

Spectrum™ Technology Platform

Version v9.0 SP3

Analytics Guide

Contents

Chapter 1: Introduction.....	5
Advanced Analytics Module.....	6
Advanced Analytics Module Components.....	6
Chapter 2: Model Evaluator.....	7
Introduction to the Model Evaluator.....	8
Deploying a Model.....	8
Re-configure Model Evaluator Settings.....	9
Chapter 3: Advanced Analytics Repository.....	11
Introduction to the Advanced Analytics Repository.....	12
View Contents of Advanced Analytics Repository.....	12
View Model Details.....	12
Add a Model to the Advanced Analytics Repository.....	13
Delete a Model from the Advanced Analytics Repository.....	14
Chapter 4: Supported Model Types.....	15
Introduction to Supported Model Formats.....	16
QMML.....	16
Miner Model.....	16
PMML.....	17
Classification Tree.....	17

Introduction

In this section:

- **Advanced Analytics Module6**
- **Advanced Analytics Module Components6**

Advanced Analytics Module

The Spectrum™ Technology Platform Advanced Analytics Module evaluates statistical and predictive models defined in either QMML (Portrait Miner's proprietary model format) or PMML (industry standard Predictive Model Markup Language). It can be used to enrich data by adding predicted or scored outputs to an existing data flow using models created by data insight teams using industry standard data modeling tools. This may include models used to calculate the churn risk for existing customers or credit scoring models to determine a consumer's credit rating.

Advanced Analytics Module Components

The Advanced Analytics Module consists of the following components.

- **Model Evaluator**—This stage can be used to evaluate any model stored in the Advanced Analytics Repository in the context of a data flow.
- **Advanced Analytics Repository**—This is the central repository for all models available to the Advanced Analytics Module. Users can manage the repository via a web client.

Model Evaluator

In this section:

- [Introduction to the Model Evaluator](#)8
- [Deploying a Model](#)8
- [Re-configure Model Evaluator Settings](#)9

Introduction to the Model Evaluator

The Model Evaluator stage is capable of evaluating analytical models which have been published to the Advanced Analytics Repository in the context of a dataflow. The evaluator operates on single data rows using the fields from each row as the inputs to the model. User selected outputs from the model are written to the output channel.

Note: For details of the supported model types and type mappings see [Introduction to Supported Model Formats](#) on page 16

Deploying a Model

This procedure describes how to configure the Model Evaluator stage to deploy an analytics model as part of a dataflow.

1. Under **Primary Stages / Deployed Stages / Advanced Analytics**, click the **Model Evaluator** stage and drag it onto the canvas, placing it where you want on the dataflow and connecting it to input and output stages.
2. Double click the Model Evaluator stage to show the **Model Evaluator Options** dialog box. By default the options dialog shows the details of the first model in the list of available models
3. Click the **Type Filter** drop-down and select the model type to filter by. Only model types which are associated with at least one model in the **Advanced Analytics Repository** are listed.
4. Click the **Model** drop-down and select the model to deploy. The details of the chosen model are displayed in the dialog.

Model type the type of the selected model as described in [Introduction to Supported Model Formats](#) on page 16

Description a short text about the purpose of the model

Inputs a table containing information about all the required input fields for the model. Each row contains information about an input field. The model input field name will automatically be mapped to a valid Spectrum field name on publish (see [Add a Model to the Advanced Analytics Repository](#) on page 13).

Model Field Name the name of the field as specified in the model definition

Spectrum Field Name the name of the field as used in the Spectrum platform

Model Field Type the type of the field as specified in the model definition

Spectrum Field Type the Spectrum field type that is mapped to the model field type as described in [QMML](#) on page 16 and [PMML](#) on page 17

5. Select the **Configuration** tab. Details about the model's **Outputs** are displayed in a table. It contains information about all the output fields for the model. Each row has information about an output field. The model output field name will automatically be mapped to a valid Spectrum field name on publish (see [Add a Model to the Advanced Analytics Repository](#) on page 13).

