ThingWorx Machine Learning API Guide

All 51.0 Releases

Document Revision History

For a detailed list of all endpoint changes, see the API Updates section.

Revision Date	Description of Changes
December 31, 2015	• The original version of Profiles has been removed and all endpoints are using the enhanced search process allowing more direct profiles and more dynamic numeric value ranges to be found. In addition, Profiles are now Mutually Exclusive and allow larger datasets to be run without performance limitations. • New samples have been added to Predictive Scoring, Prescriptive Scoring, and Predictive Evaluation endpoints to show that results are now returned via a CSV file instead of a JSON file. Important Feature results are also available through a separate request for Predictive Scoring and Predictive Evaluation. ***Note:*** *Previously run scoring jobs will no longer be accessible and will need to be re-run. The application header will need to be set to unvalued or to "text/csv" in order to return results.* • Users can now terminate jobs so a new value of Terminated has been added to the list of possible status messages.
April 26, 2016	• Updated to remove references to Neuron, where appropriate. • Corrections to resultId fields everywhere they occur to ensure they are listed as integers, not strings. • Reorganization of the document sections, including: ○ Removed the old Enhancements section. ○ Moved the following sections further up to the top of the document: Datasets, Filters, Data Query. ○ Added a new API Updates section. ○ Generated the pdf with a navigation panel for easier access to specific material.

Table of Contents

About This Guide

This guide documents the use of ThingWorx Machine Learning APIs. It is intended for developers and end users who will be implementing machine learning-powered analytics into their business model through restful APIs.

The following are documentation conventions used throughout this guide.

- A parameter placeholder is denoted as ":" followed by a descriptive name
    - Placeholders should be replaced by the parameter value
    - Colons or descriptive names should be included in the request
- Sample requests and responses are provided throughout guide. Syntax has been changed for readability and snippets should be used as a reference only.

For each service, the URI will have *IP Address* listed within the call for easier readability. This placeholder will need to be replaced with user's specific IP address.

API Updates

Status Objects

Users can now terminate a single job or all jobs in the queue for a specified applicationId and applicationKey combination. A new value of Terminated has been added to the list of possible text status messages. Jobs that have been terminated will return a status of Terminated.

Clusters Service

Endpoint	Description of Changes
Cluster Scoring	The response variable **distance** has been replaced with **dateTime**, a timestamp associated with the record.

Profiles Service

All of the old Profiles endpoints have been removed and the Profiles V2 endpoints have been renamed Profiles. The Profiles service has also been updated so that:

- All endpoints use enhanced search procedures.
- Profiles are now mutually exclusive and can allow larger datasets to run without performance limitations

Endpoint	Description of Changes
List All Profiles Jobs	Removed "V2" from the endpoint name and the request URI.
Retrieve Specific Profiles Job	Removed "V2" from the endpoint name and the request URI.
Retrieve Configuration for Specific Profiles Job	Removed "V2" from the endpoint name and the request URI.
Submit Profiles Job	Removed "V2" from the endpoint name and the request URI.
Delete Profiles Result	Removed "V2" from the endpoint name and the request URI.
Retrieve Profiles Result	Removed "V2" from the endpoint name and the request URI.

Prediction Model Generation Service

Endpoint	Description of Changes
Prediction Model Generation endpoints	The following new prediction parameters have been added: • miThreshold • iterativeTrainingRecordSampleSize • iterativeTrainingImprovementPercentageThreshold • validationHoldout
	A set of Learning Technique arguments have been added under an **arg** parameter: • relativeMiThreshold • samplingStrategy (moved from its original position outside of the **learners** parameter)

Endpoint	Description of Changes
	• samplingValue • samplingFactor • layerCount • hiddenUnitPercentage • maxDepth • numberOfTrees • numberOfIterations **Note:** These learning techniques cannot be used in combination with the **learningComplexity** parameter. The two are mutually exclusive.

Scoring Service

To improve performance when scoring results for large datasets, response payloads for the Scoring service are now returned in CSV file format instead of JSON.

Endpoint	Description of Changes
Submit Predictive Scoring Job	Added an **Additional Features** list parameter to the post body of the request.
Retrieve Predictive Scoring Result	Removed **features** from the request parameters and updated the request URI. The Accept header **application/json** can no longer be used. The header value can either be left with no value or changed to **txt/csv**.
Retrieve Predictive Scoring Important Feature Result	New endpoint. Includes parameters for **features**, **importantFeatures**, and **predictedGoals**. The Accept header value can either be left with no value or set to **txt/csv**.
Submit Prescriptive Scoring Job	Added **maximize**, a Boolean parameter, to the post body of the request.
Retrieve Prescriptive Scoring Result	The Accept header **application/json** can no longer be used. The header value can either be left with no value or changed to **txt/csv**.
Retrieve Prescriptive Scoring Metadata Result	New endpoint. The Accept header value can either be left with no value or set to **txt/csv**.

Predictive Evaluation Service

To improve performance when evaluating results for large datasets, response payloads for the Scoring service are now returned in CSV file format instead of JSON.

Endpoint	Description of Changes
Submit Predictive Evaluation Job	Added an **Additional Features** list parameter to the post body of the request.
Retrieve Predictive Evaluation Result	New endpoint. Includes parameters for **features**, **importantFeatures**, and **predictedGoals**.

Endpoint	Description of Changes
	The Accept header value can either be left with no value or set to **txt/csv**.
Retrieve Predictive Evaluation Important Feature Result	New endpoint. The Accept header value can either be left with no value or set to **txt/csv**.

Datasets Service

Endpoint	Description of Changes
Upload Data	Some additional information is provided about the post body and response for multi-part file uploads.

Job Monitor Service

This new service has been added to enable monitoring the status of pending jobs, including jobs in both queued and running states. The service provides a set of response variables that includes a text status message, job progress, and information about the job itself (such as start and end times, run time, job type, and the dataset the job ran on).

This service also provides functionality for terminating jobs.

Endpoint	Description of Changes
Retrieve All Pending Jobs	New endpoint.
Retrieve Job Status	New endpoint.
Terminate Job	New endpoint

Overview

Welcome to ThingWorx Machine Learning (formerly Neuron), a platform for building smart software. ThingWorx ML enables you to easily and quickly embed AI-based technology into your product or service. It is exposed for software developers as a library of REST services for easy integration into your software application.

Although our name has changed, this document still provides information related to Neuron, such as server endpoints, headers, and other information. These are still valid and will be updated to our new name in the future.

Standard Headers

- Content-type: application/json

- Accept: application/json
  Note: This default for this header is JSON; however, when exporting CSV results, it will need to be updated to CSV.

- neuron-application-id: "my application id".
  This header is optional but recommended. Alternative is to use query parameter of applicationId to specify the applicationId.

- neuron-application-key: "my application key".
  This header is optional but recommended. Alternative is to use query parameter of applicationKey to specify the app key.

HTTP Status Codes

200 - OK
201 - Created
202 - Accepted
204 - No Content
403 - Unauthorized
404 - Not Found
410 - Gone (Expired)
500 - Internal Server Error
503 - Service unavailable

Standard Objects

Each job submission object includes three fields in its base class: goal (string), description (string), and exclusions (string[])

Each job configuration/status call and list contains startTime, endTime, runTime, queuedStartTime, and queuedDuration.

```
{
    "lastModified": "2015-09-01T20:53:53.202Z",
    "state": "COMPLETED",
    "description": null,
    "startTime": "2015-09-01T20:53:53.289Z",
    "endTime": "2015-09-01T20:54:48.519Z",
    "runTime": "0:00:55.230",
    "queuedStartTime": "2015-09-01T20:53:53.198Z",
    "queuedDuration": "0:00:00.091"
}
```

STATUS OBJECT

Each job submission will return an object that indicates the current status of each job in ThingWorx Machine Learning.

Variable	Description	DataType	Populated
uri	The location of the job status.	String	Y
resultId	The unique identifier for a job.	Integer	Y
message	A text status message indicating the current status of the job. Possible values include: RUNNING FAILED COMPLETED EXPIRED QUEUED TERMINATED	String	Y
progress	Indicates the job completion status. Currently only 0 or 100 are provided. In the future, more granular progress measure may be implemented.	Double	Y
messageInfo	A message to display extra system information, currently used only for failure messages.	String	N
startTime	Time when the job is picked up by a grid worker	String	Y
endTime	Time the job status changed to completed or failed.	String	Y
runTime	Total job run time, excluding queued time.	String	Y
queuedStartTime	The timestamp for the moment the job was posted.	String	Y
queuedDuration	Amount of time from job post to actual start time (moment it is picked up by a worker.)	String	Y

Retrieve the Status of a Specific Job

Request

URI: https://*ip address*/1.0/status/:resultId
Method: GET

Sample Response


```
{
    "resultId": 172256,
    "uri": "https://api-staging.coldlight.corp/1.0/status/172256",
    "progress": 100,
    "message": "COMPLETED",
    "messageInfo": null,
    "startTime": "2015-09-01T20:53:53.289Z",
    "endTime": "2015-09-01T20:54:48.519Z",
    "runTime": "0:00:55.230",
    "queuedStartTime": "2015-09-01T20:53:53.198Z",
    "queuedDuration": "0:00:00.091"
}
```

STANDARD RESULTS

The following variables are used across result services:

Parameter	Description	DataType	Required
resultId	Unique identifier for the result set stored for the job.	Integer	Y
lastModified	The date and time the job configuration parameters (described below) were modified.	DateTime	Y
dataSet	The name of the data set you are operating on.	String	Y
configuration			
goal	The dependent variable (feature) name for the analytic job you are performing (e.g., CustomerChurn, Customer Lifetime Revenue, etc.)	String	Y
description	A text description that can be used to provide additional details about a job.	String	N
filter	The name of a filter (this parameter is optional). Refer to the Filters section of this document for details about creating and using filters.	String	N
exclusionList | exclusions	A list of any features to be excluded from the analysis job. Either exclusionList or exclusions can be used currently.	StringList	N
startTime	The timestamp at the moment the job is picked up by a grid worker and is actively running.	String	
endTime	The timestamp at the moment the job status changes to COMPLETED or FAILED	String	
runTime	The duration of the submitted job from start time to completion.	String	
queuedStartTime	The timestamp for the moment the job was posted.	String	Y
queuedDuration	Amount of time from job post to actual start time (moment it is picked up by a worker.)	String	Y

DATASETS

ThingWorx Machine Learning allows users to upload and configure their own datasets at any time. These services allow you to configure features, descriptions, and fields related to a specific dataset, as well as truncate and delete data.

Optimization of the dataset allows a user to take a "snapshot" of their data set, allowing certain algorithms to run faster increasing job performance. Once optimized, a data set will remain optimized as long as no further data is uploaded and no field configurations are modified. Data set optimization returns a result ID for future job polling. For jobs such as prediction and clusters, optimizing the dataset results in a dramatic increase in job performance.

