

Spectrum™ Technology Platform

Version 2018.2.0

Analytics Scoring-Handbuch



Inhalt

1 - Einführung

Advanced Scoring-Modul	4
Komponenten des Analytics Scoring-Moduls	4

2 - Advanced Scoring-Modul

Binning Lookup	6
Java Model Scoring	7
PMML Model Scoring	8
Read from Miner Dataset	24
Write to Miner Dataset	27

3 - Machine

Learning-Modellverwaltung

Zugreifen auf die Maschine	
Learning-Modellverwaltung	34
Modellbewertung	35

4 - Analytics Scoring-Repository

Einführung in das Analytics Scoring-Repository	44
Modelle anzeigen	44
Modelldetails anzeigen	45
Modell über Webbrowser-Client hinzufügen	46
Hinzufügen eines Modells über Spectrum Miner	
Spectrum Connector	48
Modell löschen	50

1 - Einführung

In this section

Advanced Scoring-Modul	4
Komponenten des Analytics Scoring-Moduls	4

Advanced Scoring-Modul

Das Advanced Scoring-Modul von Spectrum™ Technology Platform ermöglicht Vorhersagemodelle, die entweder in **QMML** auf Seite 13 (das proprietäre Modellformat von Spectrum Miner) oder in **PMML** auf Seite 14 (Predictive Model Markup Language als Industriestandard) definiert werden, um in Datenflüssen ausgewertet zu werden. Darüber ermöglicht das Modul Datenflüssen Daten in Miner-Datasets, die von der Predictive Analytics-Software Miner genutzt werden können, zu schreiben und aus ihnen abzurufen. Daten können so unter Verwendung von Modellen, die von Data Insight-Teams über Datenmodellierungstools nach Industriestandard erstellt werden, um vorhergesagte oder ausgewertete Ausgaben ergänzt werden. Beispielsweise können Modelle enthalten sein, die das Abwanderungsrisiko bei bestehenden Kunden berechnen oder die Kreditwürdigkeit eines Kunden feststellen.

Komponenten des Analytics Scoring-Moduls

Das Analytics Scoring-Modul setzt sich aus den folgenden Komponenten zusammen.

- **Binning Lookup:** Dieser Schritt kann verwendet werden, um zuvor definiertes Binning auf neue Daten anzuwenden, indem vorhandene Bins verwendet werden, die in Datenflüssen mithilfe des Binning-Schritts des Machine Learning-Moduls erstellt wurden.
- **Java Model Scoring:** Dieser Schritt kann verwendet werden, um neue Daten mithilfe der Formel zu bewerten, die beim Anpassen eines Machine Learning-Modells erstellt wird.
- **PMML Model Scoring:** Sie können diesen Schritt verwenden, um im Kontext eines Datenflusses ein beliebiges Modell auszuwerten, das im Analytics Scoring-Repository gespeichert ist.
- **Read from Miner Dataset:** Sie können diesen Schritt verwenden, um Daten aus einer Fokusdatei zu lesen, die in einem Datenfluss verwendet werden sollen.
- **Write to Miner Dataset:** Sie können diesen Schritt verwenden, um Daten aus einem Datenfluss in eine Fokusdatei zu schreiben.
- **Machine Learning-Modellverwaltung:** Dieses Repository enthält Model Assessment, mit dem Sie alle Machine Learning-Modelle auf Ihrem Spectrum™ Technology Platform-Server verwalten, und Binning Management, mit dem Sie alle Binning-Vorgänge auf Ihrem Spectrum™ Technology Platform-Server verwalten.
- **Analytics Scoring-Repository:** Dies ist das zentrale Repository für alle Modelle, die für das Analytics Scoring-Modul verfügbar sind. Benutzer können das Repository über einen Webclient verwalten.

Beispiele für überwachtes und unüberwachtes Lernen, das das Scoring von Daten mithilfe von Java Model Scoring umfasst, finden Sie im Handbuch **Machine Learning** unter „Data Science-Demonstrationsfluss“.