Model Field Name the name of the field as specified in the model definition

Spectrum Field Name the name of the field as used in the Spectrum platform

Model Field Type	the type of the field as specified in the model definition
Spectrum Field Type	the Spectrum field type that is mapped to the model field type as described in QMML on page 16 and PMML on page 17
Include	a checkbox to specify whether to use this output

- Optional: Uncheck the **Included** column of any row in the **Outputs** table to exclude the output (i.e. stop it from being written to the output channel).
At least one output must remain selected. If all outputs are excluded then a validation error symbol will appear beside the **Outputs** table. This means the current model configuration is invalid and the model cannot be deployed. The validation error symbol will remain visible until the error is corrected.
- Optional: Click the **Spectrum Field Name** column of any row in the **Outputs** table to rename the fields as required.
No two outputs can share the same **Spectrum Field Name** and **Spectrum Field Names** must follow the standard Spectrum™ Technology Platform field naming conventions. If any validation errors are detected then a validation error symbol is displayed beside the **Outputs** table; hovering the mouse cursor over a validation error symbol shows the error details.
Note: Renaming an output's **Spectrum Field Name** only affects the specific instance of the stage, it does not update the Advanced Analytics Repository.
- Finally click **OK** to save the chosen model and configuration.

Re-configure Model Evaluator Settings

This procedure describes how to re-configure the Model Evaluator stage to re-name or change the outputs that are generated from the Model Evaluator stage or to change the deployed model to be evaluated by the stage.

- Double click the Model Evaluator stage to show the **Model Evaluator Options** dialog box.
The options dialog shows the model which was previously selected, configured and deployed.
Note: If the selected model has been deleted from the Advanced Analytics Repository prior to opening the **Model Evaluator Options** dialog, a validation error symbol is displayed beside the **Model** drop-down list. This means any new changes made to the model's configuration cannot be deployed (step 5). Clicking **Cancel** will exit the **Model Evaluator Options** and allow the deleted model to be used within the dataflow as previously configured. In order to apply any new changes to the stage configuration, select a different (non-deleted) model from the **Model** drop-down list. After applying a change of model, further re-configuration of the stage will no longer have the deleted model available in the **Model** drop-down list.
- Optional: From the **Model** tab select a different model from the **Model** drop-down list to change which model is used within the dataflow.
Note: On changing the selected model in the **Model** drop-down list, all configuration changes for the previously chosen model will be discarded. Re-selecting the model will result in the default output configuration for the model. Clicking **Cancel** will undo any changes made since opening the **Model Evaluator Options** allowing any pending changes to be reverted.
Note: On selecting a model, if the model has been deleted since the **Model Evaluator Options** dialog was opened then a validation error symbol will be displayed beside the **Model** drop-down list. In this scenario the **Inputs** and **Outputs** will not be available and re-configuration of the model settings will not be able to be applied until a non-deleted model has been selected.
- Select the Configuration tab
- Make any desired changes to the selected model's outputs configuration. For example, rename a **Spectrum Field Name**, or change its **Include** checkbox status to be included/excluded from the dataflow.

Any changes to the outputs must adhere to the following validation rules: At least one output must be included. Output's **Spectrum Field Name** must be unique and must follow the standard field naming conventions. If any validation errors are detected then a validation error symbol will be displayed beside the table; hovering the mouse cursor over a validation error symbol will show the error details.

5. Once all desired changes have been made click the **OK** button to apply the new changes, or click **Cancel** to undo any changes made and close the **Model Evaluator Options** dialog.

Advanced Analytics Repository

In this section:

- [Introduction to the Advanced Analytics Repository12](#)
- [View Contents of Advanced Analytics Repository12](#)
- [View Model Details12](#)
- [Add a Model to the Advanced Analytics Repository13](#)
- [Delete a Model from the Advanced Analytics Repository .14](#)

Introduction to the Advanced Analytics Repository

The Advanced Analytics Repository is where analytical models are saved for use with the Model Evaluator stage. There is a web browser client for managing the models stored in the repository.

Note: For details of the supported model types and type mappings see [Introduction to Supported Model Formats](#) on page 16

Accessing the Advanced Analytics Repository Web Browser Client

In a web browser enter the URL `http://<servername>:<port>/analytics` (e.g. `http://localhost:8080/analytics`) to display the **Login** page. Enter a valid Spectrum™ Technology Platform username and password in the browser dialog box displayed. The browser will navigate to a page listing the existing Analytics models persisted in the repository.