Variable	Description	DataType	Required
name	The unique identifier for a data set. Cannot contain spaces. Once created, this name cannot be changed.	String	Y
description	A friendly description or text display value for the data source to be used in other applications. This description can be changed as necessary, unlike Name.	String	N
optimized	A read-only flag displaying whether or not a particular data set has been optimized.	Boolean	N/A

List Data Sets

Request

URI: https://*ip address*/1.0/datasets
Method: GET

Sample Response: List Data Sets

```
[
    {
        "name": "CustomerDatabase",
        "description": "My Customer Database",
        "optimized": false
    },
    {
        "name": "FinancialDataSet",
        "description": "Finance Dataset through 2014",
        "optimized": true
    }
]
```

Create dataset

Request

URI: https://*ip address*/1.0/datasets
Method: POST

POST Body:

```
{
    "name": "CustomerDatabase",
    "description": "My Customer Database"
}
```

Response

HTTP Status Code: 201

```
{
    "name": "CustomerDatabase",
    "description": "My Customer Database",
  "optimized": false
}
```

Configure Data Set Description

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/
Method: POST

POST Body

```
{
    "name": "CustomerDatabase",
    "description": "Updated Customer Database"
}
```

Response

HTTP 204 NO CONTENT

Note: Only the dataset description can be modified. If the user attempts to change the dataset name, they will receive the following error message "Dataset name must match name of resource. It cannot be renamed."

Retrieve Data Set

Request

URI: https://*ip address*/1.0/datasets/:dataset_name
Method: GET

Sample Response: Retrieve Data Set

```
{
    "name": "CustomerDatabase",
    "description": "My Customer Database",
  "optimized": false
}
```

Configure Data Set

Variable	Description	DataType	Required
categories	An optional tag that identifies which groupings or categories to which a data element belongs.	StringList	N
fieldName	The unique identifier for a feature in a data set. Periods (.) are not allowed in fieldName.	String	Y
dataType	Identifies the data type for the feature. Acceptable values are: BOOLEAN, INTEGER, DOUBLE, STRING, ORDEREDSTRING. For ORDEREDSTRING, the order of the Values element will be used for interpretation of the order of the values for the feature.	String	Y
description	A friendly description or text display value for the feature to be used in other applications. This description can be changed as necessary unlike fieldName.	String	N
displayOnly	Allows data to be used for internal ThingWorx Machine Learning query and data result purposes but not within the learning algorithms as an input. Default is false.	Boolean	Y
lever	A flag which indicates whether (True) or not (False) this feature can be used as a dynamic value in simulations or prescriptive scoring (e.g, a "floating" variable).	Boolean	Y
objective	Identifies whether (True) or not (False) this feature can serve as a dependent variable or goal of an analysis.	Boolean	Y
values	For ORDEREDSTRING, the sequential list order of the Values element will be used for interpretation of the order of the values for the feature. For STRING, if Values list is not empty, any String values not explicitly contained in the Values list will be rejected by ThingWorx Machine Learning. If, however, values list is empty, all values are accepted by ThingWorx Machine Learning. For example, in the following Values list "values": ["Low", "Medium", "High"], if there were a value of "Very High" in a PostRequest, an error would be returned. However, if the Values list is "values": [ ] then any value will be accepted, including "Very High"	StringList	N
range	The acceptable range of values for the feature. If the actual value is not within the acceptable range, an error is returned. This is only allowed for Integer and Double dataTypes.		N
min	The minimum acceptable value for the feature.	Double	N
max	The maximum acceptable value for the feature	Double	N

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/configuration
Method: POST

POST Body:

```
[
    {
```

```
            "categories": [
                "Pharmacy Claims History"
            ],
            "dataType": "BOOLEAN",
            "description": "Y2_RxClass_MM_Sympathomimetics",
            "displayOnly": false,
            "fieldName": "Rx Class: MM Sympathomimetics [Y2]",
            "lever": false,
            "objective": false
        },
        {
            "categories": [
                "Health Risk Attributes"
            ],
            "dataType": "ORDEREDSTRING",
            "description": "PrevYear1_SMOKING",
            "displayOnly": false,
            "fieldName": "SMOKING [PrevYear1]",
            "lever": false,
            "objective": false,
            "values": [
                "NA",
                "Unknown",
                "Low",
                "Medium",
                "High"
            ]
        },
        {
            "categories": [
                "Health Risk Attributes"
            ],
            "dataType": "DOUBLE",
            "description": "T_PHA_Completion_Flag",
            "displayOnly": false,
            "fieldName": "HasPHA Completion ",
            "lever": false,
            "objective": true,
            "range": {
                "max": 1000,
                "min": 100
            },
            "values": [ ]
        }
    ]
```


Response

HTTP 201 CREATED

Configure Data Set Feature

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/configuration/:feature_name
Method: POST

POST Body


```
{
    "categories": [
        "Health Risk Attributes"
    ],
    "dataType": "ORDEREDSTRING",
    "description": "PrevYear1_SMOKING",
    "displayOnly": false,
    "fieldName": "SMOKING [PrevYear1]",
    "lever": false,
    "objective": false,
    "values": [
        "NA",
        "Unknown",
        "Low",
        "Medium",
        "High"
    ]
}
```


Response

HTTP 204 NO CONTENT

Retrieve Data Set Configuration

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/configuration
Method: GET

Sample Response: Retrieve Data Set Configuration


```
[
    {
        "categories": [
            "Health Risk Attributes"
        ],
        "dataType": "DOUBLE",
        "description": "T_PHA_Completion_Flag",
        "displayOnly": false,
        "fieldName": "HasPHA Completion ",
        "lever": false,
        "objective": true,
        "range": {
            "max": null,
            "min": null
        },
        "values": [ ]
    },
    {
        "categories": [
            "Health Risk Attributes"
        ],
        "dataType": "ORDEREDSTRING",
        "description": "PrevYear1_HbA1CTest",
```

```
            "displayOnly": false,
            "fieldName": "Hb A1 CTest [PrevYear1]",
            "lever": false,
            "objective": false,
            "range": {
                "max": null,
                "min": null
            },
            "values": [
                "Unknown",
                "Not Compliant",
                "Compliant"
            ]
        }
    ]
```


Upload Data

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/data
Method: POST

POST Body


```
    [
        {
            "identifier": "abc",
            "entry_dt": "2014-01-01T00:00:00.0000000",
            "Zip": "97034",
            "Gender": "F",
            "Age": 33,
            "Distance": 7.3,
            "Elevation": 972,
            "AvgPace_MinPerMile": 9.41
        }
    ]
```


Response

HTTP 204 NO CONTENT

Request (Multipart file upload)

URI: https://*ip address*/1.0/datasets/:dataset_name/data
Method: POST

POST Body (file upload)

Select *form-data* under POST BODY. In the Key field, type "*file*," and select "*File*" from the *Text vs File* dropdown. Remove the Content-Type header completely before posting the job. Choose a csv/gz file that matches the previously configured meta. Currently the file needs to be comma delimited, but ThingWorx Machine Learning can accept quotes. This upload will run as a regular job and you will get a response with a resultId for status checking.

Note: Remove the Application header when uploading a file. Only the Accept header is required.

Response

See standard "status" object.

Note: For Postgres-only installs (developer edition on-prem solution), this method is now supported in version 50. Users will receive an HTTP 404 NOT FOUND with error message "Streaming data to a non-redshift database is currently not supported for versions 48 and below.

Truncate Data Set

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/data
Method: DELETE

Response

HTTP 200 OK

Delete Data Set

Request

URI: https://*ip address*/1.0/datasets/:dataset_name
Method: DELETE

Response

HTTP 204 NO CONTENT

Note: This will delete everything related to the dataset including all results, filters, data configuration, and data.

Optimize Data Set

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/optimize
Method: POST

Sample Response: Submit Signals Job

See standard "status" object

FILTERS

Variable	Description	DataType	Required
dataSet	The identifier of the data set to which the Filter belongs.	String	Y
name	The unique identifier for the name of the Filter. Spaces are not allowed. Once created, "name" cannot be changed.	String	Y
description	A friendly description or text display value for the Filter to be used in other applications. This description can be changed as necessary unlike name.	String	N
filters	List of criteria that define the filter.	Object	
fieldName	The name of the field or feature.	String	Y
expression	The evaluation criteria for the feature. For more detailed information on the filter expression syntax,	String	Y
type	Indicates whether you want the feature/expression combination to be considered an Inclusion or Exclusion condition. Allowed values for "type" field are "INCLUDE", "EXCLUDE"	String	Y

List All Filters

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/filters
Method: GET

Sample Response: List All Filters


```
[
    {
        "dataSet": "has_-_All_Data_and_External_v2",
        "name": "Only Females",
        "description": "my description",
        "filters": [
            {
                "fieldName": "Member Gender",
                "expression": "[F]",
                "type": "INCLUDE"
            }
        ]
    },
    {
        "dataSet": "has_-_All_Data_and_External_v2",
        "name": "Only_Males",
        "description": "my description",
        "filters": [
            {
                "fieldName": "Member Gender",
                "expression": "[M]",
                "type": "INCLUDE"
            }
        ]
    }
]
```

Retrieve Specific Filter

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/filters/:filter_name
Method: GET

Sample Response: Retrieve Specific Filter


```
{
    "dataSet": "has_-_All_Data_and_External_v2",
    "name": "Only_Males",
    "description": "my description",
    "filters": [
        {
            "fieldName": "Member Gender",
            "expression": "[M]",
            "type": "INCLUDE"
        }
    ]
}
```


Create Filter

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/filters
Method: POST

POST Body:


```
{
    "dataSet": "has_-_All_Data_and_External_v2",
    "name": "Only_Males",
    "description": "my description",
    "filters": [
        {
            "fieldName": "Member Gender",
            "expression": "[\"M\"]",
            "type": "INCLUDE"
        }
    ]
}
```


Sample Response: Create Filter

HTTP 201


```
{
    "dataSet": "has_-_All_Data_and_External_v2",
    "name": "Only_Males",
    "description": "my description",
    "filters": [
        {
            "fieldName": "Member Gender",
            "expression": "[M]",
            "type": "INCLUDE"
```

```
            }
        ]
    }
```

Request Example 2:

URI: https://*ip address*/1.0/datasets/:dataset_name/filters
Method: POST

POST Body:

```
{
    dataSet: "Test_datasetl",
    name: "Age_0-50",
    description: "Example Filter",
    filters: [
        {
            fieldName: "Age",
            expression: "[{\"min\":0,\"max\":50}]",
            type: "INCLUDE"
        }
    ]
}
```

Sample Response: Create Filter

HTTP 201

```
{
    dataSet: "Test_datasetl",
    name: "Age_0-50",
    description: "Example Filter",
    filters: [
        {
            fieldName: "Age",
            expression: "[{\"min\":0,\"max\":50}]",
            type: "INCLUDE"
        }
    ]
}
```

Update Filter

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/filters/:filter_name
Method: POST

POST Body:

```
{
    "dataSet": "CustomerCareDatabase",
    "name": "Males Only",
    "description": "my description",
    "filters": [
        {
            "fieldName": "Gender",
```

```
                "expression": "[\"M\"]",
                "type": "INCLUDE"
            }
        ]
    }
```

Sample Response: Update Filter

HTTP 204

Delete Filter

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/filters/:filter_name
Method: DELETE

Response

HTTP 204

DATA QUERY

The Data Query services allow you to further query the distribution of data within a dataset.