2 - Advanced Scoring-Modul

In this section

Binning Lookup	6
Java Model Scoring	7
PMML Model Scoring	8
Read from Miner Dataset	24
Write to Miner Dataset	27

Binning Lookup

Einführung in Binning Lookup

Binning Lookup wendet vorher definiertes Binning auf neue Daten an und verwendet dabei vorhandene Bins, die in Datenflüssen erstellt wurden, die den **Binning**-Schritt verwenden.

Anmerkung: Modelle müssen zunächst über die Machine Learning-Modellverwaltung verfügbar gemacht werden, bevor sie im „Binning Lookup“-Schritt verfügbar werden. Weitere Informationen finden Sie unter **Einführung in die Modellbewertung** auf Seite 35.

Definieren von Binning-Eigenschaften

1. Klicken Sie unter **Primäre Schritte/Bereitgestellte Schritte/Analytics Scoring** auf den **Binning Lookup**-Schritt und ziehen Sie ihn auf die Arbeitsfläche. Platzieren Sie ihn an einer beliebigen Stelle im Datenfluss und verbinden Sie ihn mit anderen Schritten. Beachten Sie, dass die Eingabephase die Daten enthalten muss, die gebinnt werden sollen. Für die Binning-Ausgabe ist ein Ausgabeschritt erforderlich. Sie können optional einen zweiten Ausgabeschritt verbinden, um die Binning-Zusammenfassung zu erfassen.
2. Wählen Sie den entsprechenden **Binning-Namen** aus dem Dropdown-Menü aus. Das sind die Namen vorhandener Bins, die von einem den Binning-Schritt verwendenden Datenfluss erstellt wurden.
3. Die Felder **Binning-Typ** und **Beschreibung** werden mit den Bins des Binning-Namens importiert, den Sie in Schritt 2 ausgewählt haben, und sind daher nicht bearbeitbar.
4. Das Raster **Eingaben** zeigt alle Felder, die für das Binning im Binning-Schritt eingeschlossen wurden, zusammen mit dem Datentyp.
5. Klicken Sie auf **OK**, um Ihre Einstellungen zu speichern.

Binning-Ausgabe

Auf dieser Registerkarte werden die Felder und Datentypen angezeigt, die von Ihrem „Binning Lookup“-Schritt gebinnt werden. Weitere Informationen zur Ausgabe, die durch die Verwendung eines Binning-Schritts generiert wird, finden Sie unter **Binning-Ausgabe**. Wir haben nun die Option verfügbar gemacht, gebinnte Felder aus der Registerkarte der Binning-Ausgabe zu bearbeiten. Sie

können „Spectrum Binning-Felder“ verwenden, um ein Binning-Feld umzubenennen. Sie haben nun auch die Option, Binning-Felder einzuschließen/auszuschließen.

Java Model Scoring

Einführung in Java Model Scoring

Java Model Scoring ermöglicht Ihnen die Bewertung neuer Daten mithilfe der Formel, die beim Anpassen eines Machine Learning-Modells erstellt wird.

Anmerkung: Modelle müssen zunächst über die Machine Learning-Modellverwaltung verfügbar gemacht werden, bevor sie im „Java Model Scoring“-Schritt verfügbar werden. Weitere Informationen finden Sie unter [Einführung in die Modellbewertung](#) auf Seite 35.

Sie müssen zwei Registerkarten im Dialogfeld „**Java Model Scoring**“-**Optionen** ausfüllen, um Ihre Daten bewerten zu können. Identifizieren Sie zunächst das Modell und dessen Typ, und stellen Sie anschließend sicher, dass die Felder des Modells den Spectrum™ Technology Platform-Feldern ordnungsgemäß zugeordnet wurden. Im Anschluss können Sie die Ausgabe konfigurieren, indem Sie auswählen, welche Felder eingeschlossen werden sollen, und Ihren Auftrag ausführen. Die Registerkarte **Modellausgabe** enthält das Mapping für Datentypen der Spectrum™ Technology Platform und Ihr Modell.

