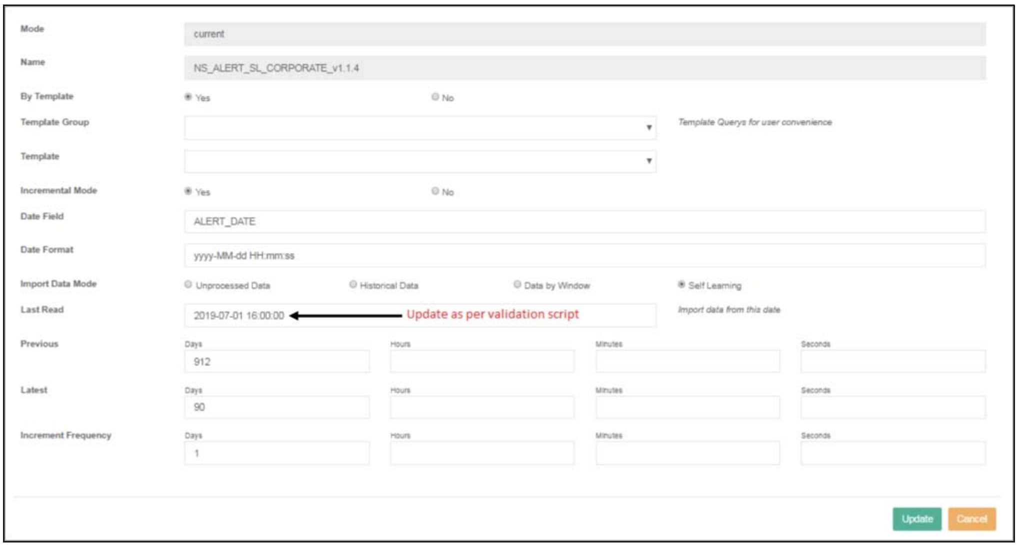
For connectors, Hive database name is **amls**, and for persisters, HBase database name is **mls** across all environments.

Modify Connector Settings for Self-Learning

To ensure the model gets sufficient data to train and validate the model, it is sometimes imperative to change the LAST_READ field in self-learning connector.

1. NS (current version name) /*_SL (for self_learning)
2. Look for connectors for both individual and corporate (NS_ALERT_SL for individual and NS_ALERT_SL_CORPORATE for corporate)
3. Data needs to be carefully chosen by running ns_validation_script for the ALERT_DATE

4 Date value should be selected appropriately to ensure the data for the training should be more than 900 records and the validation should be more than 90 records.



Mode: current

Name: NS_ALERT_SL_CORPORATE_v1.1.4

By Template: ● Yes ○ No

Template Group: ▾ *Template Querys for user convenience*

Template: ▾

Incremental Mode: ● Yes ○ No

Date Field: ALERT_DATE

Date Format: yyyy-MM-dd HH:mm:ss

Import Data Mode: ○ Unprocessed Data ○ Historical Data ○ Data by Window ● Self Learning

Last Read: 2019-07-01 16:00:00 ← Update as per validation script *Import data from this date*

	Days	Hours	Minutes	Seconds
Previous	912			
Latest	90			
Increment Frequency	1			

Add Self-Learning Model

T_V NS Corporate Alert Prioritization

1 Go to Models > Add Self learning Model Unit.

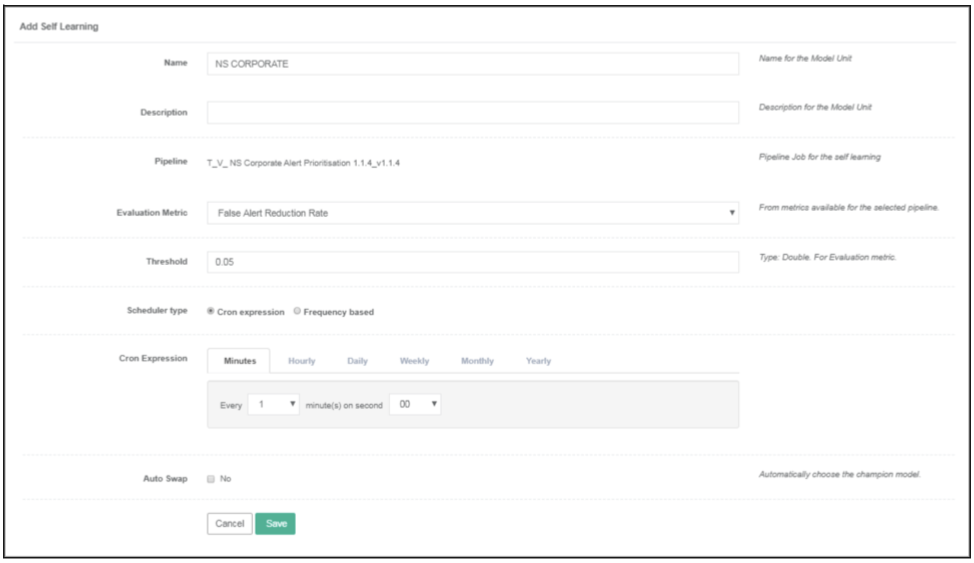
2 Add the required details in the landing page. Select the pipeline (T_V NS Corporate Alert Prioritization) > Evaluation Metric as False Alert Reduction Rate (Value: 0.05).

3 Cron expression - Yearly.

4 Select Auto Swap option.

5 Click **Save.**

6 Click **Run Self -learning Unit.**



Add Self Learning

Name: NS CORPORATE *Name for the Model Unit*

Description: *Description for the Model Unit*

Pipeline: T_V_ NS Corporate Alert Prioritisation 1.1.4_v1.1.4 *Pipeline Job for the self learning*

Evaluation Metric: False Alert Reduction Rate ▾ *From metrics available for the selected pipeline.*

Threshold: 0.05 *Type: Double. For Evaluation metric.*

Scheduler type: ● Cron expression ○ Frequency based

Cron Expression: Minutes | Hourly | Daily | Weekly | Monthly | Yearly

Every 1 ▾ minute(s) on second 00 ▾

Auto Swap: □ No *Automatically choose the champion model.*

T_V NS Individual Alert Prioritization

1 Go to Models > Add Self learning Model Unit.

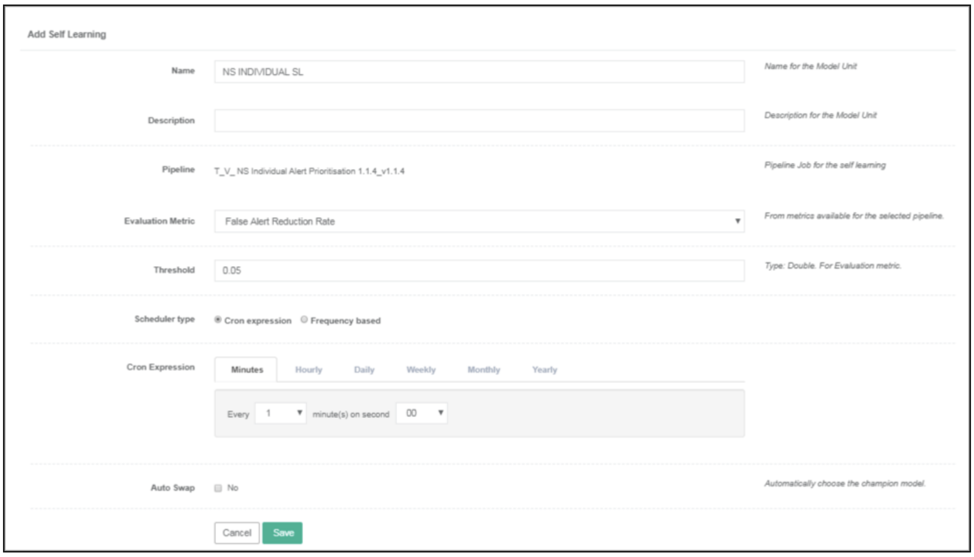
2 Add the required details in the landing page. Select the pipeline (T_V NS Individual Alert Prioritization) > Evaluation Metric as False Alert Reduction Rate (Value: 0.05).

3 Cron expression - Yearly.

4 Select Auto Swap option.

5 Click **Save**.

6 Click **Run Self -learning Unit.**



Add Self Learning

Name: NS INDIVIDUAL SL — *Name for the Model Unit*

Description: — *Description for the Model Unit*

Pipeline: T_V_ NS Individual Alert Prioritisation 1.1.4_v1.1.4 — *Pipeline Job for the self learning*

Evaluation Metric: False Alert Reduction Rate ▾ — *From metrics available for the selected pipeline.*

Threshold: 0.05 — *Type: Double. For Evaluation metric.*

Scheduler type: ● Cron expression ○ Frequency based

Cron Expression: Minutes | Hourly | Daily | Weekly | Monthly | Yearly

Every 1 ▾ minute(s) on second 00 ▾

Auto Swap: □ No — *Automatically choose the champion model.*

Prediction Pipeline for Alert Prioritization

This modification is required when prediction pipelines need to pick the deltas only and not the complete data from Hive. This comes as a part of incremental mode and would be taken care of automatically but for any customized data input requirements, this needs to be configured.

Note: Unprocessed Data gives all the data records greater than the cut off date provided as per the below screenshot. Hence, (TT_CREATED_TIME should be used as it is the load date maintained in TT AMLS).

1 Go to connectors > NS (version name to be present as postfix) / NS_INCREMENTAL_*.

2 Look for connectors for both Individual and Corporate - NS_INCREMENTAL_INDIVIDUAL and NS_INCREMENTAL_CORPORATE.

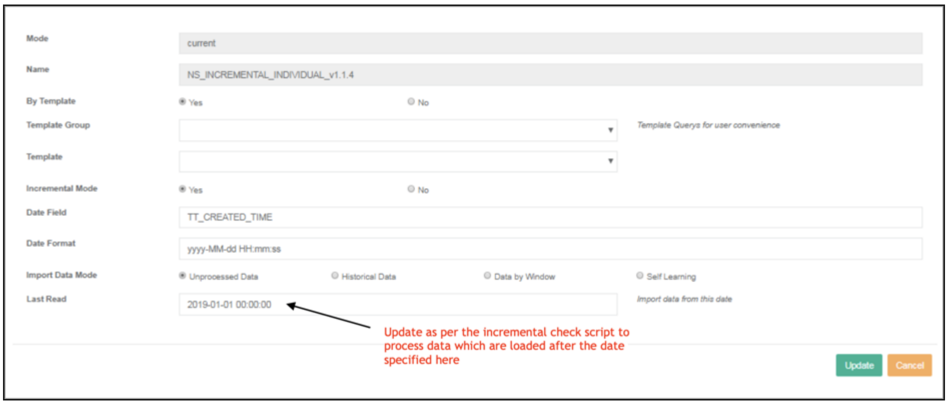
3 Data needs to be carefully chosen by running incremental_check_script for the TT_CREATED_TIME.

Incremental Check Script

1 Select distinct TT_CREATED_TIME from amls.ns_alerts.

Note: Pick the latest “TT_CREATED_TIME” and update the “LAST_READ_DATE” with the same value. In case of any customized requirement (predicting last two delta’s), the last_read_date has to be shifted/modified accordingly (for last two delta’s pick the date just before last two delta’s)

2. In order to select and run predictions only the delta load, the TT_CREATED_TIME date value should be provided properly by selecting the 'Unprocessed Data' mode.



Mode: current

Name: NS_INCREMENTAL_INDIVIDUAL_v1.1.4

By Template: ● Yes ○ No

Template Group: ▾ *Template Querys for user convenience*

Template: ▾

Incremental Mode: ● Yes ○ No

Date Field: TT_CREATED_TIME

Date Format: yyyy-MM-dd HH:mm:ss

Import Data Mode: ● Unprocessed Data ○ Historical Data ○ Data by Window ○ Self Learning

Last Read: 2019-01-01 00:00:00 *Import data from this date*

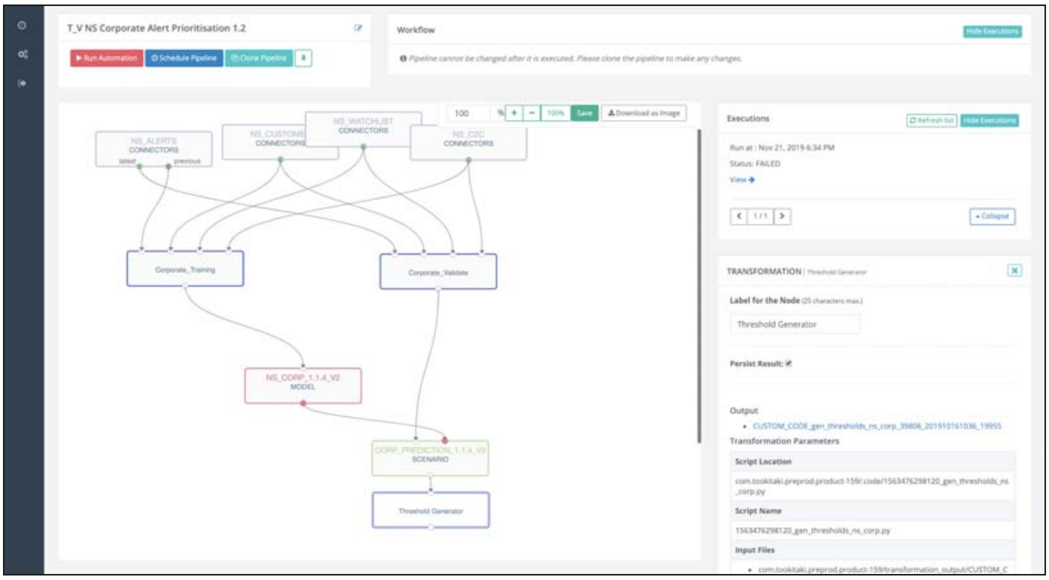
Update as per the incremental check script to process data which are loaded after the date specified here

3. Update the same for all the other connectors which have '_INCREMENTAL_'.

Prediction Pipeline Thresholds

Use the output of Threshold Generator component in the NS Training pipeline to get the L1 and L2 threshold values.

1. Go to the training pipeline and click the Threshold Generator component.
2. Click the output on the side panel.



T_V NS Corporate Alert Prioritisation 1.2

Workflow

Pipeline cannot be changed after it is executed. Please clone the pipeline to make any changes.