Parameter	Description	DataType	Required
dataSet	The source data to be used for the analysis.	String	Y
goal	The topic, outcome or dependent variable of the analysis.	String	N
feature	A maximum of 1 feature is currently allowed for binned distributions.	String	N
filter	The unique identifier for the name of the Filter.	String	N
numbins	The number of bins your values will be grouped into. Only supported for binned distributions.	Integer	N

Response Variables

Variable	Description	DataType	Populated
values	The value of the feature, for example "Male" or "Female" for the Gender feature (note that this variable is only used for nominal/string or Boolean features, not for ordinals). Items in the value list will be typed according to their source type. You may have a mix of strings, booleans and doubles in the same list. In the case of a binned set of values [15-20], values serves as a friendly String description (not useful for sorting), it is recommended to use min and max for more direct interaction with the discrete range values themselves.	String	Y
count	The number of distinct records in the cluster that match the specified value. (e.g., the total number of Gender=Male records in the cluster, the total number of Gender=Female in the cluster, etc.)	Integer	Y
averageGoal	The average goal value for all of the records in the cluster that match the specific value. (e.g., the average Customer Churn rate for Gender=Male in the cluster). Average goal is useful for both ordinal and Boolean goals.	Double	Y
totalGoal	The total (or sum) goal value for all members in the cluster that match the specified value (e.g., the sum of all revenue for Gender=Male in the cluster). Total goal value is less useful for Boolean goals and more appropriate for ordinal goals.	Double	Y
min	Only available for binned distribution query. The minimum value of a binned range.	Double	Y (only for binned ranges)
max	Only available for binned distribution query. The maximum value of a binned range.	Double	Y (only for binned ranges)

Query Distribution of Data

There are 4 potential configurations of Request/Response for Distribution Queries (this also applies to Cluster distribution queries)

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Request				
Goal	Yes	--	Yes	--
Feature	--	Yes	Yes	--
Response				
Count	Yes	Yes	Yes	Yes
Values	--	Yes	Yes	--
Avg Goal	Yes	--	Yes	--
Total Goal	Yes	--	Yes	--

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/distribution?goal=:goal&feature=:feature1&feature=:feature2&filter=:filter_name
Method: GET

Sample Response: Query Distribution of Data


```
[
    {
        "values": [
            19.45,
            "feature2_value"
        ],
        "count": 195215,
        "averageGoal": 13524.1314,
        "totalGoal": 9593859221.31
    },
    {
        "values": [
            10.2452,
            "my value 2"
        ],
        "count": 234324,
        "averageGoal": 2342.1314,
        "totalGoal": 53141.31
    }
]
```


Binned Distribution Query

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/binned_distribution?goal=:goal&feature=:feature1&filter=:filter_name&numbins=:number_of_bins

Method: GET

Sample Response: Binned Distribution Query


```
[
    {
        "values": [
            "0 - 5"
        ],
        "min": 0,
        "max": 5,
        "count": 195215,
        "averageGoal": 13524.1314,
        "totalGoal": 9593859221.31
    },
    {
        "values": [
            "5.01 - 10"
        ],
        "min": 5.01,
        "max": 10,
        "count": 234324,
        "averageGoal": 2342.1314,
        "totalGoal": 53141.31
    }
]
```

SIGNALS

Signals are those attributes or Features in a data set that ThingWorx Machine Learning finds most useful to predict outcome or explain variance in outcome. Signals can be singular (one at a time) or combinations of variables. For example, both Age and Level of Education may be good individual signals (or predictors) when analyzed individually; however, Age and Level of Education are better predictors when combined together. Signals largely utilize an information gain and mutual information algorithm that identifies the amount of information shared between independent and dependent variables.

Parameter	Description	DataType	Required
maxAtATime	An integer that indicates how many features to analyze in combination when seeking predictive strength. For example, Age and Gender together are stronger predictors of Mortality Rate, whereas each individually is less predictive than the combination of the two. Acceptable values are 1 and 2.	Integer	N Default=2

List Signals Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals
Method: GET

Sample Response: List Signals Jobs


```
[
    {
        "resultId": 2345,
        "lastModified": "2014-07-16 01:01:00 AM",
        "dataSet": "CustomerCareDatabase",
        "configuration": {
            "goal": "CustomerChurn",
            "description": "Analysis of customer churn for east coast
customers with Product X",
            "filter": "EastCoastCustomers_ProductX",
            "exclusions": [
                "Region",
                "Season"
            ],
            "maxAtATime": 2
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
}

]
```

Retrieve a Specific Signals Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals/:result_id
Method: GET

Sample Response: Retrieve Specific Signals Job


```
{
    "resultId": 2345,
    "lastModified": "2014-07-16 01:01:00 AM",
    "dataSet": "CustomerCareDatabase",
    "configuration": {
        "goal": "CustomerChurn",
        "description": "Analysis of customer churn for east coast
customers with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "exclusions": [
            "Region",
            "Season"
        ],
        "maxAtATime": 2
    },
    "state": "COMPLETED",
    "description": "The job has successfully completed."
}
```


Retrieve Configuration for a specific Signals job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals/:result_id/configuration
Method: GET

Sample Response: Retrieve Configuration for Specific Signals Job


```
{
    "goal": "CustomerChurn",
    "description": "Analysis of customer churn for east coast customers with
Product X",
    "filter": "EastCoastCustomers_ProductX",
    "exclusions": [
        "Region",
        "Season"
    ],
    "maxAtATime": 2,
    "filterProportion": 0.1,
    "calculateHighLow": false
}
```

Submit Signals Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals
Method: POST

POST Body:

```
{
    "goal": "CustomerChurn",
    "description": "Analysis of customer churn for east coast customers
with Product X",
    "filter": "EastCoastCustomers_ProductX",
    "exclusions": [
        "Region",
        "Season"
    ],
    "maxAtATime": 2
}
```

Sample Response: Submit Signals Job

See standard "status" object

Delete Signals Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Signals Result

Variable	Description	DataType	Populated
signalSetInfo	A collection of signals, one for each defined in maxAtATime. In other words, if maxAtATime was specified as 3, then 3 signalSetInfo objects will be returned	Object[]	Y
featuresAtATime	The number of features in combination that were analyzed.	Integer	Y
signals			
mi	(Mutual information score). A score between 0 and 1 which indicates the mutual information calculation. Intuitively, mutual information measures the information that *X* and *Y* share (it measures how much knowing one of these variables reduces uncertainty about the other). Essentially how useful is an individual	Double	Y

	feature or combination of features together in predicting the value of another feature.		
pearsonCorrelation	The Pearson Correlation coefficient which is measure of the linear correlation (dependence) between the Feature and the dependent variable.	Double	Y
features			
name	The name of the feature.	String	Y
featuresAtATime	The sequence of the feature when analyzed in combination.	Integer	Y
informationGain	Information gain represents how much extra information one receives by knowing a given fact or set of facts. For example, suppose two equally strong baseball teams are playing with the chance of each winning being 50%. Then you have no information regarding the outcome. If I then tell you that it is the bottom of the ninth and team A is up 11-0, then I have given you have almost 1 bit of information gain (almost because it's not over till it's over).	Double	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/signals/:result_id/results
Method: GET

Sample Response (this example is showing maxAtATime = 1): Retrieve Signals Result

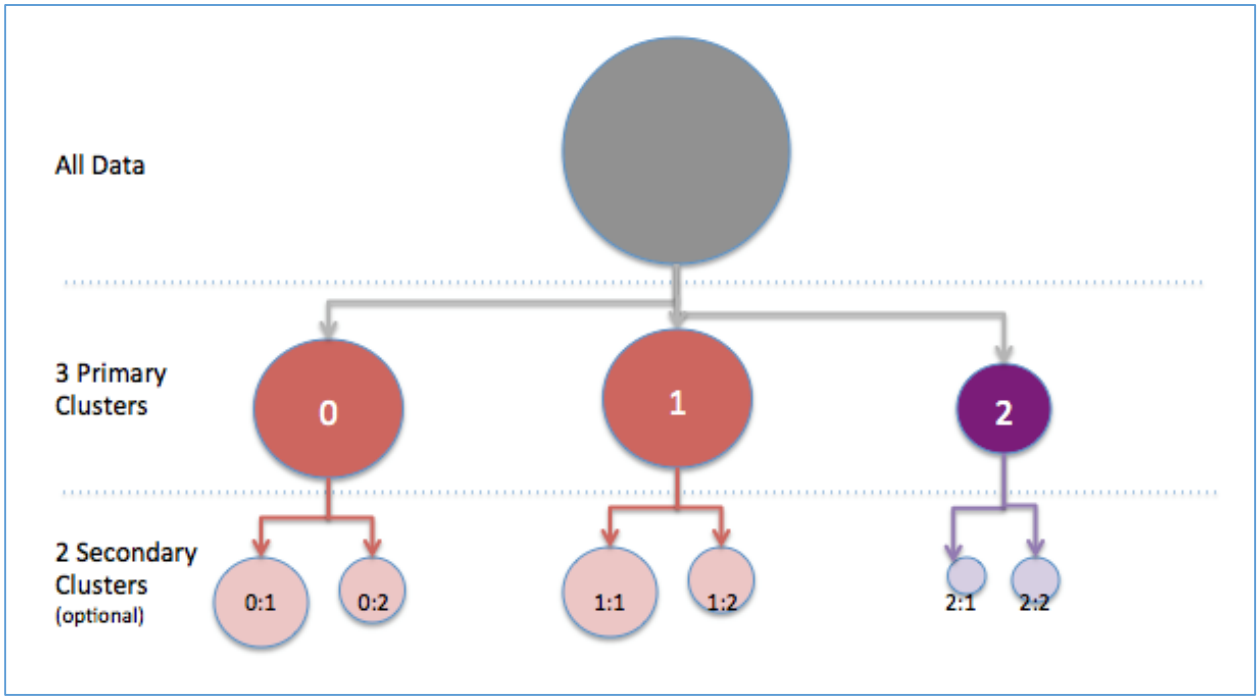

```
{
    "signalSetInfo": [
        {
            "featuresAtATime": 1,
            "signals": [
                {
                    "mi": 0.5138883646273023,
                    "pearsonCorrelation": 0,
                    "features": [
                        {
                            "name": "ZipCode",
                            "featuresAtATime": 1,
                            "informationGain": 0.5138883646273023
                        }
                    ]
                },
                {
                    "mi": 0.2494930508222044,
                    "pearsonCorrelation": -0.4553336742342987,
                    "features": [
                        {
                            "name": "Rx Adherence Risk",
                            "featuresAtATime": 1,
                            "informationGain": 0.2494930508222044
                        }
                    ]
                },
```

```
                    {
                        "mi": 0.1868925776841186,
                        "pearsonCorrelation": 0.3807231288894879,
                        "features": [
                            {
                                "name": "Respiratory Rate value",
                                "featuresAtATime": 1,
                                "informationGain": 0.1868925776841186
                            }
                        ]
                    }
                ]
            }
        ]
    }
```

CLUSTERS

Clusters partition all records in a given data set (or filtered group) into sub-segments, or Clusters, which are then organized and configured based on similar performance against the investigation Topic. For example, if your goal is Customer Churn, ThingWorx Machine Learning will cluster the customers attributes together based on similar churn rates but also on other factors that make certain records more similar to each other. For example, one Cluster of high churn customers may be made up predominantly of female executives on the East Coast while another cluster is made up predominantly of males who are middle managers on the West Coast. The greater the number of clusters, the more specific the definitions will be but the smaller the clusters will tend to be; the fewer the number of clusters, the more generic the definitions will be, but the clusters will tend to be larger.