Wenn Ihr Auftrag einen Schritt enthält, der die Ausgabe in einer Datei oder Tabelle erfasst, können Sie diese Ausgabe in einem nachfolgenden Datenfluss oder Webservice verwenden.

Definieren von Modelleigenschaften

1. Klicken Sie unter **Primäre Schritte/Bereitgestellte Schritte/Advanced Analytics** auf den **Java Model Scoring**-Schritt und ziehen Sie ihn auf die Arbeitsfläche. Platzieren Sie ihn an einer beliebigen Stelle im Datenfluss und verbinden Sie ihn mit Eingabe- und Ausgabeschritten. Beachten Sie, dass der Eingabeschritt die Datenquelle sein muss, die Ziel- und Eingabevariablenfelder für Ihr Modell enthält. Wenn Sie Ihren Auftrag im Batchmodus ausführen, benötigen Sie zudem einen Ausgabeschritt für die Erfassung von Modellbewertungen; anderenfalls können Sie Daten mithilfe eines Spectrum™ Technology Platform-Webservice in Echtzeit bewerten.
2. Doppelklicken Sie auf den „Java Model Scoring“-Schritt, um das Dialogfenster „**Model Scoring**“-**Optionen** anzuzeigen.
3. Optional: Wählen Sie im Dropdown-Menü **Typfilter** den Typ eines Modells aus, das Sie bewerten.

4. Wählen Sie den **Typfilter** aus, der für die Bewertung des Modells verwendet wird.
5. Wählen Sie den **Modellnamen** aus dem Dropdown-Menü aus.
6. Geben Sie im Feld **Modelltyp** den Typ des Modells aus, das Sie bewerten.
7. Optional: Geben Sie eine **Beschreibung** des Modells ein.
8. Aktivieren Sie **Unbekannte kategorische Ebenen ignorieren**, um Daten in der Spalte „Predicted_Value“ für Zeilen mit einer kategorischen Ebene zurückzugeben, die nicht in den Daten zur Anpassung des Modells enthalten war. Wenn Sie dieses Kästchen deaktiviert lassen, wird in der Spalte „Predicted_Value“ für diese Zeilen „Null/NA“ zurückgegeben.
9. Die Relation **Eingaben** enthält Informationen zu den Eingabefeldern des Modells. Diese Felder und die zugehörigen Datentypen ordnen Spectrum automatisch Felder und Datentypen zu.
10. Klicken Sie auf **OK**, um diese Optionen zu speichern, oder fahren Sie mit der nächsten Registerkarte fort.

Modellausgabe

Die Tabelle **Ausgaben** enthält Informationen zu den Ausgabefeldern des Modells. Diese Felder und die zugehörigen Datentypen ordnen Spectrum automatisch Felder und Datentypen zu.

1. Klicken Sie bei jedem Feld, dessen Daten im Modell eingeschlossen werden sollen, auf **Einschließen**.
2. Klicken Sie auf **OK**, um das Modell zu speichern.

PMML Model Scoring

Einführung in PMML Model Scoring

Der „PMML Model Scoring“-Schritt kann im Kontext eines Datenflusses analytische Modelle auswerten, die in der Analytics Scoring-Datenbank veröffentlicht wurden. Der Evaluator arbeitet mit einzelnen Datenzeilen und verwendet die Felder jeder Zeile als Eingabe für das Modell. Vom Benutzer festgelegte Ausgaben des Modells werden in den Ausgabekanal geschrieben.

Anmerkung: Details zu unterstützten Modelltypen und Typzuordnungen finden Sie unter **Unterstützte Modellformate** auf Seite 13.

Modell bereitstellen

Diese Schritte beschreiben, wie Sie den „PMML Model Scoring“-Schritt konfigurieren, um ein Analysemodell als Teil eines Datenflusses bereitzustellen.

1. Klicken Sie unter **Primärschritte / Bereitgestellte Schritte / Analytics Scoring** auf den **PMML Model Scoring**-Schritt, und ziehen Sie ihn auf die Arbeitsfläche. Platzieren Sie ihn an einer beliebigen Stelle im Datenfluss, und verbinden Sie ihn mit Eingabe- und Ausgabeschritten.
2. Doppelklicken Sie auf den „PMML Model Scoring“-Schritt, um das Dialogfenster „**PMML Model Scoring**“-Optionen aufzurufen.