Executions

Run at : Nov 21, 2019 6:34 PM

Status: FAILED

TRANSFORMATION | Threshold Generator

Label for the Node (25 characters max.): Threshold Generator

Persist Result: ☑

Output

- CUSTOM_CODE_gen_thresholds_ns_corp_39806_201910161036_19955

Transformation Parameters

Script Location

com.tookitaki.preprod.product-159/.code/1563476298120_gen_thresholds_ns_corp.py

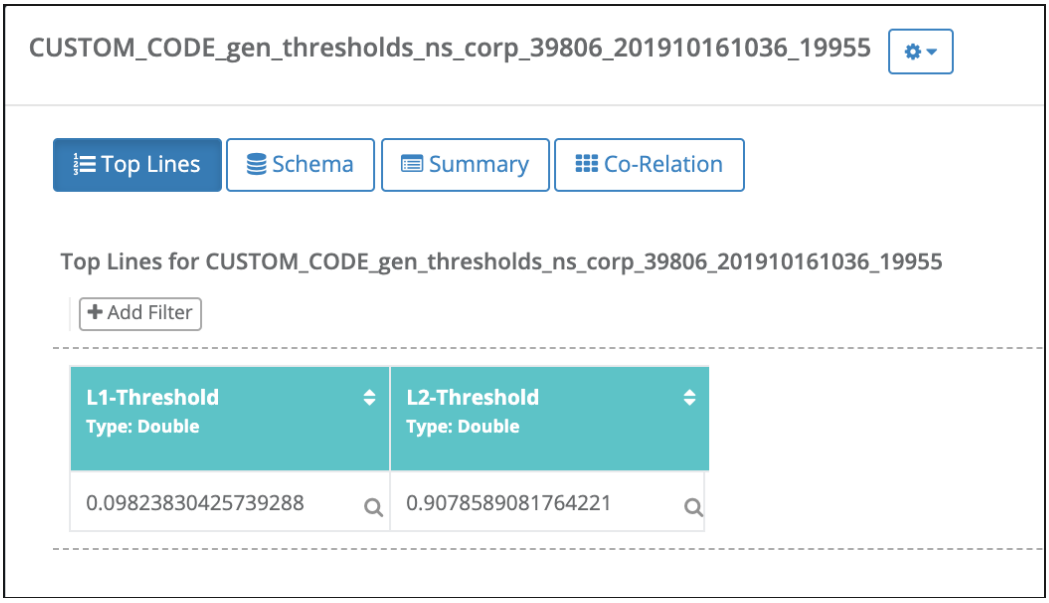
Script Name

1563476298120_gen_thresholds_ns_corp.py

Input Files

- com.tookitaki.preprod.product-159/transformation_output/CUSTOM_C

3 The L1 and L2 threshold values are displayed.

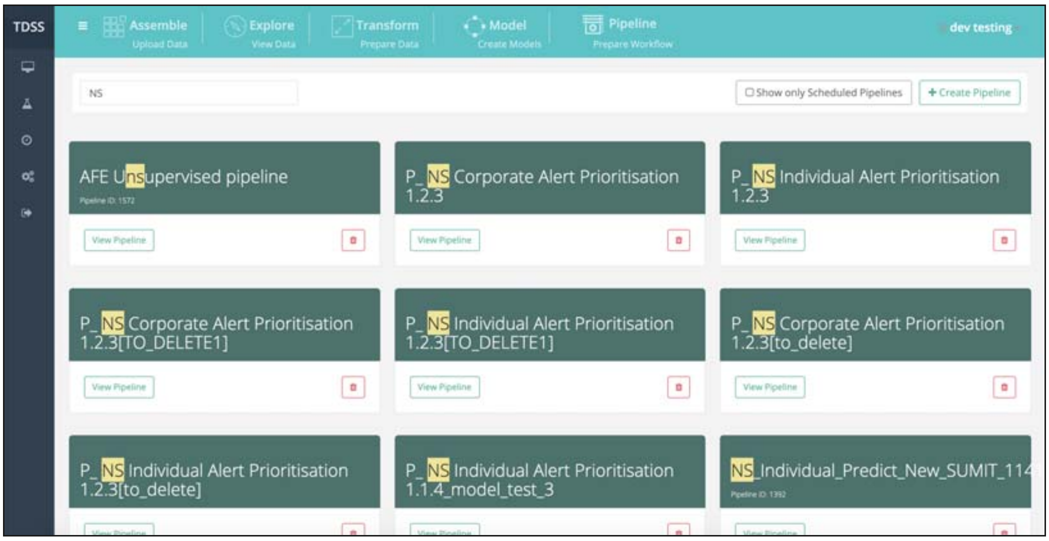


CUSTOM_CODE_gen_thresholds_ns_corp_39806_201910161036_19955

Top Lines for CUSTOM_CODE_gen_thresholds_ns_corp_39806_201910161036_19955

L1-Threshold Type: Double	L2-Threshold Type: Double
0.09823830425739288	0.9078589081764221

4 Go to **Dashboard** > **Pipelines**.



AFE Unsupervised pipeline
Pipeline ID: 1572

P_NS Corporate Alert Prioritisation 1.2.3

P_NS Individual Alert Prioritisation 1.2.3

P_NS Corporate Alert Prioritisation 1.2.3[TO_DELETE1]

P_NS Individual Alert Prioritisation 1.2.3[TO_DELETE1]

P_NS Corporate Alert Prioritisation 1.2.3[to_delete]

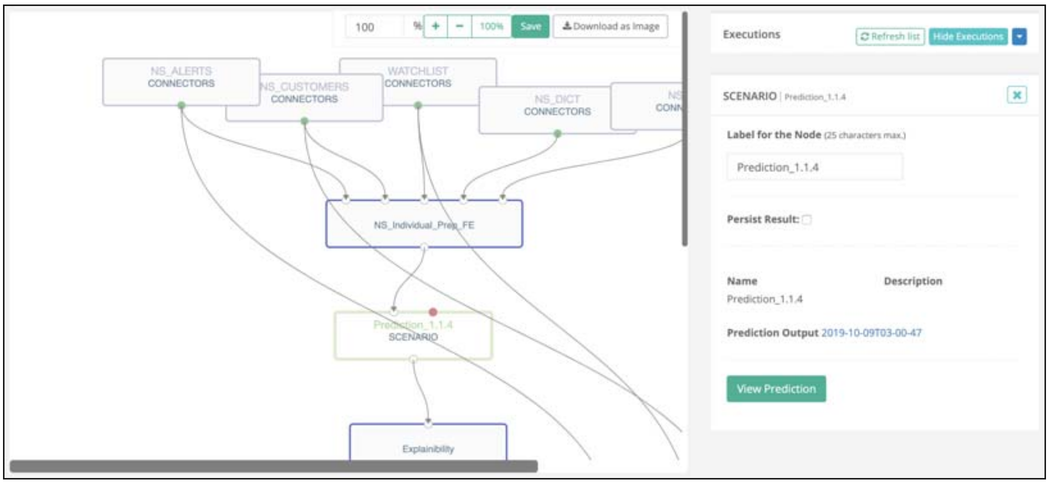
P_NS Individual Alert Prioritisation 1.2.3[to_delete]

P_NS Individual Alert Prioritisation 1.1.4_model_test_3

NS_Individual_Predict_New_SUMIT_114
Pipeline ID: 1392

5 Go to the prediction pipeline for alerts prioritization. For example, NS prediction pipeline is P NS Alert Prioritization<version_number> and click **View Pipeline**.

6 Click the Scenario component and click **View Prediction** on the side panel.



SCENARIO | Prediction_1.1.4

Label for the Node (25 characters max.)

Prediction_1.1.4

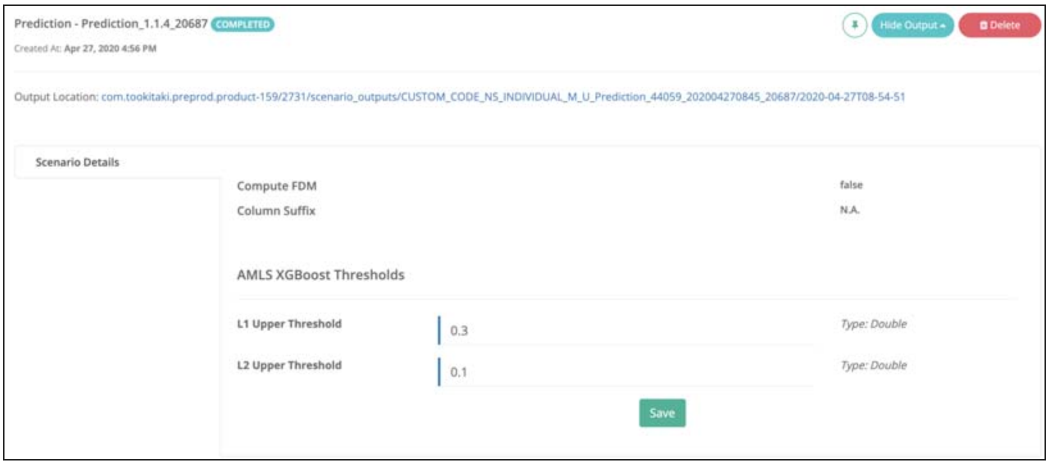
Persist Result: □

Name	Description
Prediction_1.1.4	

Prediction Output 2019-10-09T03-00-47

7 In the Predictions page, find the desired prediction and click **Show Output** to expand it.

8 Select the **Scenario Details** tab and enter the L1 and L2 threshold values from the steps above.



Prediction - Prediction_1.1.4_20687 COMPLETED

Created At: Apr 27, 2020 4:56 PM

Output Location: com.tookitaki.preprod.product-159/2731/scenario_outputs/CUSTOM_CODE_NS_INDIVIDUAL_M_U_Prediction_44059_202004270845_20687/2020-04-27T08-54-51

Scenario Details

Compute FDM	false
Column Suffix	N.A.

AMLS XGBoost Thresholds

L1 Upper Threshold	0.3	Type: Double
L2 Upper Threshold	0.1	Type: Double

9 Click **Save**.

Finally, the pipeline needs to be triggered by the following steps.

10 Go to Pipeline.

11 Select the prediction pipeline.

12 Click **Run Automation**.

Common Errors

Connectors

1. Connection failure - generic connection issue.
2. Fetches empty data - date setup is not done properly. Please verify the cut off date provided in the connector and check against the validation script.

Persisters

1. Index build failure.
2. Data type mismatch.

UDF

1. Module missing - the python module is not imported into TDSS. Ensure you can see the python module in the TDSS dynamic properties.
2. Mismatch inputs - input order is not provided properly.

Models

1. Not sufficient data for training.
2. Model not found - Clone and replace with appropriate model

Scenario

1. Scenario not found - Clone and replace with appropriate scenario.

Name Screening Module

Name Screening Module Functionalities

Alerts Investigation View

The Alerts Investigation View lists all alerts extracted from the existing NS system with the secondary scoring results. The alerts are prioritized into bucket levels of L1, L2 and L3. It also enables the user to navigate to the alert details page for further investigation purposes.

The Alerts Investigation View can be accessed by clicking **Alerts Investigation** on the side panel. See "Navigation and Filters".

Risk View

The Risk view provides detailed insights into the statistics of alerts, hits and break down of alerts by segments. It displays values for both the current period and the previous period as defined by the selected preset date or the custom date range.

The statistics are displayed on widgets categorized as:

- Alerts & Hits
- Hits Distribution
- Business Segments & Parameters
- Customer View for All Segments

The Risk View can be accessed by clicking **Risk View** on the side panel. See "Navigation and Filters".

Navigation and Filters

Icons

The following icons are used in the application:

Table 9 - Navigation and Filters

Icons	Description
🏠	Access to **Risk View**
👁	Access to **Alerts Investigation**
▼	Filter
▼✓	Filter Applied
⊞	Grid View
📊	Graphical View
☰	Tabular View
—	Minimize Window
▭	Maximize Window
⌄	Expand Details
⌃	Collapse Details

Alerts Investigation View

Go to **Side Panel** > **Alerts Investigation View**.

The Alerts Investigation View provides an overview of all prioritized NS alerts and their corresponding information at the customer level.

The following are the main features available:

1 Time Range Selection

2 Alert Status Filter

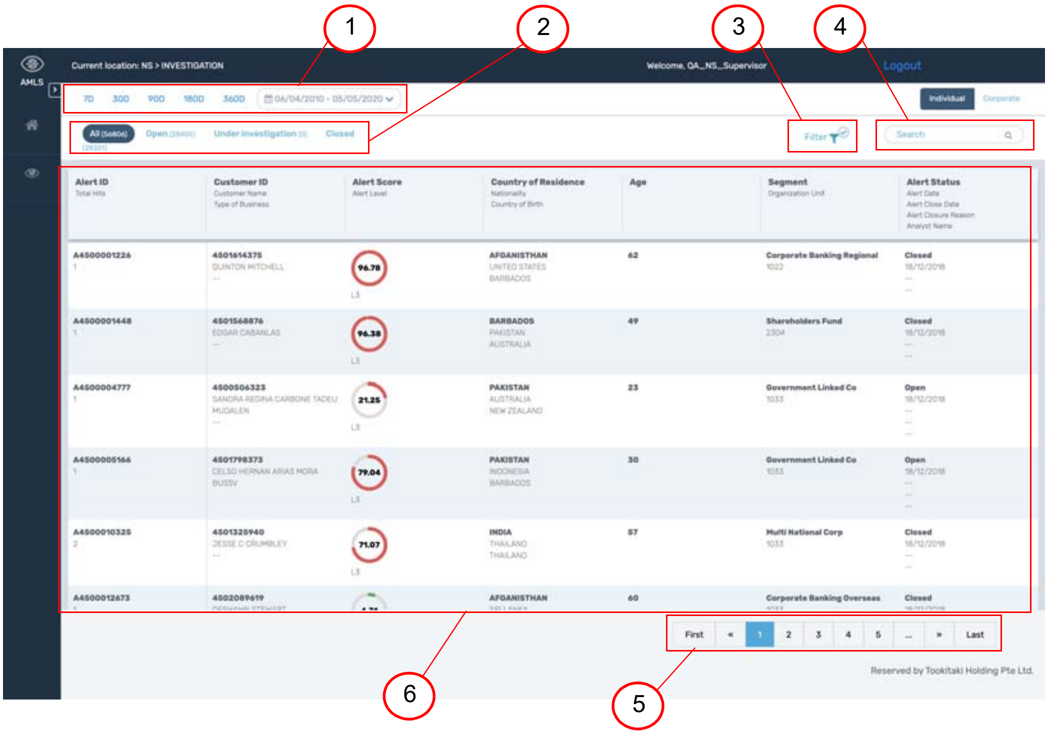
3 Column Filter

4 Search Bar

5 Pagination

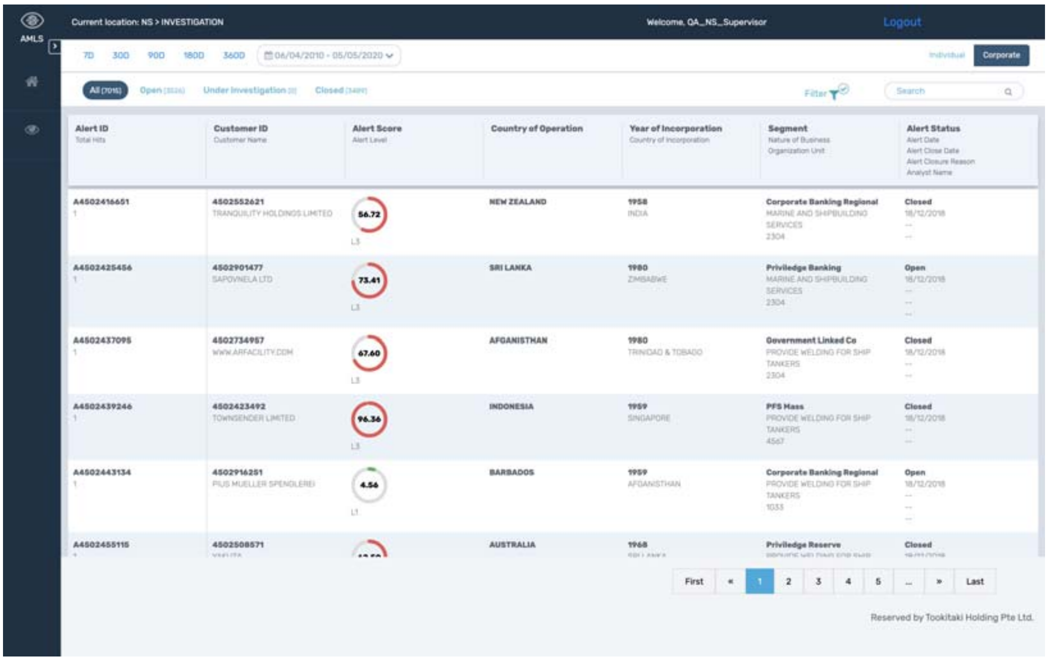
6 Alerts Listing

Figure 42 – NS Alerts Investigation View (Individual)



Alert ID Total Hits	Customer ID Customer Name Type of Business	Alert Score Alert Level	Country of Residence Nationality Country of Birth	Age	Segment Organization Unit	Alert Status Alert Date Alert Close Date Alert Closure Reason Analyst Name
A4500001226 1	4501614375 QUINTON MITCHELL ---	96.78 L3	AFGANISTHAN UNITED STATES BARBADOS	62	Corporate Banking Regional 1022	Closed 18/12/2018 --- ---
A4500001448 1	4501568876 EDGAR CABANLAS ---	96.38 L3	BARBADOS PAKISTAN AUSTRALIA	49	Shareholders Fund 2304	Closed 18/12/2018 --- ---
A4500004777 1	4500506323 SANDRA REGINA CARBONE TADEU MUDALEN ---	21.25 L3	PAKISTAN AUSTRALIA NEW ZEALAND	23	Government Linked Co 1033	Open 18/12/2018 --- --- ---
A4500005166 1	4501798373 CELSO HERNAN ARIAS MORA BUSSV	79.04 L3	PAKISTAN INDONESIA BARBADOS	30	Government Linked Co 1033	Open 18/12/2018 --- --- ---
A4500010325 2	4501325940 JESSE C CRUMBLEY ---	71.07 L3	INDIA THAILAND THAILAND	57	Multi National Corp 1033	Closed 18/12/2018 --- ---
A4500012673 1	4502089619	[illegible]	AFGANISTHAN	60	Corporate Banking Overseas	Closed

Figure 43 - NS Alerts Investigation View (Corporate)

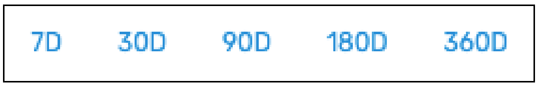


Alert ID Total Hits	Customer ID Customer Name	Alert Score Alert Level	Country of Operation	Year of Incorporation Country of Incorporation	Segment Nature of Business Organization Unit	Alert Status Alert Date Alert Close Date Alert Closure Reason Analyst Name
A4502416651 1	4502552621 TRANQUILITY HOLDINGS LIMITED	56.72 L3	NEW ZEALAND	1958 INDIA	Corporate Banking Regional MARINE AND SHIPBUILDING SERVICES 2304	Closed 18/12/2018 -- --
A4502425456 1	4502901477 SAPOVNELA LTD	73.41 L3	SRI LANKA	1980 ZIMBABWE	Priviledge Banking MARINE AND SHIPBUILDING SERVICES 2304	Open 18/12/2018 -- -- --
A4502437095 1	4502734957 WWW.ARFACILITY.COM	67.60 L3	AFGANISTHAN	1980 TRINIDAD & TOBAGO	Government Linked Co PROVIDE WELDING FOR SHIP TANKERS 2304	Closed 18/12/2018 -- --
A4502439246 1	4502423492 TOWNSENDER LIMITED	96.36 L3	INDONESIA	1959 SINGAPORE	PFS Mass PROVIDE WELDING FOR SHIP TANKERS 4567	Closed 18/12/2018 -- --
A4502443134 1	4502916251 PIUS MUELLER SPENGLEREI	4.56 L1	BARBADOS	1959 AFGANISTHAN	Corporate Banking Regional PROVIDE WELDING FOR SHIP TANKERS 1033	Open 18/12/2018 -- -- --
A4502455115	4502508571		AUSTRALIA	1968	Priviledge Reserve	Closed

Time Range Selection

The time range selection provides filtering options for the user to view alerts within defined time period. The following are the defined time periods:

Figure 44 - Time Range Selection

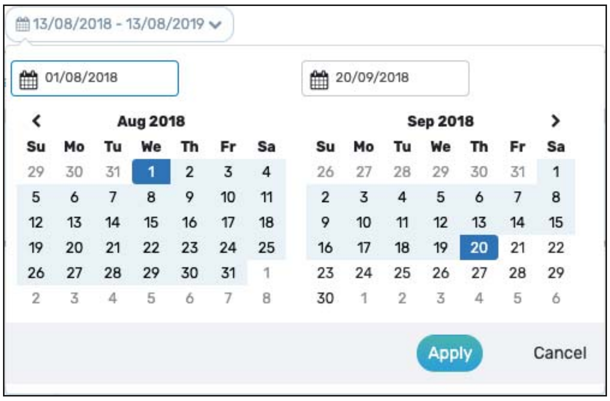


Feature	Description
7D	Filters alerts for the defined 7 days period.
30D	Filters alerts for the defined 30 days period. This is the default selection.
90D	Filters alerts for the defined 90 days period.
180D	Filters alerts for the defined 180 days period.
360D	Filters alerts for the defined 360 days period.

Custom Date Range

The custom date range filter enables the user to review alerts for a customized date range. The user needs to select the from and to dates for the customer filter and click **Apply** to modify the alerts listing based on the filter selected.

Figure 45 - Custom Date Range



13/08/2018 - 13/08/2019

01/08/2018

20/09/2018

Aug 2018

Su	Mo	Tu	We	Th	Fr	Sa
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1
2	3	4	5	6	7	8

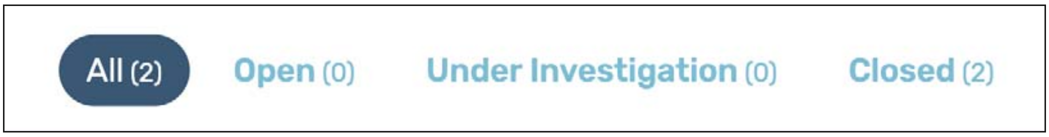
Sep 2018

Su	Mo	Tu	We	Th	Fr	Sa
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	1	2	3	4	5	6

Alert Status Filter

The Alert Status filter enables the user to review the alerts based on status listed below.

Figure 46 - Alert Status Filter

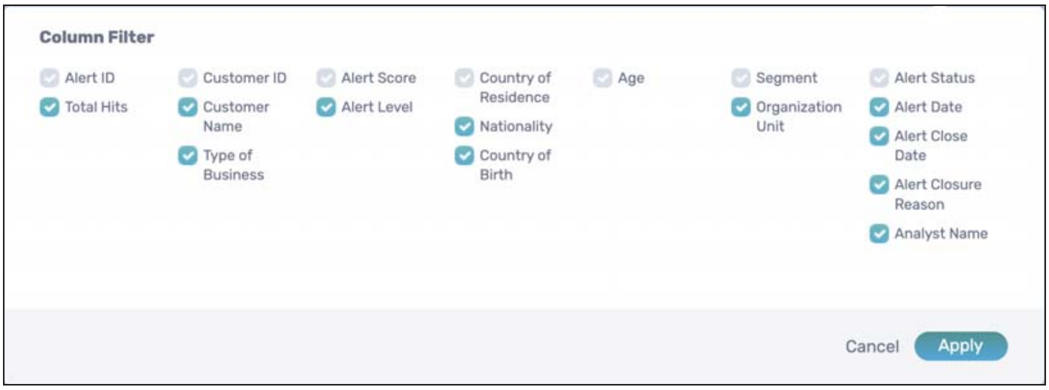


Feature	Description
All	On selection, all alerts are displayed. This is the default selection.
Open	On selection, all alerts with Open status are displayed.
Under Investigation	On selection, all alerts with Under Investigation status are displayed.
Closed	On selection, all alerts with Closed status are displayed.

Column Filter

The column filter allows the user to select the desired columns to view in the alerts listing page. Based on the filter selections, the information on the alerts listing page is tailored.

Figure 47 - Column Filter Individual



Column Filter

☑ Alert ID
☑ Total Hits
☑ Customer ID
☑ Customer Name
☑ Type of Business
☑ Alert Score
☑ Alert Level
☑ Country of Residence
☑ Nationality
☑ Country of Birth
☑ Age
☑ Segment
☑ Organization Unit
☑ Alert Status
☑ Alert Date
☑ Alert Close Date
☑ Alert Closure Reason
☑ Analyst Name

The columns available for individual profiles are listed in the tables below. The following column names displayed with disabled checked boxes.

Column	**Description**
Alert ID	Unique identifier of the alert sent by the primary screening system.
Customer ID	Unique Identifier of the customer.
Alert Score	Highest hit score of the alert.
Country of Residence	Customer country of residence.
Age	Age of customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.

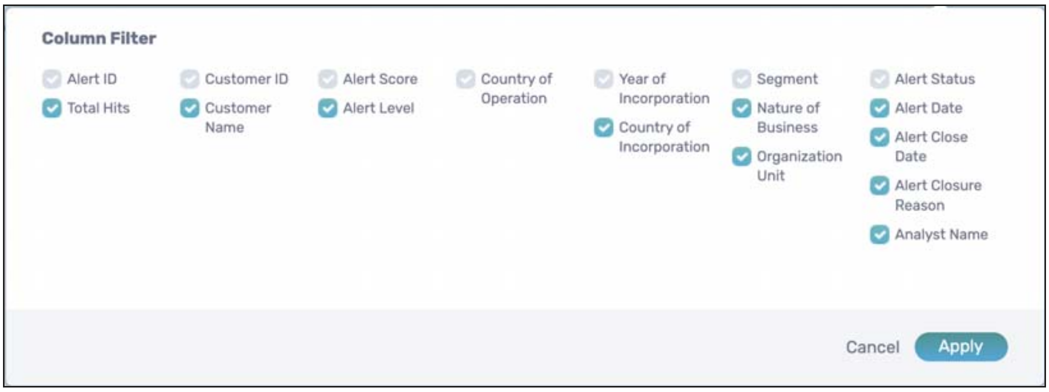
The following column names are with enabled checked boxes for customization

Column	**Description**
Total Hits	Total number of hits for the alert.
Customer Name	Name of the customer.
Type of Business	Type of business of the customer.
Alert Level	AMLS alert prioritization bucket (L1/L2/L3).
Nationality	Citizenship of customer.
Country of Birth	Country of birth.
Organization Unit	Organization unit of the bank to which the customer belongs.
Alert Date	Date on which the alert was generated.

Alert Close Date	Date on which the alert was closed.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Analyst Name	Name of investigating analyst.

The columns available for corporate profiles are listed in the tables below. The following column names displayed with disabled checked boxes.

Figure 48 - Column Filter Corporate



Column Filter

- ☑ Alert ID
- ☑ Total Hits
- ☑ Customer ID
- ☑ Customer Name
- ☑ Alert Score
- ☑ Alert Level
- ☑ Country of Operation
- ☑ Year of Incorporation
- ☑ Country of Incorporation
- ☑ Segment
- ☑ Nature of Business
- ☑ Organization Unit
- ☑ Alert Status
- ☑ Alert Date
- ☑ Alert Close Date
- ☑ Alert Closure Reason
- ☑ Analyst Name

Column	**Description**
Alert ID	Unique identifier of the alert sent by the primary screening system.
Customer ID	Unique Identifier of the customer.
Alert Score	Highest hit score of the alert.
Country of Operation	Country of operation of the corporate customer.
Year of Incorporation	Year of incorporation of the corporate customer.
Segment	Name of the business segment of the bank to which the corporate customer belongs.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.

The following column names are with enabled checked boxes for customization.

Column	**Description**
Total Hits	Total number of hits for the alert.
Customer Name	Name of customer.
Alert Level	AMLS alert prioritization bucket (L1/L2/L3).
Country of Incorporation	Country of incorporation of the corporate customer.
Nature of Business	Nature of business.
Organization Unit	Organization unit of the bank to which the customer belongs.
Alert Date	Date on which the alert was generated.

Alert Close Date	Date on which the alert was closed.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Analyst Name	Name of investigating analyst.

Search Bar

Search Bar enables user to perform browser search on Alert ID, Customer ID and Customer Name columns. It also provides the database search on Alert ID within the selected time frame (time filter).

Figure 49 - Search Bar

Pagination

The Alerts Investigation View lists 20 alerts per page. The pagination bar appears on the bottom right when the alert list contains more than 20 alerts. The user can click on the page *n*, to navigate to the next 20 alerts.