Parameter	Description	DataType	Required
hierarchy	The number of clusters you want ThingWorx Machine Learning to generate. Value must be greater than or equal to 2. OPTION 1 (without secondary clusters) : [x] OPTION 2 (with secondary clusters): [x , y] *Example:* [3] returns 3 primary clusters, [5] would returns 5 primary clusters [3,2] returns 3 primary clusters each with 2 secondary clusters (or sub clusters) derived from the primary clusters as depicted in the diagram below.	IntegerList	Y



List All Clusters jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters
Method: GET

Sample Response: List All Clusters Jobs


```
[
    {
        "resultId": 2345,
        "lastModified": "2014-07-16 01:01:00 AM",
        "dataSet": "CustomerCareDatabase",
        "configuration": {
            "goal": "CustomerChurn",
            "description": "Analysis of customer churn for east coast
customers with Product X",
            "filter": "EastCoastCustomers_ProductX",
            "exclusions": [
                "Region",
                "Season"
            ],
            "hierarchy": [
                2,
                3
            ]
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
]
```


Retrieve a Specific Clusters Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id
Method: GET

Sample Response: Retrieve Specific Clusters Job


```
{
    "resultId": 2345,
    "lastModified": "2014-07-16 01:01:00 AM",
    "dataSet": "CustomerCareDatabase",
    "configuration": {
        "goal": "CustomerChurn",
        "description": "Analysis of customer churn for east coast
customers with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "exclusions": [
            "Region",
            "Season"
        ],
        "hierarchy": [
```

```
                2,
                3
            ]
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
```

Retrieve Configuration for Specific Clusters Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id/configuration
Method: GET

Sample Response: Retrieve Configuration for Specific Clusters Job

```
    {
        "goal": "CustomerChurn",
        "description": "Analysis of customer churn for east coast customers
    with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "exclusions": [
            "Region",
            "Season"
        ],
        "hierarchy": [
            2,
            3
        ]
    }
```

Submit Clusters Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters
Method: POST

POST Body:

```
    {
        "goal": "CustomerChurn",
        "description": "Analysis of customer churn for east coast customers
    with Product X",
        "filter": "EastCoastCustomers_ProductX",
     "exclusionLis" or "exclusions": [
            "Region",
            "Season"
        ],
        "hierarchy": [
            2,
            3
        ]
    }
```

Sample Response: Submit Clusters Job

See standard "status" object.

Delete Clusters Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Clusters Result

Variable	Description	DataType	Populated
clusterId	The 0-based identification number for the cluster.	Integer	Yes
density	Measure (from 0-1) of how closely related (or similar) are records in the cluster to each other	Double	Yes
features	A collection of features and measures that describe the records in a cluster.	Object	Yes
distinctiveness	Measure (from 0-1) of how significant a specific feature is in differentiating this cluster from the other clusters in the set.	Double	Yes
featureName	The name of the data feature.	String	Yes
populationCount	The number of unique records in the cluster	Integer	Yes
subClusterDescriptionModel	A nested structure of the cluster result set that encapsulates information for the sub-nodes of a hierarchical cluster.	Object	No

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id/results
Method: GET

Sample Response: Retrieve Cluster Results


```
{
    "clusters": [
        {
            "clusterId": "0",
            "density": 0.8243708829326337,
            "features": [
                {
                    "distinctiveness": 0.05529571708685852,
                    "featureName": "Unemployment Status"
                },
                {
                    "distinctiveness": 0.04750029686892938,
```

```
                        "featureName": "Home Ownership Status"
                    }
                ],
                "populationCount": 7200,
                "subClusterDescriptionModel": null
            },
            {
                "clusterId": "1",
                "density": 0.8244831646680024,
                "features": [
                    {
                        "distinctiveness": 0.534960783727653,
                        "featureName": "Account Delinquency Status"
                    }
                ],
                "populationCount": 25505,
                "subClusterDescriptionModel": null
            }
        ]
    }
```


Query a Distribution Within a Cluster

Allows you to query any attribute in the data set for the population of a specific cluster (for example, AverageAge for the cluster, Distribution of population by State).

Variable	Description	DataType	Populated
clusterId	The cluster or cluster path that the record is closest to. In a cluster model with a single level, it will be the unique clusterId. In the case of a hierarchical cluster set, it will be the full path. For example, 2 clusters job, for 0-based, the clusterId would be either 0 or 1, but for 1-based, the clusterId would be 1 or 2.	String	Y
values	The value of the feature, for example "Male" or "Female" for the Gender feature (note that this variable is only used for nominal/string or Boolean features, not for ordinals).	String	Y
count	The number of distinct records in the cluster that match the specified value. (e.g., the total number of Gender=Male records in the cluster, the total number of Gender=Female in the cluster, etc.)	Integer	Y
averageGoal	The average goal value for all of the records in the cluster that match the specific value. (e.g., the average Customer Churn rate for Gender=Male in the cluster). Average goal is useful for both ordinal and Boolean goals.	Double	Y
totalGoal	The total (or sum) goal value for all members in the cluster that match the specified value (e.g., the sum of all revenue for Gender=Male in the cluster). Total goal value is less useful for Boolean goals and more appropriate for ordinal goals.	Double	Y

There are 3 potential configurations of Request/Response for Cluster query results (this also applies to general distribution query results as well).

	Scenario 1	Scenario 2	Scenario 3
Request			
Goal	Yes	--	Yes
Feature	--	Yes	Yes
ClusterId	Yes	Yes	Yes
Response			
Count	Yes	Yes	Yes
Values	--	Yes	Yes
Avg Goal	Yes	--	Yes
Total Goal	Yes	--	Yes

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id/distribution?clusterId=:clusterid&feature=:feature

Sample Response (Scenario 1): Query a Distribution Within Cluster


```
[
    {
        "count": 25000,
        "averageGoal": 15.10,
        "totalGoal": 377500.00
    }

]
```


Sample Response (Scenario 2): Query a Distribution Within Cluster


```
[
    {
        "values": [
            "Male"
        ],
        "count": 25000

    },
    {
        "values": [
            "Female"
        ],
        "count": 3000
    }
]
```


Sample Response (Scenario 3): Query a Distribution Within Cluster


```
[
    {
        "values": [
            "Male"
        ],
        "count": 25000,
        "averageGoal": 15.10,
        "totalGoal": 377500.00
```

```
        },
        {
            "values": [
                "Female"
            ],
            "count": 3000,
            "averageGoal": 25.50,
            "totalGoal": 76500.00
        }
    ]
```


Cluster Scoring

Allows a user to determine which cluster a new records is closest to.

Variable	Description	DataType	Populated
identifier	The identifier for the record that was scored.	Integer	Yes
clusterId	The cluster or cluster path that the record is closest to. In a cluster model with a single level, it will be the unique clusterId. In the case of a hierarchical cluster set, it will be the full path.	String	Yes
dateTime	The timestamp associated with the record.	String	Yes

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/clusters/:result_id/scores/{identifier}
Method: GET

Response


```
    {
        "identifier": "12345",
        "clusterId": "1:2",
        "dateTime": "11/13/2015"
    }
```

PROFILES

Profiles are useful for finding pockets of high performers, higher risk, low performers, etc. To identify these performers, the Profiles service searches a dataset and effectively moves the average or ordinary records out of the way to focus on the margins (exceptionally high or low outcomes). However, this activity is constrained by the fact that Profiles must meet a minimum threshold (although it can be varied) which prevents identifying Profiles that describe only a single record or very small number of records, since these are likely outliers. Once a profile is found, all the records contained within it are excluded from subsequent profiles.

Profiles enable ThingWorx Machine Learning to find more discrete profiles through a refined search process that is able to find more direct profiles and more dynamic numeric value ranges. In addition to finding more discrete profiles, Profiles returns profiles that are mutually exclusive, or grouping records into a single profile, allowing users to better target specific populations without the concern of overlapping records.

Parameters	Description	DataType	Required
maxDepth	Indicates the number of levels deep a profile search can go before it stops searching. For example, 3 levels deep would mean that ThingWorx Machine Learning can create paths that consider 3 factors concurrently, while 4 levels deep means that it can go an additional level (or one additional feature deeper in its search). Acceptable values are between 1 and 4.	Integer	Y
minPopulationPercentage	Indicates the % of the audience population threshold that must be met to return a population. If the minimum threshold is not met then ThingWorx Machine Learning will not return a profile for a given combination of factors. For example this would enable you to return profile results where any given profile represents at least 2% of the overall population. This value must be greater than 0 and less than 1.	Double	Y
maximize	Indicates to search for profiles that achieve the highest (true) or lowest (false) outcome for a profile.	Boolean	Y

List All Profiles Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles
Method: GET

Sample Response: List All Profiles Jobs


```
{
    "resultId": 172263,
    "dataSet": "Test_dataset",
```

```
        "configuration": {
            "maxDepth": 4,
            "maximize": true,
            "filter": null,
            "goal": "Customer_Churn",
            "minPopulationPercentage": 0.025,
            "description": "All Data",
            "exclusions": [ ]
        },
        "lastModified": "2015-09-02T16:33:12.528Z",
        "state": "COMPLETED",
        "description": null,
        "startTime": "2015-09-02T16:33:12.63Z",
        "endTime": "2015-09-02T16:52:06.174Z",
        "runTime": "0:18:53.544",
        "queuedStartTime": "2015-09-02T16:33:12.521Z",
        "queuedDuration": "0:00:00.109"
    }
```


Retrieve Specific Profiles Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles/:result_id
Method: GET

Sample Response: Retrieve Specific Profiles Job


```
    {
        "resultId": 172263,
        "dataSet": "Test_dataset",
        "configuration": {
            "maxDepth": 4,
            "maximize": true,
            "filter": null,
            "goal": "Customer_Churn",
            "minPopulationPercentage": 0.025,
            "description": "All Data",
            "exclusions": [ ]
        },
        "lastModified": "2015-09-02T16:33:12.528Z",
        "state": "COMPLETED",
        "description": null,
        "startTime": "2015-09-02T16:33:12.63Z",
        "endTime": "2015-09-02T16:52:06.174Z",
        "runTime": "0:18:53.544",
        "queuedStartTime": "2015-09-02T16:33:12.521Z",
        "queuedDuration": "0:00:00.109"
    }
```