Das Dialogfenster zeigt standardmäßig die Details des ersten Modells in der Liste der verfügbaren Modelle an.

3. Klicken Sie auf das Dropdown-Menü **Typfilter** und wählen Sie den Modelltyp aus, nach dem Sie filtern möchten.

In der Liste sind nur Modelltypen enthalten, die mindestens einem Modell im **Analytics Scoring-Repository** zugewiesen sind.

4. Klicken Sie auf das Dropdown-Menü **Modell** und wählen Sie das bereitzustellende Modell aus. Die Details zum ausgewählten Modell werden im Dialogfenster angezeigt.

Modelltyp Dies ist der Typ des ausgewählten Modells, wie unter **Unterstützte Modellformate** auf Seite 13 beschrieben

Beschreibung Dies ist ein kurzer Text zum Zweck des Modells

Eingaben Diese Tabelle enthält Informationen zu allen erforderlichen Eingabefeldern des Modells. Jede Zeile enthält Informationen zu einem Eingabefeld. Der Eingabefeldname des Modells wird bei Veröffentlichung automatisch einem gültigen Spectrum-Feldnamen zugeordnet (siehe **Modell über Webbrowser-Client hinzufügen** auf Seite 46 oder **Hinzufügen eines Modells über Spectrum Miner Spectrum Connector** auf Seite 48).

Modellfeldname Dies ist der Name des Feldes, wie er in der Modelldefinition festgelegt wurde

Spectrum-Feldname Dies ist der Name des Feldes, wie er auf der Spectrum-Plattform verwendet wird

Modellfeldtyp Dies ist der Typ des Feldes, wie er in der Modelldefinition festgelegt wurde

Spectrum-Feldtyp Dies ist der Spectrum-Feldtyp, der dem Modellfeldtyp, wie unter **QMML** auf Seite 13 und **PMML** auf Seite 14 beschrieben, zugeordnet wird

5. Wählen Sie die Registerkarte **Konfiguration** aus.

In einer Tabelle werden die Details zu den **Ausgaben** des Modells angezeigt. Sie enthält Informationen zu allen Ausgabefeldern des Modells. Jede Zeile enthält Informationen zu einem Ausgabefeld. Der Ausgabefeldname des Modells wird bei Veröffentlichung automatisch einem gültigen Spectrum-Feldnamen zugeordnet (siehe [Model über Webbrowser-Client hinzufügen](#) auf Seite 46 oder [Hinzufügen eines Modells über Spectrum Miner Spectrum Connector](#) auf Seite 48).

Modellfeldname	Dies ist der Name des Feldes, wie er in der Modelldefinition festgelegt wurde
Spectrum-Feldname	Dies ist der Name des Feldes, wie er auf der Spectrum-Plattform verwendet wird
Modellfeldtyp	Dies ist der Typ des Feldes, wie er in der Modelldefinition festgelegt wurde
Spectrum-Feldtyp	Dies ist der Spectrum-Feldtyp, der dem Modellfeldtyp, wie unter QMML auf Seite 13 und PMML auf Seite 14 beschrieben, zugeordnet wird
Einschließen	Über dieses Kästchen geben Sie an, ob diese Ausgabe verwendet wird

6. Optional: Deaktivieren Sie die Spalte **Eingeschlossen** in jeder Zeile der Tabelle **Ausgaben**, die von der Ausgabe ausgeschlossen werden soll. (Sie wird dann nicht mehr in den Ausgabekanal geschrieben.)

Mindestens eine Ausgabe muss ausgewählt bleiben. Wenn alle Ausgaben ausgeschlossen werden, wird ein Überprüfungsfehlersymbol neben der Tabelle **Ausgaben** angezeigt. Das bedeutet, dass die aktuelle Modellkonfiguration ungültig ist und das Modell nicht bereitgestellt werden kann. Das Überprüfungsfehlersymbol wird solange angezeigt, bis der Fehler behoben wurde.