Figure 50 - Pagination

Alerts Listing

Alert Prioritization

The alerts from the primary screening system are prioritized into L1, L2 and L3 buckets based on the scores generated by Tookitaki system. The alert buckets are displayed with the following color scheme:

- L1 - Green
- L2 - Amber
- L3 - Red

Table 10 - Individual Listing

Attributes	**Description**
Alert ID	Unique identifier of the alert sent by the primary screening system.
Total Hits	Total number of hits for the alert.
Customer ID	Unique Identifier of the customer.
Customer Name	Name of the customer.
Type of Business	Occupation of the customer.
Alert Score	Maximum value of hit scores for an alert.
Alert Level	AMLS alert prioritization bucket (L1/L2/L3).

Table 10 - Individual Listing

Nationality	Citizenship of customer.
Country of Birth	Country of birth.
Age	Age of customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Organization Unit	Organization unit of the bank to which the customer belongs.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.
Alert Date	Date on which the alert was generated.
Alert Close Date	Date on which the alert was closed.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Analyst Name	Name of investigating analyst.

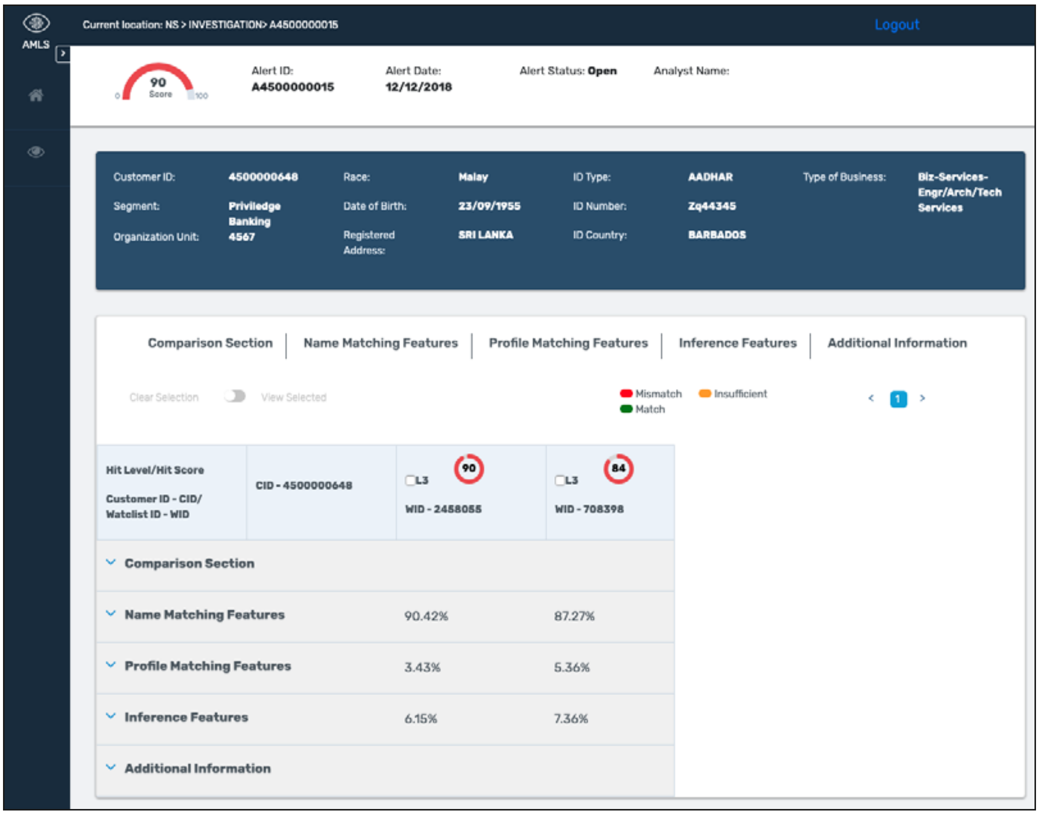
Table 11 - Corporate Listing

Attributes	**Description**
Alert ID	Unique identifier of the alert sent by the primary screening system.
Total Hits	Total number of hits for the alert.
Customer ID	Unique Identifier of the customer.
Customer Name	Name of the customer.
Alert Score	Maximum value of hit scores for an alert.
Alert Level	AMLS alert prioritization bucket (L1/L2/L3).
Country of Operation	Country of operation of the corporate customer.
Year of Incorporation	Year of incorporation of the corporate customer.
Country of Incorporation	Country of incorporation of the corporate customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Nature of Business	Nature of business.
Organization Unit	Organization unit of the bank to which the customer belongs.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.
Alert Date	Date on which the alert was generated.
Alert Close Date	Date on which the alert was closed.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Analyst Name	Name of investigating analyst.

Alert Details View

On clicking a specific alert on the Alerts Listing, the user is taken to the Alert Details page, which provides further details that are useful for the investigation. This page provides side by side comparison of the customer details against the Watchlist information. It also displays details of the features contributing to the hit scores.

Figure 51 - Alert Details View



Current location: NS > INVESTIGATION> A4500000015

90 Score

Alert ID: A4500000015

Alert Date: 12/12/2018

Alert Status: **Open**

Analyst Name:

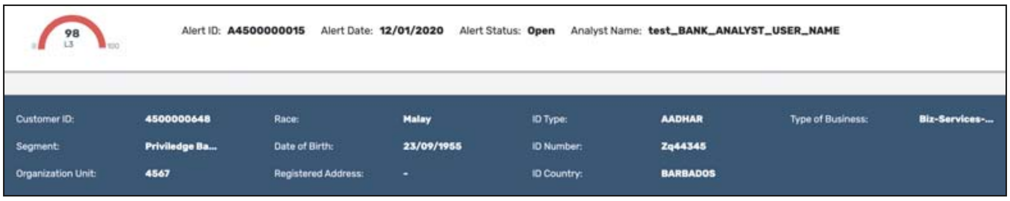
Customer ID:	**4500000648**	Race:	**Malay**	ID Type:	**AADHAR**	Type of Business:	**Biz-Services-Engr/Arch/Tech Services**
Segment:	**Priviledge Banking**	Date of Birth:	**23/09/1955**	ID Number:	**Zq44345**		
Organization Unit:	**4567**	Registered Address:	**SRI LANKA**	ID Country:	**BARBADOS**		

Hit Level/Hit Score Customer ID - CID/ Watclist ID - WID	CID - 4500000648	□L3 90 WID - 2458055	□L3 84 WID - 708398
Comparison Section			
Name Matching Features		90.42%	87.27%
Profile Matching Features		3.43%	5.36%
Inference Features		6.15%	7.36%
Additional Information			

Alert Summary

At the top of the Alert Details View, the application displays the summary of the alert and the following list of information.

Figure 52 - Alert Summary Individual



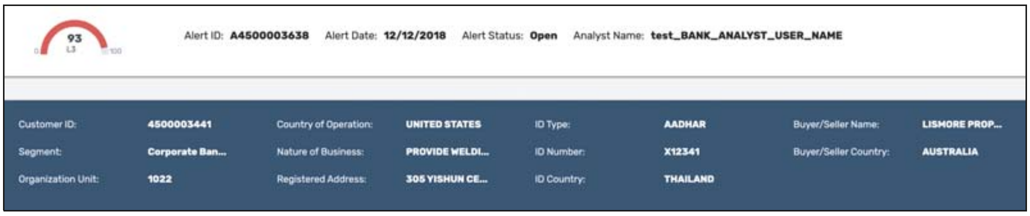
Alert ID: **A4500000015** Alert Date: **12/01/2020** Alert Status: **Open** Analyst Name: **test_BANK_ANALYST_USER_NAME**

Customer ID:	**4500000648**	Race:	**Malay**	ID Type:	**AADHAR**	Type of Business:	**Biz-Services-...**
Segment:	**Priviledge Ba...**	Date of Birth:	**23/09/1955**	ID Number:	**Zq44345**		
Organization Unit:	**4567**	Registered Address:	**-**	ID Country:	**BARBADOS**		

Table 12 - Alert Summary Individual

Attributes	**Description**
Alert Score	Maximum value of hit scores for an alert.
Alert Level	Alert prioritization bucket (L1/L2/L3).
Alert ID	Unique identifier of the alert sent by the primary screening system.
Customer ID	Unique identifier for customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Organization Unit	Organization unit of the bank to which the customer belongs.
Analyst Name	Name of investigating analyst.
Alert Date	Date on which the alert was generated.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.
ID Type	Identification document type.
ID Number	Identification document number.
ID Country	Identification document issuing country.
Race	Racial identity of customer.
Date of Birth	Date of birth of customer.
Registered Address	Registered address of the customer.
Type of Business	Occupation of the customer.

Figure 53 - Alert Summary Corporate



Alert ID: **A4500003638** Alert Date: **12/12/2018** Alert Status: **Open** Analyst Name: **test_BANK_ANALYST_USER_NAME**

Customer ID:	4500003441	Country of Operation:	UNITED STATES	ID Type:	AADHAR	Buyer/Seller Name:	LISMORE PROP...
Segment:	Corporate Ban...	Nature of Business:	PROVIDE WELDI...	ID Number:	X12341	Buyer/Seller Country:	AUSTRALIA
Organization Unit:	1022	Registered Address:	305 YISHUN CE...	ID Country:	THAILAND		

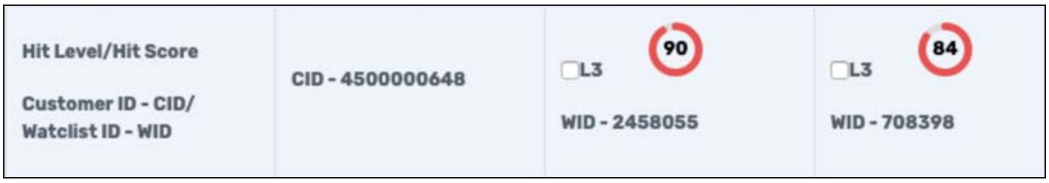
Table 13 - Alert Summary Corporate

Attributes	Description
Alert Score	Maximum value of hit scores for an alert.
Alert Level	Alert prioritization bucket (L1/L2/L3).
Alert ID	Unique identifier of the alert sent by the primary screening system.
Customer ID	Unique identifier for customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Organization Unit	Organization unit of the bank to which the customer belongs.
Analyst Name	Name of investigating analyst.
Alert Date	Date on which the alert was generated.
Alert Closure Reason	Reason for alert closure as entered by the investigator.
Alert Status	Status of the prioritized alerts. The alert status can be Open, Under Investigation or Closed.
ID Type	Identification document type.
ID Number	Identification document number.
ID Country	Identification document issuing country.
Country of Operation	Country of operation of the corporate customer.
Nature of Business	Nature of business .
Registered Address	Registered address of the customer.
Buyer/Seller Name	List of buyer and seller names associated with the customer.
Buyer/Seller Country	List of buyer and seller countries associated with the customer.
Year of Incorporation	Year of incorporation of the corporate customer.
Country of Incorporation	Country of incorporation of the corporate customer.

Hit Score

The Hit Scores for both Individual and Corporate are displayed for each Watchperson.

Figure 54 - Hit Score



Hit Level/Hit Score Customer ID - CID/ Watclist ID - WID	CID - 4500000648	□L3 90 WID - 2458055	□L3 84 WID - 708398

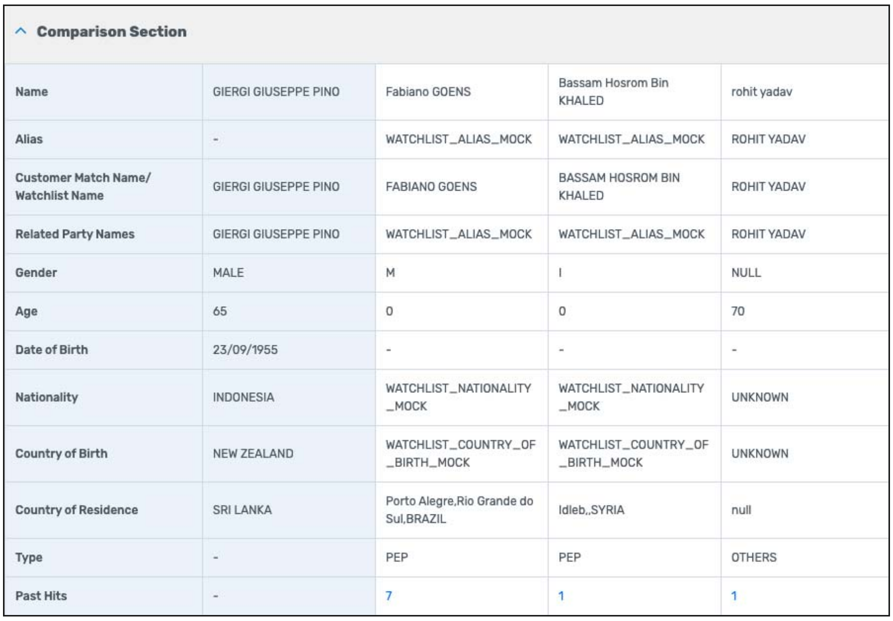
Table 14 - Hit Score Individual and Corporate

Attributes	Description
Hit Score	Score calculated by Tookitaki AMLS system for the watchperson hit comparison.

Comparison Section

The comparison section enables the investigator to view the customer information side by side with the Watchperson information. This assists the user to ascertain that the hit is true or false match.