View Contents of Advanced Analytics Repository

This procedure describes how to view the list of models that are currently in the Advanced Analytics Repository.

1. Log in to the Advanced Analytics Repository web browser client by entering a valid Spectrum™ Technology Platform username and password in the **Login** page. On successful login, the application will navigate to the **Repository** page showing a table listing the models in the Advanced Analytics Repository that are available for use by the Model Evaluator stage.

For details of how to log in via the **Login** page see [Introduction to the Advanced Analytics Repository](#) on page 12

2. Select the number of models to be shown per page from the **Models per page** drop-down below the model table. There is the option of displaying 10, 20, or 50 models per table page.
3. The page control beside the **Models per page** drop-down is used to navigate between the different pages.
The page control will only allow page changes when the value chosen in step 2 for the number of models per page is less than the total number of models in the Advanced Analytics Repository.
4. Click on the **Name** column header to toggle sorting of listed models by the **Name** value in ascending/descending order. Alternatively click on the **Type** column header to toggle sorting of listed models by the **Type** value in ascending/descending order.

Before any column headers are clicked the models by default are sorted by **Name** value in ascending order.

View Model Details

This procedure describes how to view details about a model listed in the Advanced Analytics Repository to view it's description and details of the inputs and outputs for the model.

1. Find the model you are looking for in the **Repository** page.
For details on how to browse the Advanced Analytics Repository see [View Contents of Advanced Analytics Repository](#) on page 12
2. Navigate to the model's detail page by either clicking on model's name in the **Name** column of the table or checking the model's checkbox and clicking the **Model details** button above the table.
The details page displays detailed information about the model.

Name	the name of the model
Type	type of the model as described in Introduction to Supported Model Formats on page 16
Description	a short text about the purpose of the model
Inputs	a table containing information about all the required fields for the model. Each row contains a field's name and type as described in Introduction to Supported Model Formats on page 16
Outputs	a table containing information about all the output fields produced by the model. Each row contains a field's name and type as described in Introduction to Supported Model Formats on page 16

3. To navigate back to the **Repository** page click the **Close** button in the top-right corner.

Add a Model to the Advanced Analytics Repository

This procedure describes how to add a new model in to the Advanced Analytics Repository in order to make the model available for deployment in a Spectrum™ Technology Platform dataflow.

1. Using the [View Contents of Advanced Analytics Repository](#) on page 12 procedure, navigate to the **Repository** page.
2. Click the **Publish a model** button (located directly above the model table) to navigate to the **Add new model** page.

The **Publish a model** button will be disabled if any models are selected in the model table's current page. De-select these models to re-enable the **Publish a model** button.

3. On the **Add new model** page there are three input fields: **Name**, **Description** and **Upload model**. Write the name of the new model in to the text box labeled **Name**.

Model names must be unique. If the model name inputted is already in use by another non-deleted model stored in the Advanced Analytics Repository, the text box will be highlighted with a validation error.

Note: White space at the start or end of a model's name is always removed. This means if the name entered is already in use, adding white space to the start or end of the name will not resolve the error.

Note: Model names are case sensitive. This means models can share the same name so long as they have different uppercase and lowercase characters. For example models named "my Model", "my model" and "mY ModEl" can co-exist simultaneously within the repository.

Note: Name validation only detects conflicts against the contents of the Advanced Analytics Repository when the page was opened. This means if the repository has changed since navigating to the **Add new model** page e.g. models have been added or deleted by other users, then the page validation will not have these changes. However, on clicking the **Publish** button, the new model will be re-validated against the up-to date version of the Advanced Analytics Repository.

4. Write a description for the new model in to the text area labeled **Description**. Models must have a description.
5. Add a model definition by clicking the **Choose file** button and selecting a model stored in either QMML or PMML file formats.

Note: The selected model must be less than 50MB in size.

Note: When using an older browser that does not support file picking, the XML model definition must be input into the text area labeled **Upload model**.

Note: The model definition will not be validated until clicking the **Publish** button.