Retrieve Configuration for Specific Profiles Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles/:result_id/configuration
Method: GET

Sample Response: Retrieve Configuration for Specific Profiles Job


```
{
    "goal": "Customer_Churn",
    "description": "All Data",
    "filter": null,
    "exclusions": [ ],
    "maxDepth": 4,
    "minPopulationPercentage": 0.025,
    "maximize": true
}
```


Submit Profiles Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles
Method: POST

POST Body:


```
{
    "maxDepth": 4,
    "maximize": true,
    "filter": null,
    "goal": "Customer_Churn",
    "minPopulationPercentage": 0.025,
    "description": "All Data",
    "exclusions": []
}
```


Sample Response: Submit Profiles Job

See standard "status" object

Delete Profiles Result

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Profiles Result

Variable	Description	DataType	Populated
totalPopulationCount	Total number of records in a population matching all conditions in the profile. (This is the same result as the last condition in a profile).	Double	Y
conditions			

populationCount	The total number of unique records matching this condition and any prior conditions in this profile condition set.	Integer	Y
populationPercentage	The % of records matching this condition and any prior conditions in this profile condition set.	Double	Y
averageGoal	The average goal value for records matching this condition and any prior conditions in this profile condition set.	Double	Y
featureName	The name of the field or feature that describes the condition in the set. (e.g., Gender, Age)	String	Y
values	The value(s) of the field or feature that describes the condition in the set. (e.g., Gender=Male, Age=18-25 or 85-90)	StringList	Y
evaluation	The z-score of all the records that lie within this profile. This value can be positive or negative and shows the signed distance from the population mean in units of standard deviation. A positive evaluation value represents a profile mean that is greater than the population mean, while a negative value represents a mean less than the population mean.	Integer	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/profiles/:result_id/results
Method: GET

Sample Response: Retrieve Profiles Result


```
{
    "totalPopulationCount": 1000000,
    "profiles": [
        {
            "conditions": [
                {
                    "featureName": "HyperTensionFlag",
                    "values": [
                        "1"
                    ],
                    "evaluation": 33.88057494845566,
                    "populationCount": 237262,
                    "populationPercentage": 0.237262,
                    "averageGoal": 0.007557046640422824
                },
                {
                    "featureName": "UW_Exclusion_Flag",
                    "values": [
                        "0"
                    ],
                    "evaluation": 57.944905321099306,
                    "populationCount": 104204,
                    "populationPercentage": 0.104204,
                    "averageGoal": 0.014020575025910713
```

```
                    },
                    {
                        "featureName": "NewHypertensionDiagnosis",
                        "values": [
                            "0"
                        ],
                        "evaluation": 58.72513886276003,
                        "populationCount": 84756,
                        "populationPercentage": 0.084756,
                        "averageGoal": 0.015326348577091886
                    },
                    {
                        "featureName": "Prev_90Days_OtherDrugAbuseFlag_1",
                        "values": [
                            "0"
                        ],
                        "evaluation": 59.71872635376868,
                        "populationCount": 82658,
                        "populationPercentage": 0.082658,
                        "averageGoal": 0.015679063127585956
                    }
                ]
            }
        ]
    }
```

PREDICTION MODEL GENERATION

When analyzing any Group or any Topic, ThingWorx Machine Learning will automatically build and refine a prediction model using a variety of machine learning, Artificial Intelligence, and statistical modeling techniques and approaches to achieve the best results, also called an ensemble learning process. ThingWorx Machine Learning self-validates its predictions by putting aside a relevant set of data (holdout set) to verify against. Prediction accuracy is qualified on several performance metrics, primarily RMSE and Pearson Correlation. Once Prediction Models are created, they are then consumed by Scoring functions.

Parameter	Description	DataType	Required
learningComplexity	The depth of learning that you would like ThingWorx Machine Learning to perform. Acceptable values range from 1 to 5. Level 1 is the simplest and has the fastest learning time. The amount of time to completion increases substantially as the value increases but the higher the value, the deeper the learning performed on the dataset. **Note**: If using learningComplexity, do not add any of the training parameters below. Training parameters and learningComplexity are mutually exclusive.	Integer	N
ensembleTechnique	The ensemble learning technique used by ThingWorx Machine Learning to minimize prediction error. Possible values are "ELITE_AVERAGE", "BEST", "AVERAGE" or "SOLOIST"	String	N
comparisonMetric	Sets how ThingWorx Machine Learning compares sets of learners. Available choices include: "PEARSONS" (Pearson's Correlation), "RMSE" (Root Mean Square Error), and "ROC" (Receiver Operating Characteristic area interpolated from a single ROC Point; can only be used with boolean goals). Default value is PEARSONS.	String	N
miThreshold	Allows the user to select the threshold for which features are possible for inclusion based on their signal strength. Must be greater than or equal to 0. Default miThresold = 0.01. Lower thresholds mean more features are added into the model while higher thresholds mean only higher quality features are selected for inclusion into the model.	Double	N
iterativeTrainingRecordSampleSize	The number of records added with every iteration of training. Must be greater than 0. Default value is 8192.	Integer	N

iterativeTrainingImprovement PercentageThreshold	The Threshold value used to determine when an optimal training size has been reached. Must be between 0 and 1, exclusive to inclusive. Default value is 0.001 (0.1%).	Double	N
samplingStrategy	**Deprecated**. The Sampling Strategy parameter has been relocated under learners/args (see below).	String	N
validationHoldout	The percentage of data that is held back and used to validate the prediction model. Acceptable values are from 0 to 1, inclusive to exclusive. Default value is 0.2; i.e. 20% of data is used for benchmarking the performance of the newly trained model.	Double	N
learners	A collection of learners or modeling techniques can be used when building a prediction model. You can include one or more of the learning techniques, but you must have at least one defined.	StringList	N
learningTechnique	Indicates the specific learning techniques to be included in the array of learners. Available learners include: LINEAR_REGRESSION, LOGISTIC_REGRESSION (Recommended to only be used with Boolean goals), DECISION_TREE, RANDOM_FOREST, GRADIENT_BOOST, and NEURAL_NET.	String	N
args	LearningTechnique dependent additional parameters. An argument (arg) not supported by the selected learningTechnique will be ignored. An unexpected or misspelled arg will throw an exception.		N
relativeMiThreshold	All techniques. A fraction of the above miThreshold for the specific learning technique to use. Must be between 0 and 1, inclusively. **Default value is 1.** Example: If miThreshold is set to 0.8 and relativeMiThreshold is set to 1, the learningTechnique will train with an miThreshold of 0.8 * 1 = 0.8.	Double	N
samplingStrategy	All techniques. The Sampling Strategy used. Possible values are "SAMPLE_WITH_REPLACEMENT", "SAMPLE_WITHOUT_REPLAC EMENT", "NONE", "UPSAMPLE_AND_SAMPLE_WITHOUT_REPLACEMENT", "DOWNSAMPLE_AND_SAMPLE_WITHOUT_REPLACEMENT". Upsampling and Downsampling require additional parameters	String	N

	(see below). **Default value is "SAMPLE_WITHOUT_REPLAC EMENT"**		
samplingValue	All techniques, samplingStrategy="UPSAMPLE_ AND_SAMPLE_WITHOUT_REP LACEMENT" and "DOWNSAMPLE_AND_SAMPL E_WITHOUT_REPLACEMENT", boolean goal only. Determines which value is going to become more populous. This field is only required if your goal is boolean and you are using the above two samplingStrategies. For UPSAMPLE_AND_SAM PLE_WITHOUT_REPLACE MENT, duplicate all records equal to samplingValue by an amount samplingFactor . For DOWNSAMPLE_AND_SAM PLE_WITHOUT_REPLACE MENT, keep only samplingF actor * 100% of records not equal to samplingValue	Boolean	Y*
samplingFactor	All techniques, samplingStrategy="UPSAMPLE_ AND_SAMPLE_WITHOUT_REP LACEMENT" and "DOWNSAMPLE_AND_SAMPL E_WITHOUT_REPLACEMENT", boolean goal only. For UPSAMPLE_AND_SAM PLE_WITHOUT_REPLACE MENT, the number specified is the number of times a record with the desired samplingValue will be duplicated. It can be an integer or a double, although doubles will be rounded to integers. For example, passing in 3.1 will result in the value being rounded to 3 and all records with goal values equal to sa mplingValue are changed to 3 identical records. Other records are unchanged. Valid values are 0.5 (rounds to 1) or greater. Default value is 5. For DOWNSAMPLE_AND_ SAMPLE_WITHOUT_REPL ACEMENT, the number specified is the percentage of records that are not the s amplingValue that will be kept. For example, passing in 0.3 means that 30% of records with a goal value not equal to samplingValue are kept while 100% of records with a goal equal to samplingValue are kept. Valid values are greater than 0 and less than or equal to 1. Default value is 0.5	Double	N

layerCount	NEURAL_NET. The number of total layers to use in the Neural Network. Acceptable values are 2, 3, and 4. Default value is 3.	Integer	N
hiddenUnitPercentage	NEURAL_NET. The percentage of the # of input nodes to use in each hidden layer. Must be between 0 and 1, exclusive to inclusive. Default value is 0.2. Example: If the default value is used and the input layer has 10 nodes, each hidden layer will contain 10 * 0.2 = 2 hidden nodes.	Double	N
maxDepth	DECISION_TREE, RANDOM_FOREST. The maximum tree depth. Must be greater than 0. DECISION_TRE E default value is 12. RANDOM_FOREST default value is 3.	Integer	N
numberOfTrees	RANDOM_FOREST. The number of trees to use. Must be greater than 1. Default value is 25.	Integer	N
numberOfIterations	GRADIENT_BOOST. The number of iterations to use. Must be greater than 0. Default value is 100.	Integer	N

List All Prediction Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction
Method: GET

Sample Response: List All Prediction Jobs


```
[
    {
        "resultId": 76539,
        "dataSet": "coldlight!!test_dataset_clinical_bool_2",
        "configuration": {
            "filter": null,
            "goal": "Hospital Readmission",
            "learners": [
                {
                    "learningTechnique": "NEURAL_NET",
                    "args": { }
                },
                {
                    "learningTechnique": "GRADIENT_BOOST",
                    "args": { }
                },
                {
                    "learningTechnique": "DECISION_TREE",
                    "args": { }
                },
                {
                    "learningTechnique": "LINEAR_REGRESSION",
                    "args": { }
                }
```

```
                ],
                "samplingStrategy": "SAMPLE_WITHOUT_REPLACEMENT",
                "iterativeTrainingRecordSampleSize": 8192,
                "ensembleTechnique": "ELITE_AVERAGE",
                "description": "All Data",
                "iterativeTrainingImprovementPercentageThreshold": 0.001,
                "exclusions": [ ],
                "miThreshold": 0.3
            },
            "lastModified": "2015-07-10T20:49:36.474Z",
            "state": "COMPLETED",
            "description": null
        }
    ]
```