Figure 55 - Comparison Section Individual



^ Comparison Section

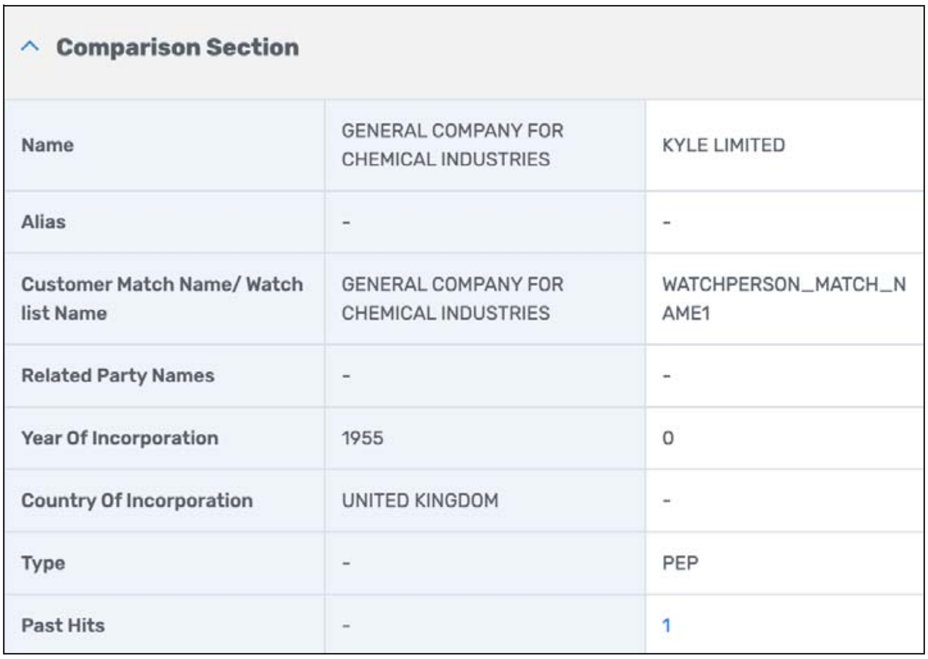
Name	GIERGI GIUSEPPE PINO	Fabiano GOENS	Bassam Hosrom Bin KHALED	rohit yadav
Alias	-	WATCHLIST_ALIAS_MOCK	WATCHLIST_ALIAS_MOCK	ROHIT YADAV
Customer Match Name/ Watchlist Name	GIERGI GIUSEPPE PINO	FABIANO GOENS	BASSAM HOSROM BIN KHALED	ROHIT YADAV
Related Party Names	GIERGI GIUSEPPE PINO	WATCHLIST_ALIAS_MOCK	WATCHLIST_ALIAS_MOCK	ROHIT YADAV
Gender	MALE	M	I	NULL
Age	65	0	0	70
Date of Birth	23/09/1955	-	-	-
Nationality	INDONESIA	WATCHLIST_NATIONALITY_MOCK	WATCHLIST_NATIONALITY_MOCK	UNKNOWN
Country of Birth	NEW ZEALAND	WATCHLIST_COUNTRY_OF_BIRTH_MOCK	WATCHLIST_COUNTRY_OF_BIRTH_MOCK	UNKNOWN
Country of Residence	SRI LANKA	Porto Alegre,Rio Grande do Sul,BRAZIL	Idleb,,SYRIA	null
Type	-	PEP	PEP	OTHERS
Past Hits	-	7	1	1

Listed below are the field definitions available in the customer and watchperson information comparison section for individual.

Table 15 - Comparison Section Individual

Attributes	**Description**
Customer ID	Unique hit identifier for customer.
Watchlist ID	Watchperson unique identifier in Watchlist.
Name	Comparison of primary name between customer and watchperson.
Alias	Comparison of alias name between customer and Watchperson.
Customer Match Name/Watchlist Match Name	Match name that triggers the hit.
Related Party Names	Semi colon separated names of all C2C related data.
Type	Type of hits indicating the severity of hit (Adverse, PEP, Sanction).
Past Hits	Number of past hits for the customer against the same watchperson information. The following details are displayed: ▪ Alert ID ▪ Hit ID ▪ Type ▪ Alert Close Date ▪ Analyst Name ▪ Alert Closure Reason
Age	Comparison of age between customer and watchperson.
Date of Birth	Comparison of date of birth between customer and watchperson.
Gender	Comparison of gender between customer and watchperson.
Country of Birth	Comparison of country of birth between customer and watchperson.
Nationality	Comparison of nationality between customer and watchperson.
Country of Residence	Comparison of the country inferred from residential address of the customer against the watchperson location.

Figure 56 - Comparison Section Corporate



^ Comparison Section

Name	GENERAL COMPANY FOR CHEMICAL INDUSTRIES	KYLE LIMITED
Alias	-	-
Customer Match Name/ Watch list Name	GENERAL COMPANY FOR CHEMICAL INDUSTRIES	WATCHPERSON_MATCH_NAME1
Related Party Names	-	-
Year Of Incorporation	1955	0
Country Of Incorporation	UNITED KINGDOM	-
Type	-	PEP
Past Hits	-	1

Listed below are the field definitions available in the customer and watchperson information comparison section for corporate.

Table 16 - Comparison Section Corporate

Attributes	**Description**
Customer ID	Unique hit identifier for customer.
Watchlist ID	Watchperson unique identifier in Watchlist.
Name	Comparison of primary name between customer and watchperson.
Alias	Comparison of alias name between customer and watchperson.
Customer Match Name/Watchlist Match Name	Data displayed under customer and watchperson hit data
Related Party Names	Semi colon separated names of all C2C related data.
Type	Type of hits indicating the severity of hit (Adverse, PEP, Sanction).

Table 16 - Comparison Section Corporate

Attributes	Description
Past Hits	Number of past hits for the customer against the same watchperson information. The following details are displayed: ▪ Alert ID ▪ Hit ID ▪ Type ▪ Alert Close Date ▪ Analyst Name ▪ Alert Closure Reason
Year of Incorporation	Comparison of year of incorporation between customer and watchperson.
Country of Incorporation	Comparison of country of incorporation between customer and watchperson.

Feature Matching Section

Feature matching section displays the features and its contributions to the prediction. Each feature matching section will display the relevant features. Against each feature the following data will be displayed:

- Feature Value
- Feature Contribution

Feature value computed by prediction pipelines. All the values are displayed in 2 decimal places if they are numerical.

Feature contribution is displayed in percentage contribution (+ve/-ve) to the hits score and level prediction by prediction pipelines. A positive prediction contribution indicates how strongly the feature contributes to the prediction where as a negative prediction contribution indicates how strongly the feature does not participate in the prediction.

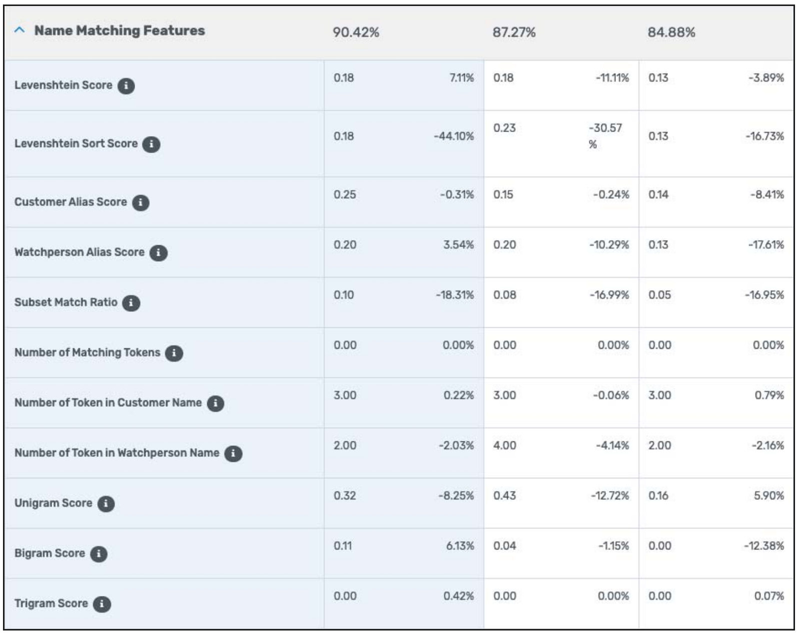
Following are sections displayed under Feature Matching Sections:

- Name Matching Features Section
- Profile Matching Features Section
- Inference Features Section

Name Matching Features

Percentile against the features within the group indicates their contribution in feature group prediction. All feature percentage contributions categorized under Name Matching Feature add up to the Name Matching Feature percentage contribution.

Figure 57 - Name Matching



^ Name Matching Features	90.42%		87.27%		84.88%	
Levenshtein Score	0.18	7.11%	0.18	-11.11%	0.13	-3.89%
Levenshtein Sort Score	0.18	-44.10%	0.23	-30.57 %	0.13	-16.73%
Customer Alias Score	0.25	-0.31%	0.15	-0.24%	0.14	-8.41%
Watchperson Alias Score	0.20	3.54%	0.20	-10.29%	0.13	-17.61%
Subset Match Ratio	0.10	-18.31%	0.08	-16.99%	0.05	-16.95%
Number of Matching Tokens	0.00	0.00%	0.00	0.00%	0.00	0.00%
Number of Token in Customer Name	3.00	0.22%	3.00	-0.06%	3.00	0.79%
Number of Token in Watchperson Name	2.00	-2.03%	4.00	-4.14%	2.00	-2.16%
Unigram Score	0.32	-8.25%	0.43	-12.72%	0.16	5.90%
Bigram Score	0.11	6.13%	0.04	-1.15%	0.00	-12.38%
Trigram Score	0.00	0.42%	0.00	0.00%	0.00	0.07%

Table 17 - Name Matching Features

Attributes	**Description**
Levenshtein Score	Algorithm measures similarity between customer primary name and watchperson name. The higher the score, the stronger the match.
Levenshtein Sort Score	Algorithm measures more detailed similarity between customer primary name and watchperson name to generate a score. The names are sorted and matched. The higher the score, the stronger the match.
Alias Score Customer	Algorithm measures similarity between customer alias and watchperson name to generate a score. The higher the score, the stronger the match.
Alias Score Watchlist	Algorithm measures similarity between customer primary name and watchperson alias name to generate a score. The higher the score, the stronger the match.
Subset Match Ratio	Statistic depicts the match between customer primary name and watchperson name. The higher the score, the stronger the match.
Number of Matching Tokens	Number of matching words in customer primary name and watchperson name. The higher the score, the stronger the match.

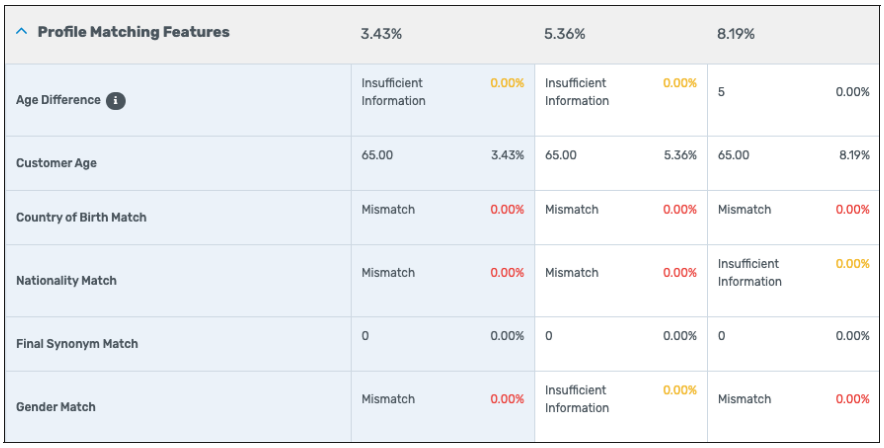
Table 17 - Name Matching Features

Attributes	**Description**
Unigram Score	Algorithm measures similarity customer primary name and Watchperson name to generate a score. The measure is done at unigram (1) level. The higher the score, the stronger the match.
Bigram Score	Algorithm measures similarity between customer primary name and watchperson name to generate a score. The measure is done at bigram (2) level. The higher the score, the stronger the match.
Trigram Score	Algorithm measures similarity between customer primary name and watchperson name to generate a score. The measure is done at trigram (3) level. The higher the number, the stronger the match.

Profile Matching Features

All feature percentage contributions categorized under Profile Matching Feature add up to the Profile Matching Feature percentage contribution.

Figure 58 - Profile Matching - Individual

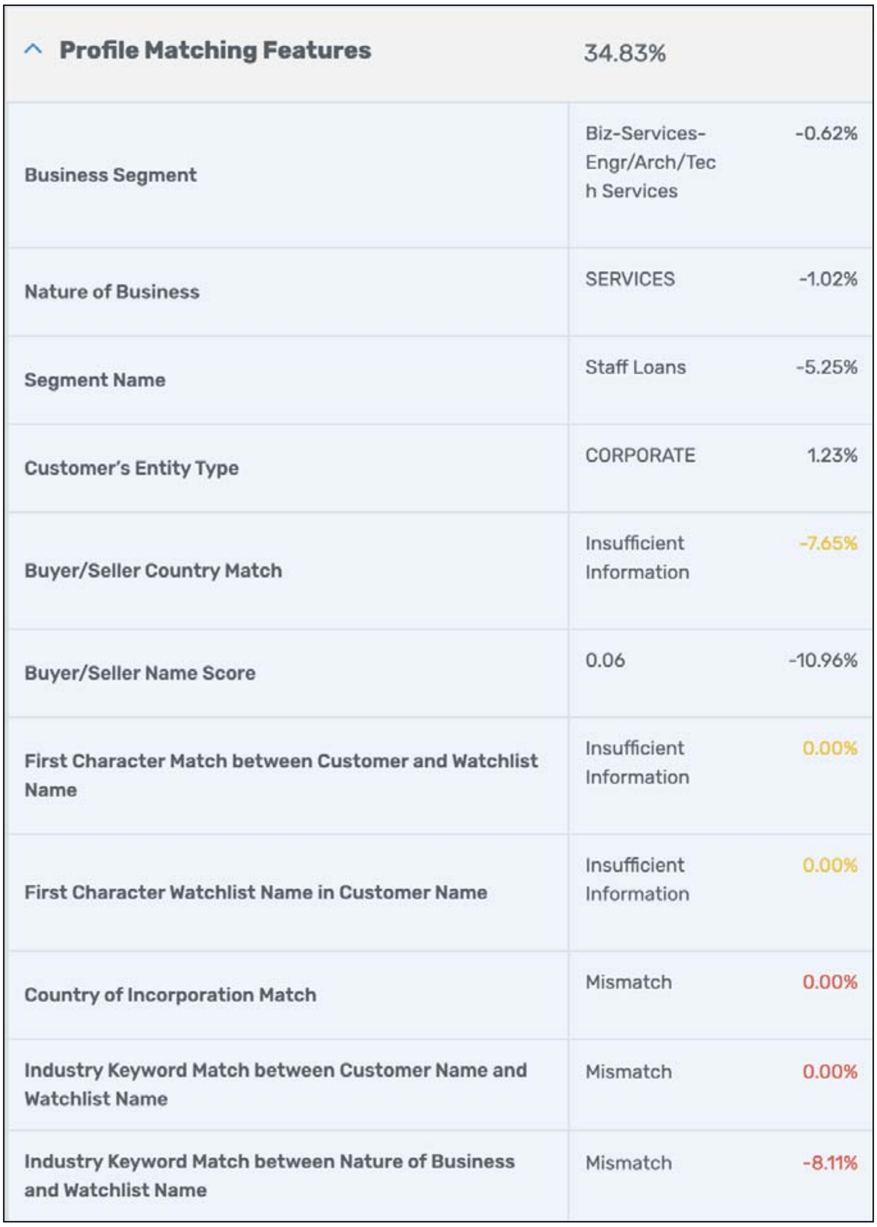


Profile Matching Features	3.43%		5.36%		8.19%	
Age Difference	Insufficient Information	0.00%	Insufficient Information	0.00%	5	0.00%
Customer Age	65.00	3.43%	65.00	5.36%	65.00	8.19%
Country of Birth Match	Mismatch	0.00%	Mismatch	0.00%	Mismatch	0.00%
Nationality Match	Mismatch	0.00%	Mismatch	0.00%	Insufficient Information	0.00%
Final Synonym Match	0	0.00%	0	0.00%	0	0.00%
Gender Match	Mismatch	0.00%	Insufficient Information	0.00%	Mismatch	0.00%

Table 18 - Profile Matching - Individual

Attributes	**Description**
Age Difference	Age difference between customer and closest match in watchperson data.
Customer Age	Age of customer derived from date of birth. Same value is displayed for all watchperson data but contribution (%) varies for each hit prediction.
Country Of Birth Match	Categorical value to indicate country of birth match between customer and watchperson.
Nationality Match	Nationality match between customer and watchperson.
Final Synonym Match	Number of synonyms token match between watchperson name and customer primary name.
Gender Match	Gender match between customer and watchperson.