7. Optional: Klicken Sie in der Spalte **Spectrum-Feldname** beliebiger Zeilen in der Tabelle **Ausgaben**, um die Felder nach Bedarf umzubenennen.

Spectrum-Feldnamen können nicht doppelt vergeben werden, darüber hinaus müssen die **Spectrum-Feldnamen** die Standardkonventionen zur Feldbenennung von Spectrum™ Technology Platform einhalten. Wenn Überprüfungsfehler erkannt werden, wird ein Überprüfungsfehlersymbol neben der Tabelle **Ausgaben** angezeigt. Wenn Sie den Mauszeiger über ein Überprüfungsfehlersymbol bewegen, werden die Fehlerdetails angezeigt.

Anmerkung: Umbenennen des **Spectrum-Feldnamens** einer Ausgabe wirkt sich nur auf die spezifische Instanz des Schritts aus. Das Analytics Scoring-Repository wird dadurch nicht aktualisiert.

Anmerkung: Feldnamen von Spectrum™ Technology Platform:

- müssen eindeutig sein,
- dürfen nicht leer sein,
- dürfen nur alphanumerische Zeichen, Punkte und Unterstriche enthalten,

- dürfen nicht mit einem Punkt beginnen.
8. Klicken Sie zum Abschluss auf **OK**, um das ausgewählte Modell und die Konfiguration zu speichern.

Rekonfigurieren von „PMML Model Scoring“-Einstellungen

Diese Schritte beschreiben, wie der „PMML Model Scoring“-Schritt neu konfiguriert werden kann, um die vom „PMML Model Scoring“-Schritt generierten Ausgaben umzubenennen oder zu ändern oder das bereitgestellte Modell, das vom Schritt ausgewertet werden soll, zu ändern.

1. Doppelklicken Sie auf den „PMML Model Scoring“-Schritt, um das Dialogfenster „**PMML Model Scoring**“-Optionen aufzurufen.

Das Dialogfeld mit den Optionen zeigt das Modell an, das zuvor ausgewählt, konfiguriert und bereitgestellt wurde.

Anmerkung: Wenn das ausgewählte Modell vor dem Öffnen des Dialogfeldes „**PMML Model Scoring**“-Optionen aus der Analytics Scoring-Datenbank gelöscht wurde, wird ein Überprüfungsfehlersymbol neben der Dropdown-Liste **Modell** angezeigt. Das bedeutet, dass alle neuen Änderungen an der Konfiguration des Modells nicht bereitgestellt werden können (Schritt 5). Klicken Sie auf **Abbrechen**, um die „**PMML Model Scoring**“-Optionen zu verlassen. Sie können dann das gelöschte Modell innerhalb des Datenflusses verwenden, wie es vorher konfiguriert wurde. Um neue Änderungen an der Schrittkonfiguration umzusetzen, wählen Sie ein anderes (nicht gelöscht) Modell aus der Dropdown-Liste **Modell** aus. Nachdem Sie Änderungen am Modell angewendet haben, ist das gelöschte Modell nicht mehr in der Dropdown-Liste **Modell** für weitere Konfigurationsänderungen des Schrittes verfügbar.

2. Optional: Wählen Sie auf der Registerkarte **Modell** ein anderes Modell aus der Dropdown-Liste **Modell** für die Verwendung im Datenfluss aus.

Anmerkung: Bei der Auswahl eines anderen Modells in der Dropdown-Liste **Modell** gehen alle Konfigurationsänderungen am vorher ausgewählten Modell verloren. Wenn Sie das vorherige Modell erneut auswählen, ist die standardmäßige Ausgabekonfiguration des Modells aktiv. Klicken Sie auf **Abbrechen**, um alle Änderungen, die Sie seit dem Öffnen der „**PMML Model Scoring**“-Optionen vorgenommen haben, rückgängig zu machen und die vorherige Konfiguration wiederherzustellen.