Retrieve Specific Prediction Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction/:result_id
Method: GET

Sample Response: Retrieve Specific Prediction Job


```
    {
        "resultId": 76539,
        "dataSet": "coldlight!!test_dataset_clinical_bool_2",
        "configuration": {
            "filter": null,
            "goal": "Hospital Readmission",
            "learners": [
                {
                    "learningTechnique": "NEURAL_NET",
                    "args": { }
                },
                {
                    "learningTechnique": "GRADIENT_BOOST",
                    "args": { }
                },
                {
                    "learningTechnique": "DECISION_TREE",
                    "args": { }
                },
                {
                    "learningTechnique": "LINEAR_REGRESSION",
                    "args": { }
                }
            ],
            "samplingStrategy": "SAMPLE_WITHOUT_REPLACEMENT",
            "iterativeTrainingRecordSampleSize": 8192,
            "ensembleTechnique": "ELITE_AVERAGE",
            "description": "All Data",
            "iterativeTrainingImprovementPercentageThreshold": 0.001,
            "exclusions": [ ],
            "miThreshold": 0.3
        },
```

```
    "lastModified": "2015-07-10T20:49:36.474Z",
    "state": "COMPLETED",
    "description": null
}
```


Retrieve Configuration for Prediction Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction/:result_id/configuration
Method: GET

Sample Response: Retrieve Configuration for Prediction Job


```
{
    "goal": "Hospital Readmission",
    "description": "All Data",
    "filter": null,
    "exclusions": [ ],
    "learningComplexity": null,
    "ensembleTechnique": "ELITE_AVERAGE",
    "learners": [
        {
            "learningTechnique": "NEURAL_NET",
            "args": { }
        },
        {
            "learningTechnique": "GRADIENT_BOOST",
            "args": { }
        },
        {
            "learningTechnique": "DECISION_TREE",
            "args": { }
        },
        {
            "learningTechnique": "LINEAR_REGRESSION",
            "args": { }
        }
    ],
    "miThreshold": 0.3,
    "iterativeTrainingRecordSampleSize": 8192,
    "iterativeTrainingImprovementPercentageThreshold": 0.001,
    "samplingStrategy": "SAMPLE_WITHOUT_REPLACEMENT"
}
```


Submit Prediction Model Generation Job

Request 1

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction
Method: POST

POST Body:


```
{
```

```
        "ensembleTechnique": "ELITE_AVERAGE",
        "learners": [
            {
                "learningTechnique": "NEURAL_NET",
                "args": { }
            },
            {
                "learningTechnique": "GRADIENT_BOOST",
                "args": { }
            },
            {
                "learningTechnique": "DECISION_TREE",
                "args": { }
            },
            {
                "learningTechnique": "LINEAR_REGRESSION",
                "args": { }
            }
        ],
        "description": "All Data",
        "iterativeTrainingRecordSampleSize": 8192,
        "exclusions": [ ],
        "filter": null,
        "goal": "Hospital Readmission",
        "miThreshold": 0.3
    }
```

Sample Response: Submit Prediction Model Generation Job

See standard "status" object

Request 2: Boolean model

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction
Method: POST

POST Body:

```
    {
        "ensembleTechnique": "ELITE_AVERAGE",
        "learners": [
            {
                "learningTechnique": "NEURAL_NET",
                "args": { }
            },
            {
                "learningTechnique": "GRADIENT_BOOST",
                "args": { }
            },
            {
                "learningTechnique": "DECISION_TREE",
                "args": { }
            },
            {
                "learningTechnique": "LOGISTIC_REGRESSION",
                "args": { }
            }
```

```
        ],
        "description": "All Data",
        "iterativeTrainingRecordSampleSize": 1024,
        "exclusions": [ ],
        "filter": null,
        "comparisonMetric": "ROC",
        "goal": "WillBuyMyProduct_Flag",
        "miThreshold": 0.2
    }
```


Sample Response: Submit Prediction Model Generation Job

See standard "status" object

Delete Prediction Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Describe Prediction Model Result

Variable	Description	DataType	Populated
numberOfRecords	The total number of records used in building the prediction model	Integer	Y
rmse	The Root Mean Square Error (RMSE) for the entire data set.	Double	Y
pearsonCorrelation	The Pearson Correlation Coefficient for the entire data set.	Double	Y
binErrors	A container of binned range of goal values that are helpful in determining model performance within specific data sub-segments. A model may perform very well on certain data ranges and poorly on others. 10 bins (evenly distributed) are returned for Ordinal goals; 2 bins (0,1) are returned for Boolean goals.	Collection	Y
binRange	Indicates the range of goal values being described / validated. [x,y] where 'x' is the starting value for the range and 'y' is the ending value for the range. First value in the list represents the minimum value of the bin range; the second value	DoubleList	Y

	represents the maximum value of bin range.		
numberOfRecords	The number of records or examples that fall into a given binRange.	Integer	Y
rmse	The Root Mean Square Error (RMSE) for each data bin.	Double	Y
pearsonCorrelation	The Pearson Correlation Coefficient for each data bin.	Double	Y
modelAccuracy	A measure of how well the model trained, between 0 and 1.	Double	Y
validationStatistics	A container of various statistics calculated from the scores produced on the validation set. These statistics are meant to show a quick overview of how well the model trained.	DoubleList	Y
count	The total number of records in the validation set.	Integer	Y
mean	The Average of all the validation set scores.	Double	Y
variance	The Variance of all the validation set scores.	Double	Y
standardDeviation	The Standard Deviation of all the validation set scores.	Double	Y
coefficientVariation	The Coefficient of Variation of all the validation set scores.	Double	Y
sumWeights	The sum of the Weights of all the validation set scores. (Weights should be one, so total should be equal to 'count')	Double	Y
sumObservations	The sum of all the validation set scores.	Double	Y
correctedSumSquares	The corrected Squared Sum of all the validation set scores. (Score - mean)^2	Double	Y
uncorrectedSumSquares	The uncorrected Squared Sum of all the validation set scores. (Score)^2	Double	Y
standardErrorMean	The Standard Error of the Mean of all the validation set scores.	Double	Y
skewness	The Skewness of all the validation set scores.	Double	Y
kurtosis	The Kurtosis of all the validation set scores.	Double	Y
percentilesHistogram	A container for the major percentiles of scores produced from the validation set. This histogram provides a quick overview of the distribution of scores.	DoubleList	Y

percentileN	The Nth percentile of the validation set score distribution	Double	Y
trainingParameters			
ensembleTechnique	The ensemble learning technique used by ThingWorx Machine Learning to minimize prediction error. Possible values are "ELITE_AVERAGE", "BEST", "AVERAGE" or "SOLOIST".	String	Y
learners	A collection of learners or modeling techniques can be used when building a prediction model. You can include one or more of the learning techniques, but you must have at least one defined.	Object	Y
learningTechnique	Indicates the specific learning techniques to be included in the array of learners. Available learners include: LINEAR_REGRESSION, LOGISTIC_REGRESSION (Recommended to only be used with Boolean goals), DECISION_TREE, NEURAL_NET (Neural networks), GRADIENT_BOOST	String	Y
miThreshold	Allows the user to select the threshold for which features are possible for inclusion based on their signal strength. Default miThresold = 0.01.	Double	N
iterativeTrainingRecordSampleSize	The number of records added with every iteration of training	Integer	Y
iterativeTrainingImprovement PercentageThreshold	The Threshold value used to determine when an optimal training size has been reached	Double	Y
samplingStrategy	The Sampling Strategy used	String	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction/:result_id/results
Method: GET

Sample Response: Describe Prediction Model Result


```
{
    "evaluationSummary": {
        "numberOfRecords": 6570,
        "rmse": 0.4755834356469914,
        "pearsonCorrelation": 0.09177739457607152,
        "binErrors": [
            {
                "numberOfRecords": 5089,
```

```
                "binRange": [
                    "0"
                ],
                "rmse": 0.12058676222298643,
                "pearsonCorrelation": 0,
                "modelAccuracy": 0.985458832776577
            },
            {
                "numberOfRecords": 1481,
                "binRange": [
                    "1"
                ],
                "rmse": 0.976427088012165,
                "pearsonCorrelation": 0,
                "modelAccuracy": 0.04659014179608367
            }
        ],
        "modelAccuracy": 0.7738203957382039,
        "validationStatistics": {
            "count": 6570,
            "mean": 0.02176560121765598,
            "variance": 0.021295101084773255,
            "standardDeviation": 0.1459284108211052,
            "coefficientVariation": 6.704543070592046,
            "sumWeights": 6570,
            "sumObservations": 143,
            "correctedSumSquares": 139.88751902587552,
            "uncorrectedSumSquares": 143,
            "standardErrorMean": 0.001800350933352361,
            "skewness": 6.554369996787157,
            "kurtosis": 40.959613847782606
        },
        "percentilesHistogram": {
            "percentile100": 1,
            "percentile99": 1,
            "percentile95": 0,
            "percentile90": 0,
            "percentile75": 0,
            "percentile50": 0,
            "percentile25": 0,
            "percentile10": 0,
            "percentile5": 0,
            "percentile1": 0,
            "percentile0": 0
        }
    },
    "trainingParameters": {
        "ensembleTechnique": "ELITE_AVERAGE",
        "learners": [
            {
                "learningTechnique": "NEURAL_NET",
                "args": { }
            },
            {
                "learningTechnique": "GRADIENT_BOOST",
                "args": { }
            },
```

```
                {
                    "learningTechnique": "DECISION_TREE",
                    "args": { }
                },
                {
                    "learningTechnique": "LINEAR_REGRESSION",
                    "args": { }
                }
            ],
            "miThreshold": 0.3,
            "iterativeTrainingRecordSampleSize": 8192,
            "iterativeTrainingImprovementPercentageThreshold": 0.001,
            "samplingStrategy": "SAMPLE_WITHOUT_REPLACEMENT"
        },
        "trainingModelFeatures": [
            "Respiratory Rate value",
            "ZipCode",
            "RiskScore Mortality",
            "Rx Adherence Risk",
            "Hospital Readmission",
            "White Blood Cell Count value"
        ]
    }
```

Query Predicted Results vs. Actual Prediction Results

Variable	Description	DataType	Populated
count	The number of unique records in the validation data set that exactly match the predicted and actual values.	Integer	Y
pva	A container comparing ThingWorx Machine Learning's Predicted results versus Actual results	Object	Y
actual	The actual goal value (or true outcome) for a record (or group of records) in the validation data set.	Double	Y
predicted	The predicted goal value (or predicted outcome) for a record (or group of records) in the validation data set.	Double	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prediction/:result_id/pva
Method: GET

Sample Response: Query Predicted Results vs. Actual Prediction Results

```
    {
        "goal": "CustomerChurn",
        "values": [
            {
                "count": 200,
                "pva": {
                    "actual": "0.25",
                    "predicted": "0.19"
                }
            },
            {
                "count": 1,
```

```
                "pva": {
                    "actual": "0.018",
                    "predicted": "0.005"
                }
            }
        ]
    }
```

SCORING

The scoring service allows you to request and retrieve individual record-level prediction scores for a defined data set for a set of prediction topics (or goals).