Figure 59 - Profile Matching - Corporate



Profile Matching Features	34.83%	
Business Segment	Biz-Services-Engr/Arch/Tech Services	-0.62%
Nature of Business	SERVICES	-1.02%
Segment Name	Staff Loans	-5.25%
Customer's Entity Type	CORPORATE	1.23%
Buyer/Seller Country Match	Insufficient Information	-7.65%
Buyer/Seller Name Score	0.06	-10.96%
First Character Match between Customer and Watchlist Name	Insufficient Information	0.00%
First Character Watchlist Name in Customer Name	Insufficient Information	0.00%
Country of Incorporation Match	Mismatch	0.00%
Industry Keyword Match between Customer Name and Watchlist Name	Mismatch	0.00%
Industry Keyword Match between Nature of Business and Watchlist Name	Mismatch	-8.11%

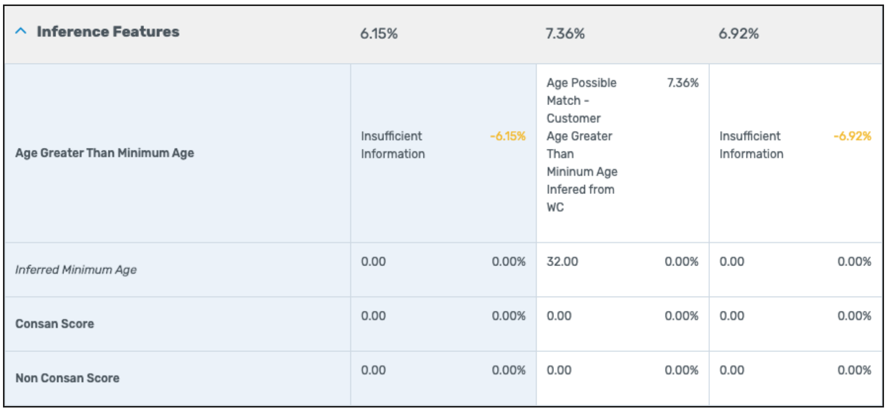
Table 19 - Profile Matching - Corporate

Attributes	Description
Business Segment	Type of business the customer is in. Same value is displayed for all watchperson data but contribution (%) varies for each hit prediction.
Nature of Business	Nature of business the customer is in. Same value is displayed for all watchperson data but contribution (%) varies for each hit prediction.
Segment Name	Banking segment name customer belongs to. Same value is displayed for all watchperson data but contribution (%) varies for each hit prediction.
Customer's Entity Type	Entity type description of customer. Same value is displayed for all watchperson data but contribution (%) varies for each hit prediction.
Buyer/Seller Name Score	Measure of highest Levenshtein score calculated between customer's buyer and seller names and watchperson primary name. Higher value indicates stronger match.
Buyer/Seller Country Match	Measure of highest Levenshtein score calculated between customer's buyer and seller country of incorporation and watchperson country of incorporation. Higher value indicates stronger match.
First Character Match Between Customer and Watchlist Name	Matches first character of each name part between customer primary name and watchperson name.
First Character Watchlist In Customer Name Match	Match first character name of watchperson is a subset of customer primary name.
Country of Incorporation Match	Match country of incorporation between corporate customer and watchperson.
Industry Keyword Match Between Customer Name And Watchlist Name	Check for the presence of predefined set of keywords in customer primary name and watchperson name. Keywords are picked to check if the two corporate entity belongs to the same industry.
Industry Keyword Match Between Industry Segment And Watchlist Name	Check for the presence of predefined set of keywords in customer nature of business field and watchperson name. Keywords are picked to check if the two corporate entity belongs to the same industry.

Inference Features

All feature percentage contributions categorized under the Inference Matching Feature add up to the Inference Matching Feature percentage contribution.

Figure 60 - Inference Feature Individual

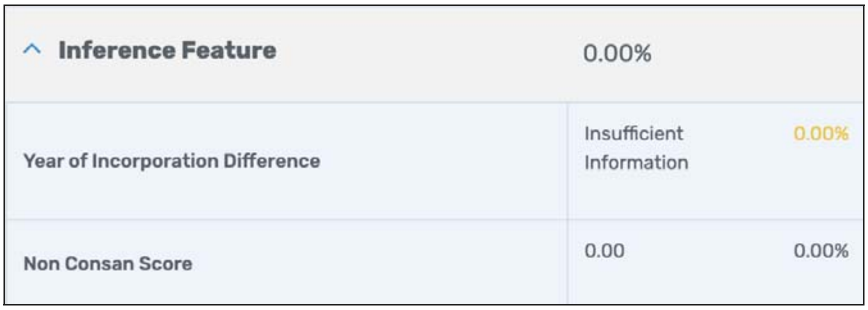


^ Inference Features	6.15%		7.36%		6.92%	
Age Greater Than Minimum Age	Insufficient Information	-6.15%	Age Possible Match - Customer Age Greater Than Mininum Age Infered from WC	7.36%	Insufficient Information	-6.92%
Inferred Minimum Age	0.00	0.00%	32.00	0.00%	0.00	0.00%
Consan Score	0.00	0.00%	0.00	0.00%	0.00	0.00%
Non Consan Score	0.00	0.00%	0.00	0.00%	0.00	0.00%

Table 20 - Inference Feature - Individual

Attributes	Description
Age Greater Than Minimum Age	Categorical field to indicate if customer age exceeds minimum inferred age of watchlist customer.
Inferred Minimum Age	Inferred Minimum Age from watchlist data (using regular expressions).
Consan Score	Relationship score (blood releated) that determine if there is any common entity between customer and watchlist data. Presence of common entity is more likely to be a true hit.
Non Consan Score	Relationship score (non-blood / non-consanguinity) that determine if there is any common entity between customer and watchlist data. Presence of common entity is more likely to be a true hit.

Figure 61 - Inference Feature Corporate



^ Inference Feature	0.00%	
Year of Incorporation Difference	Insufficient Information	0.00%
Non Consan Score	0.00	0.00%

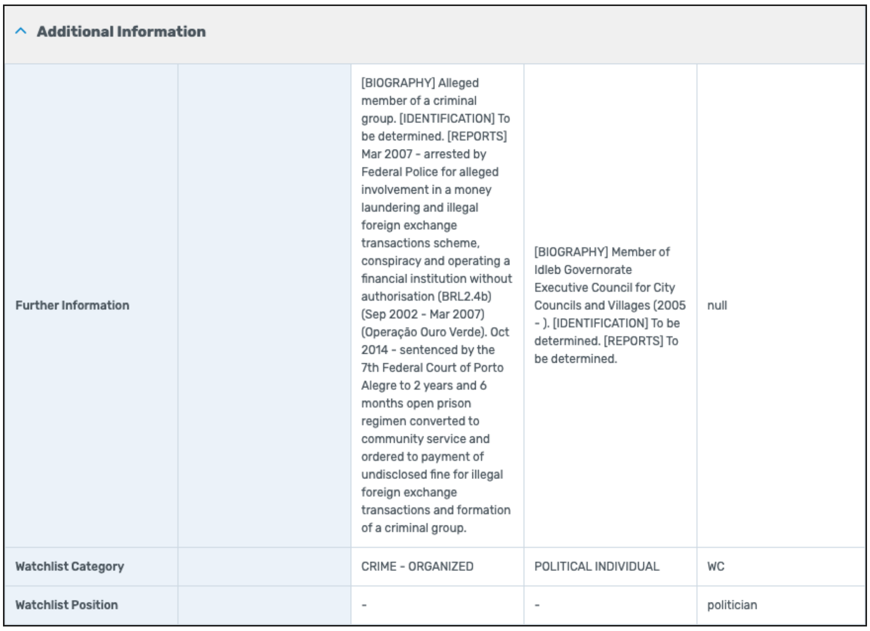
Table 21 - Inference Feature - Corporate

Attributes	**Description**
Year Of Incorporation Difference	Year of incorporation difference between customer and watchperson.
Non Consan Score	Relationship score (non-blood / non-consanguinity) that determine if there is any common entity between customer and watchperson data. Presence of common entity is more likely to be a true hit.

Additional Information

In addition to the above, the alerts and hits detail provide additional analysis information specific to watchlist profile that supplement the alert decision for analysts.

Figure 62 - Additional Information - Individual



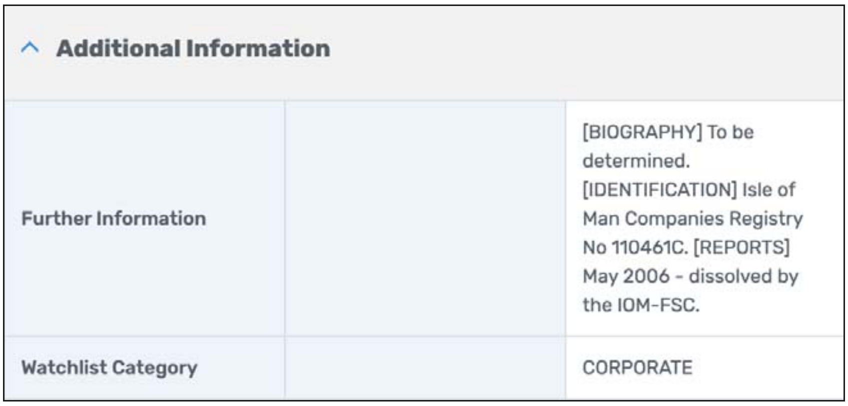
^ **Additional Information**

Further Information		[BIOGRAPHY] Alleged member of a criminal group. [IDENTIFICATION] To be determined. [REPORTS] Mar 2007 - arrested by Federal Police for alleged involvement in a money laundering and illegal foreign exchange transactions scheme, conspiracy and operating a financial institution without authorisation (BRL2.4b) (Sep 2002 - Mar 2007) (Operação Ouro Verde). Oct 2014 - sentenced by the 7th Federal Court of Porto Alegre to 2 years and 6 months open prison regimen converted to community service and ordered to payment of undisclosed fine for illegal foreign exchange transactions and formation of a criminal group.	[BIOGRAPHY] Member of Idleb Governorate Executive Council for City Councils and Villages (2005 -). [IDENTIFICATION] To be determined. [REPORTS] To be determined.	null
Watchlist Category		CRIME - ORGANIZED	POLITICAL INDIVIDUAL	WC
Watchlist Position		-	-	politician

Table 22 - Additional Information - Individual

Attributes	**Description**
Further Information	Watchlist text field to get further information.
Watchlist Category	Category of watchlist customer.
Watchlist Position	Watchperson's position in the watchlist.

Figure 63 - Additional Information - Corporate



Additional Information

Further Information		[BIOGRAPHY] To be determined. [IDENTIFICATION] Isle of Man Companies Registry No 110461C. [REPORTS] May 2006 - dissolved by the IOM-FSC.
Watchlist Category		CORPORATE

Table 23 - Additional Information - Corporate

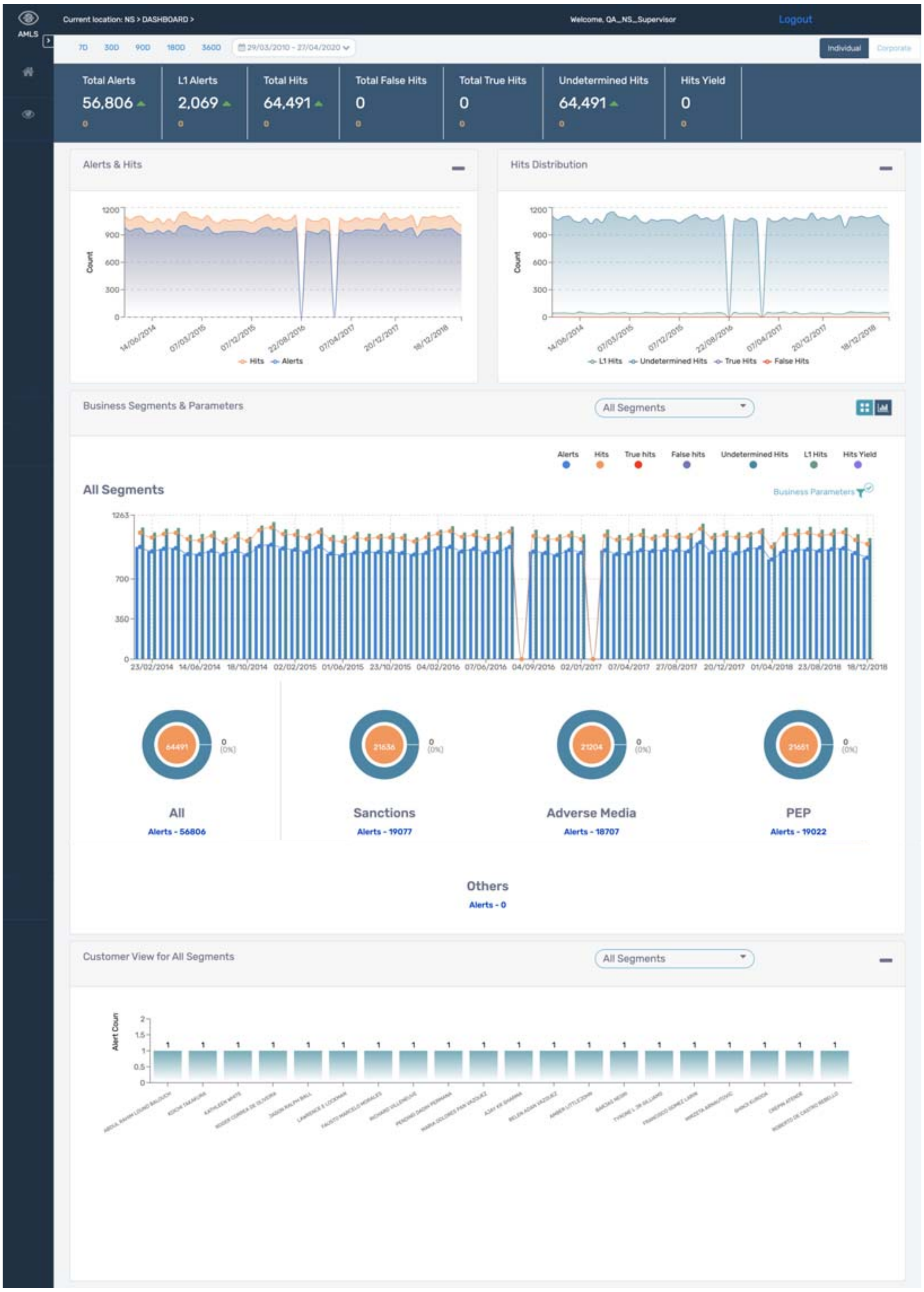
Attributes	**Description**
Further Information	Watchlist text field to get further information.
Watchlist Category	Category of watchlist customer.