6. Click the **Publish** button to begin uploading the model to the Advanced Analytics Repository.
The **Publish** button will be disabled until the new model has a valid name and a description and model definition specified.
 7. Wait as the **Publish** button's busy icon is shown. No changes can be made to the new model's name, description or model definition while the model is being published. Once the model has been successfully published in to the Advanced Analytics Repository the page will navigate to the new model's **Details** page.
The **Cancel** button will be disabled while the model is being published.
- Note:** If a problem is found while publishing the model e.g. the XML model definition is invalid or the model name is already in use, then the model will not be published and an alert error will be displayed detailing the problem. Click **OK** to close the error alert and correct the problem before attempting to publish the model again.

Delete a Model from the Advanced Analytics Repository

This procedure describes how to delete models from the Advanced Analytics Repository. Deleting a model from the repository will prevent the model from being deployed to any new Spectrum™ Technology Platform dataflows. Deleting a model will not affect the runtime operation of any existing dataflows that are configured to use the model although no modifications can be made to the settings of a Model Evaluator stage within a dataflow that is configured to use the model.

1. Using the [View Contents of Advanced Analytics Repository](#) on page 12 procedure, navigate to the **Repository** page.
The **Repository** page displays models stored within the Advanced Analytics Repository via a model table. If a large number of models are stored within the Advanced Analytics Repository then the model table may display models across several pages.
2. Using the model table select models to be deleted from the Advanced Analytics Repository.
Note: Only models selected on the model table's current page will be deleted. For example, if the model table has two pages of models and models have been selected on both pages, then clicking the delete button will only delete the models on the current page and ignore any models selected on the other page.
3. Click the **Delete model** button, located directly above the model table.
The **Delete model** button will be disabled when no models are selected on the model table's current page.
4. A pop-up message is displayed asking for confirmation to delete the selected model or models from the Advanced Analytics Repository. Click the **Yes** button to confirm and delete the selected models from the Advanced Analytics Repository. Click the **No** button to cancel the delete operation.
After successfully deleting the selected models from the Advanced Analytics Repository the **Repository** page will refresh and update the model table.

Supported Model Types

In this section:

- **Introduction to Supported Model Formats**16
- **QMML**16
- **PMML**17

Introduction to Supported Model Formats

The Advanced Analytics Module supports the deployment of analytical models saved in both QMML and PMML file formats.

QMML models can be created and exported from Portrait Miner. All types of analytical models and segmentations exported from Portrait Miner are supported by the Model Evaluator stage and can be deployed within a Spectrum™ Technology Platform dataflow.

PMML models can be created and exported from many commercial and open source modeling tools. A list of the supported model types and their outputs is detailed in the [PMML](#) on page 17 section.

QMML

QMML is a proprietary XML based file format used to represent model results generated from Portrait Miner.

All types of analytical models and segmentations exported from Portrait Miner are supported by the Model Evaluator stage and can be deployed within a Spectrum™ Technology Platform dataflow.

Type Mapping

QMML model inputs and outputs are automatically mapped to Spectrum™ Technology Platform field types.

QMML Field Type	Spectrum™ Technology Platform Field Type
integer	integer
real	double
string	string
date	date time

Supported Models

All model types constructed within Portrait Miner (including decision tree, scorecard, cluster analysis and naive bayes models) are interpreted by the Advanced Analytics Module as a Miner Model.

Miner Model

A Miner Model is any type of QMML model exported from Portrait Miner, such as those generated from the decision tree, scorecard, cluster analysis or naive bayes modeling tools.

Unsupported Features

All compiled QMML exported from Portrait Miner is supported.

Model Outputs

Field	Description
<dynamic fields>	A field is output for each QMML model output.

PMML

Predictive Model Markup Language (PMML) is an XML-based file format developed by the Data Mining Group to provide a way for applications to describe and exchange models produced by data mining and machine learning algorithms. PMML files can be created and exported from many commercial and open source modeling tools.

Type Mapping

PMML model inputs and outputs are automatically mapped to Spectrum™ Technology Platform field types. All PMML models inputs and outputs must be of supported types.