List All Predictive Scoring Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_scores
Method: GET

Sample Response: List All Predictive Scoring Jobs


```
[
    {
        "resultId": 2345,
        "lastModified": "2014-07-16 01:01:00 AM",
        "dataSet": "CustomerCareDatabase",
        "configuration": {
            "resultId": 1234,
            "description": "Analysis of customer churn for east coast
customers with Product X",
            "filter": "EastCoastCustomers_ProductX",
            "additionalFeatures": [
                "Ethnicity",
                "Age"
            ],
            "importantFeatureCount": 0
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
]
```


Retrieve Specific Predictive Scoring Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_scores/:result_id
Method: GET

Sample Response: Retrieve Specific Scoring Job


```
{
    "resultId": 2345,
    "lastModified": "2014-07-16 01:01:00 AM",
    "dataSet": "CustomerCareDatabase",
    "configuration": {
        "resultId": 1234,
        "description": "Analysis of customer churn for east coast
customers with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "additionalFeatures": [
            "Ethnicity",
```

```
                "Age"
            ],
            "importantFeatureCount": 0
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
```

Retrieve Specific Predictive Scoring Job Configuration

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_scores/:result_id/configuration
Method: GET

Sample Response: Retrieve Specific Predictive Scoring Job Configuration

```
    {
        "goal": null,
        "description": "Analysis of customer churn for east coast customers
    with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "exclusions": [ ],
        "resultId": 1234,
        "importantFeatureCount": 0,
        "additionalFeatures": [
            "Ethnicity",
            "Age"
        ],
        "causalTechnique": "FULL_RANGE"
    }
```

Submit Predictive Scoring Job

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the training prediction model to be used for scoring. The model must be for the same dataset and completed successfully	Integer	Y
filter	The name of a filter	String	N
importantFeatureCount	Indicates the number of causal factors you want ThingWorx Machine Learning to return with each scored result. This is a non-negative whole number which represents a reasonable number of features available for their dataset.	Integer	N
causalTechnique	Options are FULL_RANGE (default), DISTANCE_FROM_MAX, and DISTANCE_FROM_MIN	String	N
additionalFeatures	List of additional features that need to be included in predictive scoring results	List[]	N

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_scores
Method: POST

Post Body:

```
{
    "resultId": 2785,
    "description": "scores for all east coast customer using product
x",
    "filter": "EastCoastCustomers_ProductX",
    "importantFeatureCount": 0,
    "additionalFeatures": [
        "SMOKING [PrevYear1]",
        "HasPHA Completion"
    ]
}
```

Response

See standard "status" object

Delete Predictive Scoring Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_scores/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Predictive Scoring Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the scoring job	Integer	Y

Request

URI: /datasets/:dataset_name/predictive_scores/:result_id/results
Method: GET
Accept: text/csv

Note: In order to receive Predictive Scoring Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Sample Response: Retrieve Predictive Scoring Result

Example CSV response without features:

IDENTIFIER,DATE TIME,PREDICTED GOAL 'ALLOWEDAMOUNT' SCORE,PREDICTED GOAL 'ALLOWEDAMOUNT' MIN,PREDICTED GOAL 'ALLOWEDAMOUNT' MAX
id-1002,2014-06-10T13:52:43Z,13690.4346,15583.2906,11797.5785
id-1353,2014-06-10T13:52:43Z,2539.7005,4432.55657,646.8444924
id-881,2014-06-10T13:52:43Z,2530.2191,4423.07517,637.363092

Example CSV response with features:

IDENTIFIER,DATE TIME,PREDICTED GOAL 'RISKSCORE STROKE' SCORE,PREDICTED GOAL 'RISKSCORE STROKE' MIN,PREDICTED GOAL 'RISKSCORE STROKE' MAX,FEATURE 'ETHNICITY'
20959,2014-09-02T17:40:34.1626000-04:00,0.15674095646588648,0.05368285917465575,0.25979905375711726,African American

Retrieve Predictive Scoring Important Feature Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the scoring job	Integer	Y

Request

URI: /datasets/:dataset_name/predictive_scores/:result_id/results_important_features
Method: GET
Accept: text/csv

Note: In order to receive Predictive Scoring Important Feature Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Example CSV response with important features:

IDENTIFIER,IMPORTANT FEATURE NAME,IMPORTANT FEATURE VALUE
20959,RiskScore Mortality,0.36734415514715124
20959,Creatinine value,0.12419053916184486

Variable	Description	DataType	Populated
identifier	The identifier for the record that was scored.	String	Y
dateTime	The timestamp associated with the record.	String	N
features	A list of additional features requested as part of the export.	Object	Y
importantFeatures	A list of features and weights that were used in the scoring process. May contain between 0 and 5 important features. ThingWorx Machine Learning will return the most important features for the score up to the requested amount. If features had no influence on	Object	Y

	the score, it will not be returned. Only features that had an influence on the score will be returned.		
predictedGoals			
name	feature name of the goal for the prediction model	String	Y
score	The predicted value for the given record.	Double	Y
max	The predicted maximum value for the score. Only present for integer or double datatype goals.	Double	N
min	The predicted minimum value for the score. Only present for integer or double datatype goals.	Double	N

Note: Retrieve scoring results support Accept-Encoding header of "gzip" to allow the data to be compressed in transit.

List All Prescriptive Scoring Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores
Method: GET

Sample Response: List All Prescriptive Scoring Jobs


```
[
    {
        "resultId": 76540,
        "dataSet": "coldlight!!test_dataset_clinical_bool_2",
        "configuration": {
            "maximize": true,
            "filter": "",
            "goal": "Hospital Readmission",
            "resultId": 76539,
            "prescriptiveFeatures": [
                "RiskScore Mortality",
                "Rx Adherence Risk",
                "Hospital Readmission"
            ],
            "description": null
        },
        "lastModified": "2015-07-10T21:02:55.968Z",
        "state": "COMPLETED",
        "description": null
    }
]
```


Retrieve Specific Prescriptive Scoring Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores/:result_id
Method: GET

Sample Response: Retrieve Specific Prescriptive Scoring Job


```
{
```

```
        "resultId": 76540,
        "dataSet": "coldlight!!test_dataset_clinical_bool_2",
        "configuration": {
            "maximize": true,
            "filter": "",
            "goal": "Hospital Readmission",
            "resultId": 76539,
            "prescriptiveFeatures": [
                "RiskScore Mortality",
                "Rx Adherence Risk",
                "Hospital Readmission"
            ],
            "description": null
        },
        "lastModified": "2015-07-10T21:02:55.968Z",
        "state": "COMPLETED",
        "description": null
    }
```


Retrieve Specific Prescriptive Scoring Job Configuration

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores/:result_id/configuration
Method: GET

Sample Response: Retrieve Specific Prescriptive Scoring Job Configuration


```
    {
        "goal": null,
        "description": null,
        "filter": null,
        "exclusions": [ ],
        "resultId": 76539,
        "prescriptiveFeatures": [
            "RiskScore Mortality",
            "Rx Adherence Risk",
            "Hospital Readmission"
        ],
        "maximize": true
    }
```


Submit Prescriptive Scoring Job

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the training prediction model to be used for scoring. The model must be for the same dataset and completed successfully	Integer	Y
filter	The name of a filter	String	N
prescriptiveFeatures	List of features from the training model	List []	Y
maximize	Determines whether the prescription should maximize or minimize the goal value. Available options are true and false. **Default value set to true.**	Boolean	N

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores
Method: POST

Post Body:

```
{
    "description": null,
    "resultId": 76539,
    "filter": "",
    "prescriptiveFeatures": [
        "RiskScore Mortality",
        "Rx Adherence Risk",
        "Hospital Readmission"
    ],
 "maximize": true
}
```

Response

See standard "status" object

Delete Prescriptive Scoring Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Prescriptive Scoring Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the prescriptive scoring job	Integer	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores/:result_id/results
Method: GET
Accept: text/csv

Note: In order to receive Prescriptive Scoring Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Sample Response: Retrieve Prescriptive Scoring Result

Example CSV response:

ROW ID,GOAL NAME,ORIGINAL SCORE,OPTIMIZED SCORE,FEATURE 'HOSPITAL READMISSION' ORIGINAL VALUE,FEATURE 'HOSPITAL READMISSION' OPTIMAL VALUE,FEATURE 'RISKSCORE MORTALITY' ORIGINAL VALUE,FEATURE 'RISKSCORE MORTALITY' OPTIMAL VALUE
1,0,1,0,0,0.0872000009,0.0872000009
2,0,1,0,1,0.0864999965,0.5563782732127028

Retrieve Prescriptive Scoring Metadata Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the prescriptive scoring job	Integer	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/prescriptive_scores/:result_id/results_metadata
Method: GET
Accept: text/csv

Note: In order to receive Predictive Scoring Metadata Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Sample Response: Retrieve Prescriptive Scoring Metadata Result

Example CSV response metadata:

GOAL NAME,FEATURE NAME,IS INCLUDED
Hospital Readmission,Hospital Readmission,true
Hospital Readmission,RiskScore Mortality,true

PREDICTIVE EVALUATION

Predictive Evaluation allows users to evaluate how accurately ThingWorx Machine Learning scored their data. Predictive Evaluation is used when running Predictive Scoring on a known dataset allowing users to receive the scored results and statistics measuring how accurate the scores are. Customers may find using Predictive Evaluation can help determine if previously created models are still accurate for new data.