Risk View

The Risk View provides statistical insights for alert and hit information from different perspectives. The page displays values for both the current period and the previous period as defined by the selected date range or the custom date range. These values are displayed graphically in the following widgets:

- Alerts & Hits
  See “Alerts and Hits” on page 78.
- Hits Distribution
  See “Hits Distribution” on page 79.
- Business Segments & Parameters
  See “Business Segments & Parameters” on page 80.
- Customer View for All Segments
  See “Customer View for All Segments” on page 82.

Figure 64 - Risk View



Current location: NS > DASHBOARD >

Welcome, QA_NS_Supervisor

Logout

7D 30D 90D 180D 360D 29/03/2010 - 27/04/2020

Individual Corporate

Total Alerts	L1 Alerts	Total Hits	Total False Hits	Total True Hits	Undetermined Hits	Hits Yield
56,806	2,069	64,491	0	0	64,491	0
0	0	0	0	0	0	0

Alerts & Hits

Hits Distribution

Business Segments & Parameters

All Segments

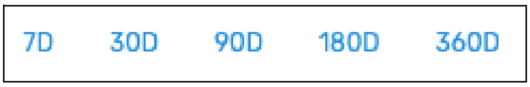
Customer View for All Segments

All Segments

Time Range Selection

Time range filters are available for the user to view the values of the Risk View widgets according to desired time frames as follows:

Figure 65 - Time Range Selection

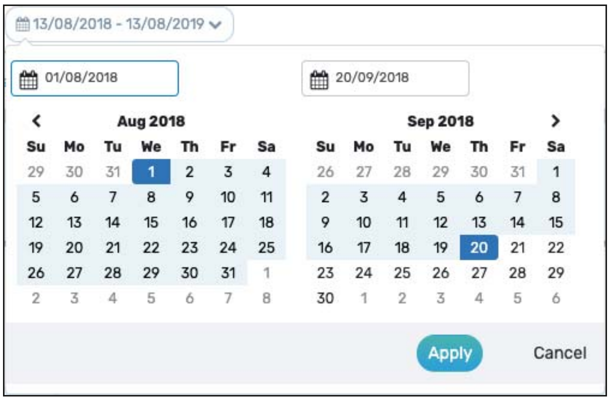


Feature	Description
7D	Filters alerts for the defined 7 days period.
30D	Filters alerts for the defined 30 days period. This is the default selection.
90D	Filters alerts for the defined 90 days period.
180D	Filters alerts for the defined 180 days period.
360D	Filters alerts for the defined 360 days period.

Custom Date Range

The custom date range filter enables the user to filter the statistical information for a customized date range. The user needs to select the from and to date for the customer filter and click **Apply** to modify the Risk View page based on the filter selected.

Figure 66 - Custom Date Range



13/08/2018 - 13/08/2019

01/08/2018

20/09/2018

Aug 2018

Su	Mo	Tu	We	Th	Fr	Sa
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1
2	3	4	5	6	7	8

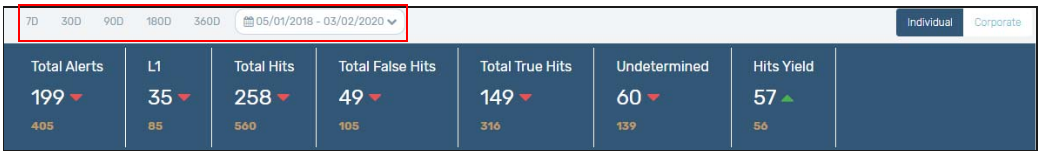
Sep 2018

Su	Mo	Tu	We	Th	Fr	Sa
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	1	2	3	4	5	6

Summary

The information displayed on each widget is specific to the time frame selected on the “Time Range Selection” on page 75.

Figure 67 - Risk View Summary

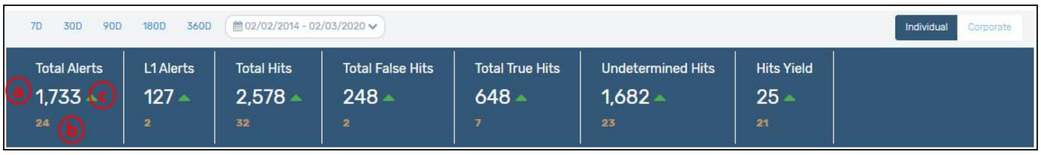


7D 30D 90D 180D 360D 05/01/2018 - 03/02/2020 Individual Corporate

Total Alerts	L1	Total Hits	Total False Hits	Total True Hits	Undetermined	Hits Yield
199 ▼	35 ▼	258 ▼	49 ▼	149 ▼	60 ▼	57 ▲
405	85	560	105	316	139	56

On the summary bar, each statistical category is summarized into two values, total count of current and previous time frame.

Figure 68 - Statistics Bar



7D 30D 90D 180D 360D 02/02/2014 - 02/03/2020 Individual Corporate

Total Alerts	L1 Alerts	Total Hits	Total False Hits	Total True Hits	Undetermined Hits	Hits Yield
1,733 ▲	127 ▲	2,578 ▲	248 ▲	648 ▲	1,682 ▲	25 ▲
24	2	32	2	7	23	21

a) Total count of selected timeframe.

For instance,
Total Alerts: Calendar day 17 March 2020 and user selects a timeframe of 30D; then the stats show total count of all the alerts dated from 17 Feb 2020 to 17 March 2020.

b) Total count of similar length of previous timeframe.

For instance,
Total Alerts: Calendar day 17 March 2020 and user selects timeframe of 30D; then the stats show total count of all the alerts generated from 18 Jan 2020 to 16 Feb 2020.

c) The arrow icon besides each group of stats indicates increase or decrease in count of selected timeframe and previous timeframe.

Example 1: Total count for selected timeframe is 3000
Total count for previous timeframe is 4000

Statistics are displayed as:
3000 v
4000

Example 2: Total count for selected timeframe is 3000
Total count for previous timeframe is 2500

Statistics are displayed as:
3000 ^
2500

Listed below are the field definitions available in the high-level statistics section on alerts and hits.

Table 24 - Risk View Summary Bar

Attribute	**Description**
Total Alerts	■ Count of total number of all unique alerts for selected time frame. ■ Count of total number of all unique alerts for similar previous time frame.
L1 Alerts	■ Count of total number of all L1 for selected time frame. ■ Count of total number of all L1 for similar previous time frame.
Total Hits	■ Count of total number of all hits for selected time frame. ■ Count of total number of all hits for similar previous time frame.
Total False Hits	■ Count of total number of all false Hits for selected time frame. ■ Count of total number of all false Hits for similar previous time frame.
Total True Hits	■ Count of total number of all true Hits for selected time frame. ■ Count of total number of all true Hits for similar previous time frame.
Undetermined Hits	■ Count of total number of all undetermined for selected time frame. ■ Count of total number of all undetermined for similar previous time frame.
Hits Yield	■ Hit Yield for selected time frame. ■ Hit Yield for similar previous time frame.

Alerts and Hits

The Alerts and Hits widget shows a graphical count of Name Screening Hits and Alerts generated by the primary screening system and as ingested into the AMLS application.

Mouse over the graph to display the below stated information on the top and bottom of the pop-up at the specific time.

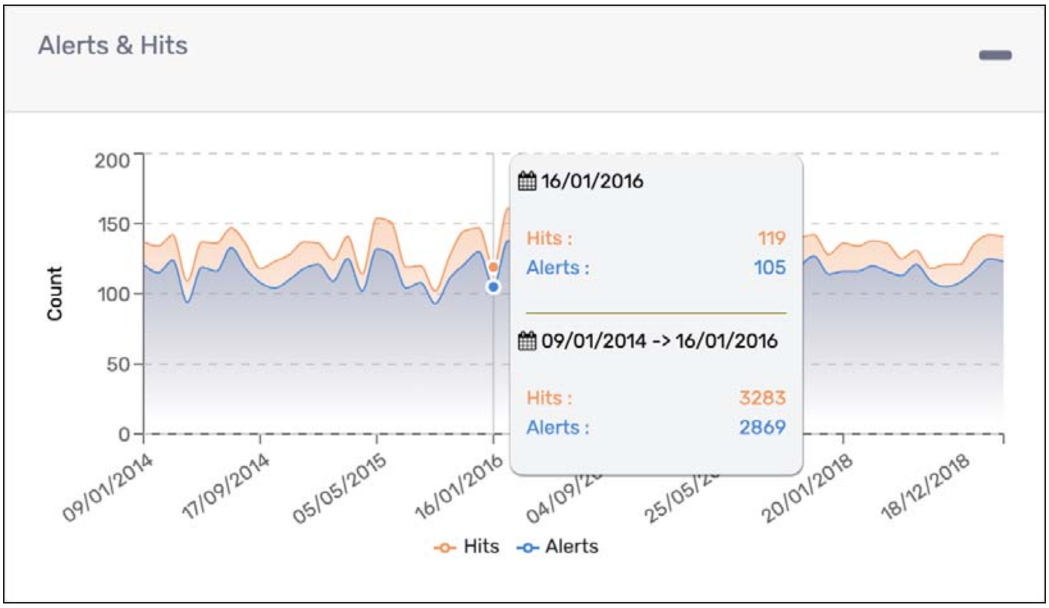
Top

- Date of the selected data point.
- Alerts — number of alerts generated and loaded into AMLS for the specific date.
- Hits — number of hits generated and loaded into AMLS for the specific date.

Bottom

- Date range of the selected data point.
- Alerts — total number of alerts generated and loaded into AMLS within the date range.
- Hits — total number of hits generated and loaded into AMLS within the date range.

Figure 69 - Alerts and Hits



The graph shows the following, plotted on the axes:

- X-Axis — Date per the selected time frame
- Y-Axis — Number of Alerts/Hits

Hits Distribution

The Hits Distribution widget shows the breakdown of the Name Screening Hits based on status - True, False, Undetermined and L1.

Mouse over the graph to display the below stated information on the top and bottom of the popup at the specific time.

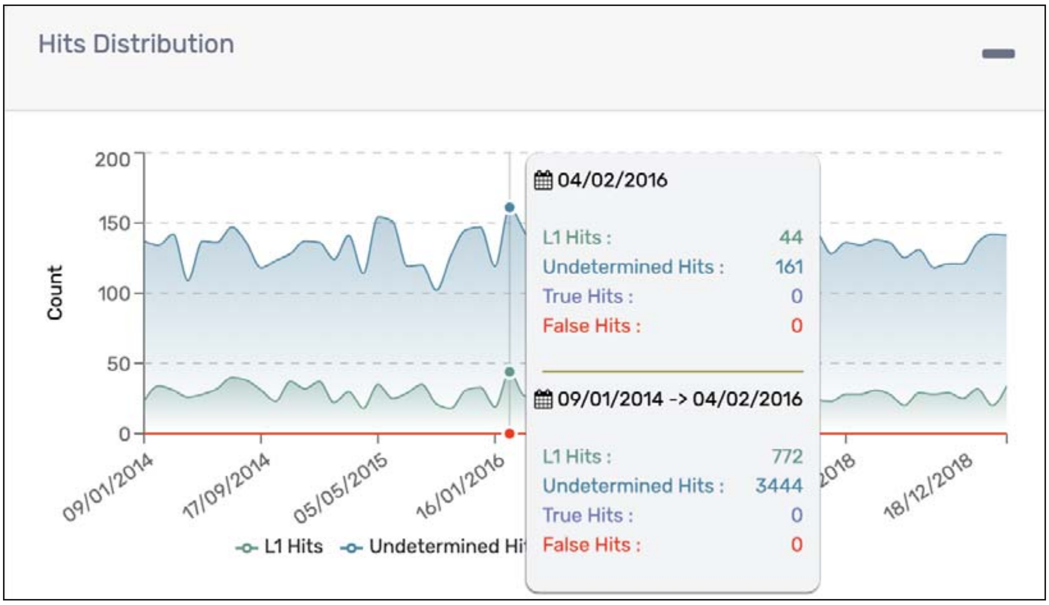
Top

- Date of the selected data point.
- True Hits — number of true status hits for the specific date.
- False Hits — number of false status hits for the specific date.
- Undetermined Hits — number of undetermined status hits for the specific date.
- L1 Hits — number of low quality alerts for the specific date.

Bottom

- Date range of the selected data point.
- True Hits — total number of true status hits within the date range.
- False Hits — total number of false status hits within the date range.
- Undetermined Hits — total number of undetermined status hits within the date range.
- L1 Hits — total number of low quality alerts within the date range.

Figure 70 - Hits Distribution



Hits Distribution

04/02/2016

L1 Hits : 44
Undetermined Hits : 161
True Hits : 0
False Hits : 0

09/01/2014 -> 04/02/2016

L1 Hits : 772
Undetermined Hits : 3444
True Hits : 0
False Hits : 0

The graph shows the following, plotted on the axes:

- X-Axis — Date per the selected time frame
- Y-Axis — Number of Hits

Business Segments & Parameters

The Business Segments & Parameters widget provides a graphical presentation of the total alert details across business segments by the following categories:

- Alerts and Hits as individual graph statistics elements plotted across the stack bar graph
- Count of Alerts, Hits, L1 Hits, Undetermined Hits, True Hits and False Hits as a statistical stack bar.
- Pie charts

Mouse over the graph to pinpoint a specific date and display the below details.