PMML Field Type	Spectrum™ Technology Platform Field Type
string	string
integer	integer
float	float
double	double
boolean	boolean
date	date
dateDaysSince[1960]	date
dateDaysSince[1970]	date
dateDaysSince[1980]	date
dateDaysSince[0]	<i>not supported</i>
time	time
timeSeconds	time
dateTime	date time
dateTimeSecondsSince[1960]	date time
dateTimeSecondsSince[1970]	date time
dateTimeSecondsSince[1980]	date time
dateTimeSecondsSince[0]	<i>not supported</i>

Supported Models

The Advanced Analytics Module currently supports PMML classification tree models only.

Classification Tree

A PMML classification tree model predicts membership of a categorical dependent variable from one or more independent variables.

Unsupported Features

Classification trees with a missing value strategy of "aggregateNodes" or "weightedConfidence" are not supported.

Model Outputs

Model Output Feature	Description
Target field	If specified, the target field will be available as an output feature - this is a synonym for the predictedValue feature.
predictedValue	The categorical dependent variable that we are predicting membership of.
probability	The statistical probability of the predicted value. Multiple probability outputs can be specified in the model, one for each predicted category or by rank.
predictedDisplayValue	The human readable value used to represent the predicted value from the model.
transformedValue	A value generated via a transformation expression applied to the predicted model output.
decision	A value generated via an expression applied to the predicted model output resulting in a categorized value.
entityId	If present, the ID of the tree node of the predicted result.

Notices

© 2014 Pitney Bowes Software Inc. All rights reserved. MapInfo and Group 1 Software are trademarks of Pitney Bowes Software Inc. All other marks and trademarks are property of their respective holders.

USPS® Notices

Pitney Bowes Inc. holds a non-exclusive license to publish and sell ZIP + 4® databases on optical and magnetic media. The following trademarks are owned by the United States Postal Service: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS^{Link}, NCOA^{Link}, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite^{Link}, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, and ZIP + 4. This list is not exhaustive of the trademarks belonging to the Postal Service.

Pitney Bowes Inc. is a non-exclusive licensee of USPS® for NCOA^{Link}® processing.

Prices for Pitney Bowes Software's products, options, and services are not established, controlled, or approved by USPS® or United States Government. When utilizing RDI™ data to determine parcel-shipping costs, the business decision on which parcel delivery company to use is not made by the USPS® or United States Government.

Data Provider and Related Notices

Data Products contained on this media and used within Pitney Bowes Software applications are protected by various trademarks and by one or more of the following copyrights:

© Copyright United States Postal Service. All rights reserved.

© 2014 TomTom. All rights reserved. TomTom and the TomTom logo are registered trademarks of TomTom N.V.

© Copyright NAVTEQ. All rights reserved

Data © 2014 NAVTEQ North America, LLC

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

Based upon electronic data © National Land Survey Sweden.

© Copyright United States Census Bureau

© Copyright Nova Marketing Group, Inc.

Portions of this program are © Copyright 1993-2007 by Nova Marketing Group Inc. All Rights Reserved

© Copyright Second Decimal, LLC

© Copyright Canada Post Corporation

This CD-ROM contains data from a compilation in which Canada Post Corporation is the copyright owner.

© 2007 Claritas, Inc.

The Geocode Address World data set contains data licensed from the GeoNames Project (www.geonames.org) provided under the Creative Commons Attribution License ("Attribution License") located at <http://creativecommons.org/licenses/by/3.0/legalcode>. Your use of the GeoNames data (described in the Spectrum™ Technology Platform User Manual) is governed by the terms of the Attribution License, and any conflict between your agreement with Pitney Bowes Software, Inc. and the Attribution License will be resolved in favor of the Attribution License solely as it relates to your use of the GeoNames data.

ICU Notices

Copyright © 1995-2011 International Business Machines Corporation and others.

All rights reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, and/or sell copies of the

Software, and to permit persons to whom the Software is furnished to do so, provided that the above copyright notice(s) and this permission notice appear in all copies of the Software and that both the above copyright notice(s) and this permission notice appear in supporting documentation.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE BE LIABLE FOR ANY CLAIM, OR ANY SPECIAL INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