List All Predictive Evaluation Jobs

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation
Method: GET

Sample Response: List All Predictive Evaluation Jobs


```
[
    {
        "resultId": 2345,
        "lastModified": "2014-07-16 01:01:00 AM",
        "dataSet": "CustomerCareDatabase",
        "configuration": {
            "resultId": 1234,
            "description": "Analysis of customer churn for east coast
customers with Product X",
            "filter": "EastCoastCustomers_ProductX",
            "additionalFeatures": [
                "Ethnicity",
                "Age"
            ],
            "importantFeatureCount": 0
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
]
```


Retrieve Specific Predictive Evaluation Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id
Method: GET

Sample Response: Retrieve Specific Evaluation Job


```
{
    "resultId": 2345,
    "lastModified": "2014-07-16 01:01:00 AM",
    "dataSet": "CustomerCareDatabase",
    "configuration": {
        "resultId": 1234,
```

```
            "description": "Analysis of customer churn for east coast
    customers with Product X",
            "filter": "EastCoastCustomers_ProductX",
            "additionalFeatures": [
                "Ethnicity",
                "Age"
            ],
            "importantFeatureCount": 0
        },
        "state": "COMPLETED",
        "description": "The job has successfully completed."
    }
```


Retrieve Specific Predictive Evaluation Job Configuration

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id/configuration
Method: GET

Sample Response: Retrieve Specific Predictive Evaluation Job Configuration


```
    {
        "goal": null,
        "description": "Analysis of customer churn for east coast customers
    with Product X",
        "filter": "EastCoastCustomers_ProductX",
        "exclusions": [ ],
        "resultId": 1234,
        "importantFeatureCount": 0,
        "additionalFeatures": [
            "Ethnicity",
            "Age"
        ],
        "causalTechnique": "FULL_RANGE"
    }
```


Submit Predictive Evaluation Job

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the training prediction model to be used for scoring. The model must be for the same dataset and completed successfully	Integer	Y
filter	The name of a filter	String	N
importantFeatureCount	Indicates the number of causal factors you want ThingWorx Machine Learning to return with each scored result. This is a non-negative whole number which represents a reasonable number of features available for their dataset.	Integer	N
causalTechnique	Options are FULL_RANGE (default), DISTANCE_FROM_MAX, and DISTANCE_FROM_MIN	String	N

additionalFeatures	List of additional features that need to be included in predictive evaluation results	List[]	N

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation
Method: POST

Post Body:

```
{
    "resultId": 2785,
    "description": "scores for all east coast customer using product
x",
    "filter": "EastCoastCustomers_ProductX",
    "importantFeatureCount": 0,
    "additionalFeatures": [
        "SMOKING [PrevYear1]",
        "HasPHA Completion"
    ]
}
```

Response

See standard "status" object

Delete Predictive Evaluation Job

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id
Method: DELETE

Response

HTTP 204 NO CONTENT

Retrieve Predictive Evaluation Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the evaluation job	Integer	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id/results
Method: GET
Accept: text/csv

Note: In order to receive Predictive Evaluation Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Sample Response: Retrieve Predictive Evaluation Result

Example CSV response without features:

IDENTIFIER,DATE TIME,PREDICTED GOAL 'ALLOWEDAMOUNT' SCORE,PREDICTED GOAL 'ALLOWED AMOUNT' MIN,PREDICTED GOAL 'ALLOWED AMOUNT' MAX
id-1002,2014-06-10T13:52:43Z,13690.4346,15583.2906,11797.5785
id-1353,2014-06-10T13:52:43Z,2539.7005,4432.55657,646.8444924
id-881,2014-06-10T13:52:43Z,2530.2191,4423.07517,637.363092

Example CSV response with features:

IDENTIFIER,DATE TIME,PREDICTED GOAL 'RISKSCORE STROKE' SCORE,PREDICTED GOAL 'RISKSCORE STROKE' MIN,PREDICTED GOAL 'RISKSCORE STROKE' MAX,FEATURE 'ETHNICITY'
20959,2014-09-02T17:40:34.1626000-04:00,0.15674095646588648,0.05368285917465575,0.25979905375711726,African American

Variable	Description	DataType	Populated
identifier	The identifier for the record that was scored.	String	Y
dateTime	The timestamp associated with the record.	String	N
features	A list of additional features requested as part of the export.	Object	Y
importantFeatures	A list of features and weights that were used in the scoring process. May contain between 0 and 5 important features. ThingWorx Machine Learning will return the most important features for the score up to the requested amount. If features had no influence on the score, it will not be returned. Only features that had an influence on the score will be returned.	Object	Y
predictedGoals			
name	feature name of the goal for the prediction model	String	Y
score	The predicted value for the given record.	Double	Y
max	The predicted maximum value for the score. Only present for integer or double datatype goals.	Double	N
min	The predicted minimum value for the score. Only present for integer or double datatype goals.	Double	N

Retrieve Predictive Evaluation Important Feature Result

Request Parameters

Parameter	Description	DataType	Required
resultId	The identifier to the evaluation job	Integer	Y

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id/results/results_important_features
Method: GET
Accept: text/csv

Note: In order to receive Predictive Evaluation Important Feature Results, the header must not be "Accept: application/JSON" or a 406 error will result. The Accept header can either be left unvalued or set to "text/csv."

Sample Response: Retrieve Predictive Evaluation Important Features Result

Example CSV response with important features:

IDENTIFIER,IMPORTANT FEATURE NAME,IMPORTANT FEATURE VALUE
20959,RiskScore Mortality,0.36734415514715124
20959,Creatinine value,0.12419053916184486

Note: Retrieve scoring results support Accept-Encoding header of "gzip" to allow the data to be compressed in transit.

Retrieve Predictive Evaluation Statistics

Request

URI: https://*ip address*/1.0/datasets/:dataset_name/predictive_evaluation/:result_id/statistics
Method: GET

Sample Response: Describe Prediction Model Result


```
{
    "numberOfRecords": 6570,
    "rmse": 0.4755834356469914,
    "pearsonCorrelation": 0.09177739457607152,
    "binErrors": [
        {
            "numberOfRecords": 5089,
            "binRange": [
                "0"
            ],
            "rmse": 0.12058676222298643,
            "pearsonCorrelation": 0,
            "modelAccuracy": 0.985458832776577
        },
        {
            "numberOfRecords": 1481,
            "binRange": [
                "1"
            ],
            "rmse": 0.976427088012165,
            "pearsonCorrelation": 0,
            "modelAccuracy": 0.04659014179608367
        }
    ],
    "modelAccuracy": 0.7738203957382039,
    "validationStatistics": {
        "count": 6570,
        "mean": 0.02176560121765598,
        "variance": 0.021295101084773255,
        "standardDeviation": 0.1459284108211052,
```

```
            "coefficientVariation": 6.704543070592046,
            "sumWeights": 6570,
            "sumObservations": 143,
            "correctedSumSquares": 139.88751902587552,
            "uncorrectedSumSquares": 143,
            "standardErrorMean": 0.001800350933352361,
            "skewness": 6.554369996787157,
            "kurtosis": 40.959613847782606
        },
        "percentilesHistogram": {
            "percentile100": 1,
            "percentile99": 1,
            "percentile95": 0,
            "percentile90": 0,
            "percentile75": 0,
            "percentile50": 0,
            "percentile25": 0,
            "percentile10": 0,
            "percentile5": 0,
            "percentile1": 0,
            "percentile0": 0
        }
    }
```


For more information about these statistics, please see Predictive Model Generation.

JOB MONITOR

This is a group of services that allows the user to determine the list of pending jobs (which includes jobs both in the QUEUED and RUNNING state). It will also provide the ability to terminate jobs as required.

Variable	Description	DataType	Populated
uri	The location of the job status.	String	Y
resultId	The unique identifier for a job.	Integer	Y
message	A text status message indicating the current status of the job. Possible values include: RUNNING FAILED COMPLETED EXPIRED QUEUED TERMINATED	String	Y
progress	Indicates the job completion status. Currently only 0 or 100 are provided. In the future, more granular progress measure may be implemented.	Double	Y
messageInfo	A message to display extra system information, currently used only for failure messages.	String	N
startTime	Time when the job is picked up by a grid worker	String	Y
endTime	Time the job status changed to completed or failed.	String	Y
runTime	Total job run time, excluding queued time.	String	Y
queuedStartTime	The timestamp for the moment the job was posted.	String	Y
queuedDuration	Amount of time from job post to actual start time (moment it is picked up by a worker.)	String	Y
dataset	The name of the dataset being operated on.	String	Y
jobType	A text message explaining the type of job running. Possible values include: SIGNALS CLUSTER_MODEL CLUSTER_DESCRIPTION_MODEL PROFILES TRAINING_DESCRIPTION_MODEL SCORING_MODEL SCORING DATA_VALIDATION DATASET_SNAPSHOT PMML_MODEL DATA_UPLOAD PRESCRIP_SCORING PREDICTIVE_EVALUATION	String	Y

Retrieve All Pending Jobs

Pending jobs are defined jobs that in either a QUEUED or RUNNING state. This service will return an array of all currently pending jobs within the context of the Application ID provided in the request header. Completed, Failed, and Terminated jobs will not be included.

Request

URI: https://*ip address*/1.0/jobs
Method: GET

Sample Response: Pending Jobs Query


```
[
    {
        "resultId": 809341,
        "uri": "http://internal-api-staging-
vpc.local.coldlight.com/1.0/status/809341",
        "progress": 0,
        "message": "RUNNING",
        "messageInfo": null,
        "startTime": null,
        "endTime": null,
        "runTime": null,
        "queuedStartTime": null,
        "queuedDuration": null,
        "dataset": "60471780-7c50-46ba-abf3-83d505358e12",
        "jobType": "DATASET_SNAPSHOT"
    },
    {
        "resultId": 809467,
        "uri": "http://internal-api-staging-
vpc.local.coldlight.com/1.0/status/809467",
        "progress": 0,
        "message": "RUNNING",
        "messageInfo": null,
        "startTime": "2015-11-19T14:02:03.872Z",
        "endTime": null,
        "runTime": null,
        "queuedStartTime": "2015-11-19T14:02:01.884Z",
        "queuedDuration": "0:00:01.988",
        "dataset": "test_dataset1",
        "jobType": "SIGNALS"
    }
]
```


Retrieve Job Status

This service retrieves the job status for a single job. This service will be able to load statuses for jobs that are no longer pending including Completed, Failed, or Terminated jobs.

Request

URI: https://*ip address*/1.0/jobs/:result_id
Method: GET

Sample Response: Retrieve Job Status Query


```
{
    "resultId": 809467,
    "uri": "http://internal-api-staging-
vpc.local.coldlight.com/1.0/status/809467",
    "progress": 0,
    "message": "TERMINATED",
    "messageInfo": null,
    "startTime": "2015-11-19T14:25:36.303Z",
    "endTime": "2015-11-19T14:25:36.305Z",
    "runTime": "0:00:00.002",
    "queuedStartTime": "2015-11-19T14:02:01.884Z",
    "queuedDuration": "0:23:34.419",
    "dataset": "test_dataset1",
    "jobType": "SIGNALS"
}
```


Terminate Job

This request allows you to terminate a currently running job.

Request

URL: https://*ip address*/1.0/jobs/:result_id
Method: POST

POST Body:


```
{
    "action": "TERMINATE"
}
```


Sample Response: Terminate Job

Responds with the new job status.


```
{
    "resultId": 802924,
    "uri": "http://internal-api-staging-
vpc.local.coldlight.com/1.0/status/802924",
    "progress": 0,
    "message": "TERMINATED",
    "messageInfo": null,
    "startTime": "2015-11-18T19:18:03.642Z",
    "endTime": "2015-11-18T19:18:03.644Z",
    "runTime": "0:00:00.002",
    "queuedStartTime": "2015-11-16T21:37:41.676Z",
    "queuedDuration": "45:40:21.966"
}
```

INFORMATION SERVICES

A family of endpoints providing information such as build version of REST application.

Retrieve Version Information

This end point provides a way to get a JSON payload back containing version information related to the deployment of the REST application.

Request

URI: https://*ip address*/1.0/about/versioninfo
Method: GET

Sample Response: Version Information Query

```
{
    "implementationVersion": "3.3.0-SNAPSHOT"
}
```