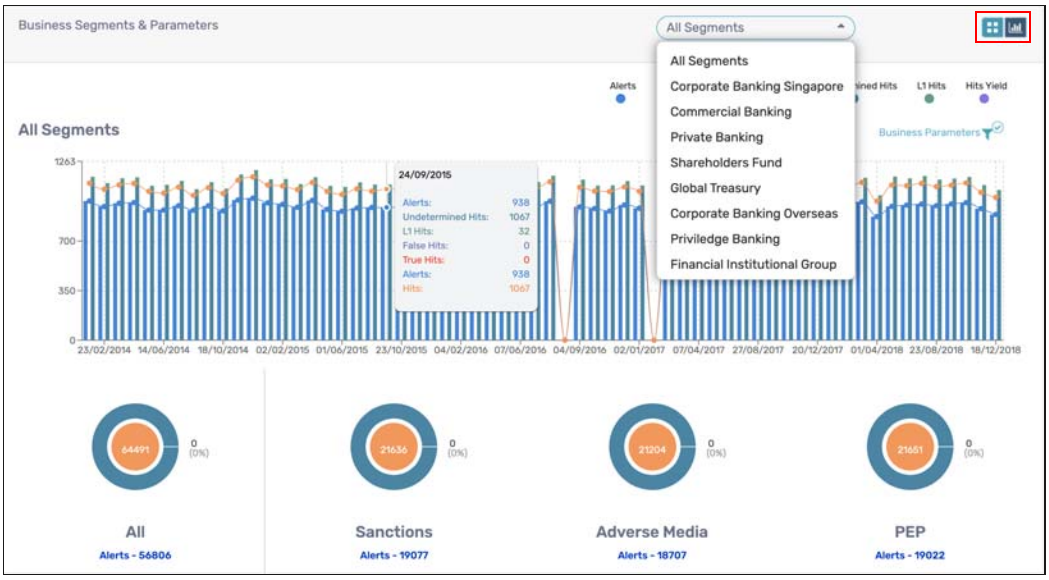
- Alerts
- Hits
- L1 Hits
- Undetermined Hits
- True Hits
- False Hits

User is also able to have three types of views such as **Graphical** and **Tabular** views as shown below. Click one of the options on the top right to toggle the views.

Graphical View

The graphical view is the default option. The filter option is available to view the information specific to a segment.

Figure 71 - Graphical View



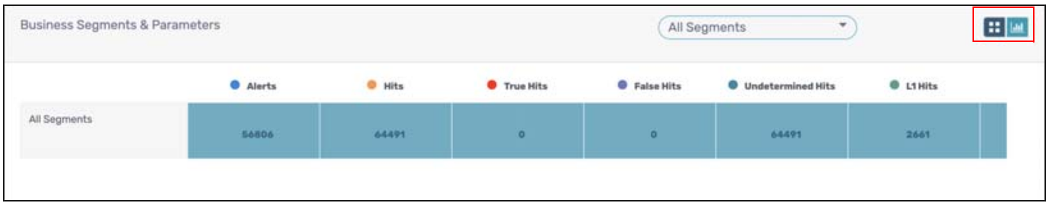
The graph shows the following, plotted on the axes:

- X-Axis — Date per the selected time frame
- Y-Axis — Statistics of alerts and hits

Tabular View

The widget also features a grid view to view the information in a tabular format.

Figure 72 - Grid View



Business Segments & Parameters

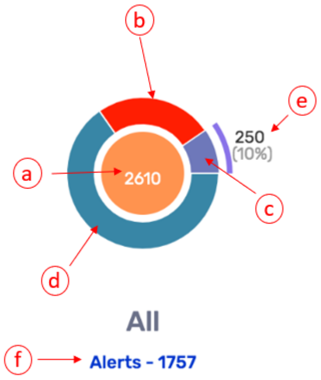
All Segments ▾

	● Alerts	● Hits	● True Hits	● False Hits	● Undetermined Hits	● L1 Hits
All Segments	56806	64491	0	0	64491	2661

Pie Charts

The different pie charts represent the statistics based on hit types - All, Sanctions, Adverse Media, PEP, and Others.

Figure 73 - Pie Chart



Each pie chart section as shown above:

a) Count of Total Hits for all or detection check filtered data.

b) Count of Total True Hits for all or detection check filtered data.

c) Count of Total False Hits for all or detection check filtered data.

d) Count of Total Undetermined Hits for all or detection check filtered data.

e) Hit yield of selected hit category section over total hits.
e.g. If user has selected the pie chart graph section of true hits then;
True Hit yield % = Total true hits for the selected graph only/
Total hits for the selected graph only.

Alert Count:

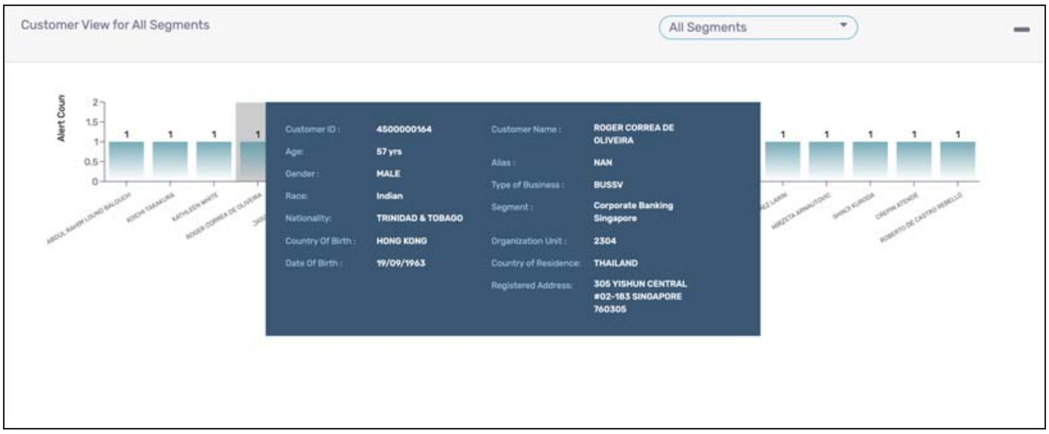
f) Total Alert count for all or detection check filtered data.

Customer View for All Segments

This view displays the top 20 customers based on the number of alerts over the selected time frame. See “Time Range Selection” on page 53. By default, all segments are displayed.

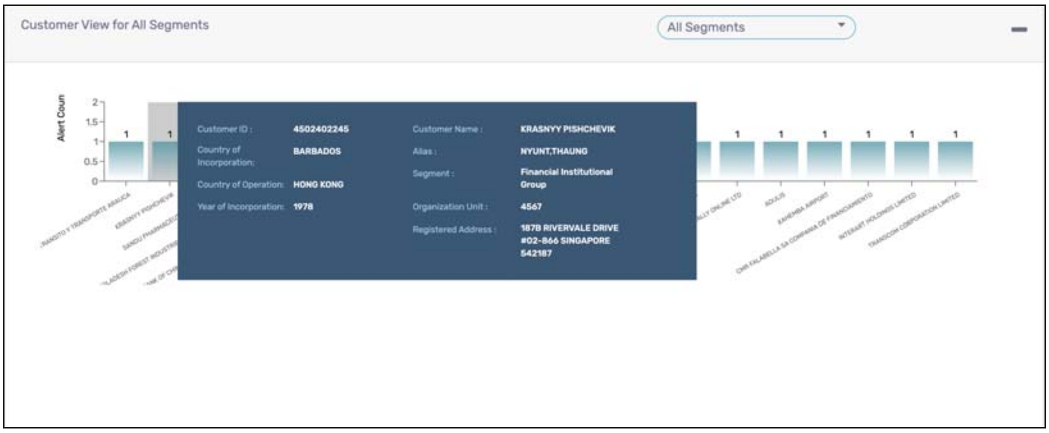
Mouse over the customer bar to produce a pop-up display of the profile of the customer.

Figure 74 - Customer View for All Segments (Individual)



Customer ID : 4500000164
Age: 57 yrs
Gender : MALE
Race: Indian
Nationality: TRINIDAD & TOBAGO
Country Of Birth : HONG KONG
Date Of Birth : 19/09/1963
Customer Name : ROGER CORREA DE OLIVEIRA
Alias : NAN
Type of Business : BUSSV
Segment : Corporate Banking Singapore
Organization Unit : 2304
Country of Residence: THAILAND
Registered Address: 305 YISHUN CENTRAL #02-183 SINGAPORE 760305

Figure 75 - Customer View for All Segments (Corporate)


Customer View for All Segments
All Segments
Alert Coun
2
1.5
1
0.5
0
1
KRASNYY PISHCHEVIK
SANDU PHARMACEU
ADULIS
KAHEMBA AIRPORT
CMR FALABELLA SA COMPANIA DE FINANCIAMIENTO
INTERART HOLDINGS LIMITED
TRANSCOM CORPORATION LIMITED

Customer ID : 4502402245
Country of Incorporation: BARBADOS
Country of Operation: HONG KONG
Year of Incorporation: 1978
Customer Name : KRASNYY PISHCHEVIK
Alias : NYUNT,THAUNG
Segment : Financial Institutional Group
Organization Unit : 4567
Registered Address : 1878 RIVERVALE DRIVE #02-866 SINGAPORE 542187

The graph shows the following, plotted on the axes:

- X-Axis — Customer Name
- Y-Axis — Alert Count

The user has the option to filter the information specific to a segment that they would focus on.

Listed below are customer profile details for individual and corporate:

Table 25 - Individual

Attributes	Description
Customer ID	Unique identifier for customer.
Customer Name	Name of customer.
Alias	List of alias names for customer.

Table 25 - Individual

Attributes	Description
Registered Address	Registered address of the customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Organization Unit	Organization unit of the bank to which the customer belongs.
Age	Age of customer.
Gender	Gender identity of customer.
Race	Race of customer.
Nationality	Citizenship of individual customer.
Country of Birth	Country of birth of individual customer.
Date of Birth	Date of birth of customer.
Type of Business	Occupation of the individual customer.
Country of Residence	Country of residence of individual.

Table 26 - Corporate

Attribute	Description
Customer ID	Unique identifier for customer.
Customer Name	Name of customer.
Alias	List of alias names for customer.
Registered Address	Registered address of the customer.
Segment	Name of the business segment of the bank to which the individual customer belongs.
Organization Unit	Organization unit of the bank to which the customer belongs.
Country of Incorporation	Country of incorporation of the corporate customer.
Country of Operation	Country of operation of the corporate customer.
Year of Incorporation	Year of incorporation for corporate customers.

Terminology

Abbreviations

The following terminologies are used in this document.

Term	Definition
AI	Artificial Intelligence
AML	Anti-Money Laundering
AMLS	Anti-Money Laundering Suite
CR	Credit
CRM	Customer Relationship Management
CDD	Customer Due Diligence
CDH	Cloudera Distribution Hadoop
CSV	Comma-Separated Values
DDL	Data Definition Language
DML	Data Manipulation Language
DR	Debit
FI	Financial Institutions
HDP	Hortonworks Distribution Hadoop
ID	Identification
KYC	Know Your Client
ML	Machine Learning
NS	Name Screening
PCA	Principal Component Analysis
POC	Proof of Concept
RM	Relationship Manager
RPC	Remote Procedure Call
SAR	Suspicious Activity Report
SIT	System Integration Testing
SLA	Service Level Agreement
SSO	Single Sign On
STR	Suspicious Transaction Report
SVD	Single Value Decomposition
UAT	User Acceptance Testing

UDF	User-Defined Functions
URI	Uniform Resource Identifier
TDSS	Tookitaki Data Science Studio
TM	Transaction Monitoring

Glossary

Term	Definition
Binning	Data binning is a way to group a number of more or less continuous values into a smaller number of bins.
Challenger	The model which is challenging (with respect to user defined performance metric) the current champion machine learning model after incremental data learning.
Champion	The current machine learning model used for prediction in solution.
Confusion Matrix	A confusion matrix is a table that is often used to describe the performance of a classification model (or classifier) on a set of test data for which the true values are known.
Correlation Matrix	A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used as a way to summarize data, as an input into a more advanced analysis and as a diagnostic for advanced analyses.
Dataframe	Dataframe is a 2-dimensional labeled data structure with columns of potentially different types such as a spreadsheet or SQL table.
Dataset	A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.
Dependent Variable	The column which is being predicted or leaned on.
Denormalize	Denormalization is a database optimization technique in which we add redundant data to one or more tables.
Evaluation Metric	The metric to define the performance metric of machine learning model.
F-Measure	The F-score (or F-measure) considers both the precision and the recall of the test to compute the score.
Feature	Columns used for machine learning model to learn patterns.
Machine Learning	Machine Learning (ML) is the scientific study of algorithms and statistical models that computer systems use to progressively imprive their performance on a specific task.
Majority Class	The majority class is simply having the greatest frequency in the class distribution of training examples reaching the leaf.
Minority Class	The minority class is simply having the least frequency in the class distribution of training examples reaching the leaf.

Normalize	In the simples cases, normalization of ratings means adjusting values measured on different scales to notionally common scale, often prior to averaging.
Oversampling	Oversampling is capable of improving resolution, reducing noise and can be helpful in avoiding aliasing and phase distortion by relaxing anti-aliasing filter performance requirements.
Pearson's Correlation Matrix	Correlation matrix between variables using Pearson's Correlation Coefficient.
Predictive Model	Predictive modeling is a process that uses data mining and probability to forecast outcomes. Each model is made up of a number of predictors, which are variables that are likely to influence future results. Once data has been collected for relevant predictors, a statistical model is formulated.
Recall	It is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.
Regular Expression	A sequence of symbols and characters expressing a string or pattern to be searched for within a longer piece of text.
Precision	Precision is the fraction of relevant instances among the retrieved instances.
Principal Component Analysis	A method of analysis which involves finding the linear combination of a set of variables that has maximum variance and removing its effects, repeating this successively.
Singular Value Decomposition	It is a generalization of the Eigen decomposition of a positive semidefinite normal matrix to any m*n matrix via an extension of polar decomposition.
Supervised Classification	Supervised learning is the data mining task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal).
Term Frequency-Inverse Document Frequency	It is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.
Undersampling	Undersampling in data analysis are techniques used to adjust the class distribution of a data set.
XGBoost	XGBoost is machine learning algorithm. It is an implementation of gradient boosted decision trees designed for speed and performance.