Anmerkung: Wenn das Modell nach dem Öffnen der „**PMML Model Scoring**“-Optionen gelöscht wurde, wird bei der Auswahl des Modells ein Überprüfungsfehlersymbol neben der Dropdown-Liste **Modell** angezeigt. In diesem Fall sind **Eingaben** and **Ausgaben** nicht verfügbar und eine erneute Konfiguration der Modelleinstellungen kann nicht angewendet werden, bis ein nicht gelöscht Modell ausgewählt wurde.

- Wählen Sie die Registerkarte Konfiguration aus.
- Führen Sie alle gewünschten Änderungen an der Ausgabenkonfiguration des ausgewählten Modells durch. Benennen Sie beispielsweise einen **Spectrum-Feldnamen** um oder aktivieren/deaktivieren Sie das Kästchen **Einschließen** des Feldes, um es in den Datenfluss einzuschließen oder es vom Datenfluss auszuschließen.

Alle Änderungen an den Ausgaben müssen folgende Überprüfungsregeln befolgen: Mindestens eine Ausgabe muss eingeschlossen sein. Der **Spectrum-Feldname** der Ausgabe muss eindeutig sein und die Standards der Feldnamenskonventionen von Spectrum™ Technology Platform erfüllen. Wenn Überprüfungsfehler erkannt werden, wird ein Überprüfungsfehlersymbol neben der Tabelle angezeigt. Wenn Sie den Mauszeiger über ein Überprüfungsfehlersymbol bewegen, werden die Fehlerdetails angezeigt.

Anmerkung: Feldnamen von Spectrum™ Technology Platform:

- müssen eindeutig sein,
 - dürfen nicht leer sein,
 - dürfen nur alphanumerische Zeichen, Punkte und Unterstriche enthalten,
 - dürfen nicht mit einem Punkt beginnen.
- Klicken Sie, sobald Sie alle gewünschten Änderungen vorgenommen haben, auf die Schaltfläche **OK**, um die neuen Änderungen anzuwenden, oder klicken Sie auf **Abbrechen**, um alle Änderungen rückgängig zu machen und das Dialogfeld „**PMML Model Scoring**“-Optionen zu schließen.

Ausgabe

Der „PMML Model Scoring“-Schritt gibt die Ausgabefelder des ausgewählten Modells zurück. Wenn der „PMML Model Scoring“-Schritt einen Datensatz nicht verarbeiten kann, gibt er die Felder „Status“, „Status.Code“ und „Status.Description“ zurück. Diese Felder bieten Informationen darüber, warum der Schritt den Datensatz nicht verarbeiten konnte. Details dazu:

Feldname	Beschreibung
Status	Meldet den Erfolg oder das Fehlschlagen des Auswertungsversuchs.
	null Erfolg
	F Fehler

Feldname	Beschreibung
Status.Code	Fehlergrund: InputConversionFailed Typen von Spectrum-Eingabefeldern konnten nicht in die erforderlichen Modellfeldtypen konvertiert werden. ModelEvaluationFailed Für den angegebenen Datensatz ist die Modellauswertung fehlgeschlagen.
Status.Description	Beschreibung des Problems

Unterstützte Modellformate

PMML Model Scoring unterstützt die Bereitstellung von analytischen Modellen, die in den Dateiformaten **QMML** auf Seite 13 und **PMML** auf Seite 14 gespeichert wurden.

QMML auf Seite 13-Modelle können in Spectrum Miner erstellt und daraus exportiert werden. Alle aus Spectrum Miner exportierten Typen von Analysemodellen und Segmentierungen werden vom „PMML Model Scoring“-Schritt unterstützt und können innerhalb eines Datenflusses von Spectrum™ Technology Platform bereitgestellt werden.

PMML auf Seite 14-Modelle können in vielen kommerziellen und Open-Source-Modellierungstools erstellt und aus ihnen exportiert werden.

QMML

QMML ist ein proprietäres, XML-basiertes Dateiformat, das zur Darstellung von Modellergebnissen aus Spectrum Miner verwendet wird.

Alle aus Spectrum Miner exportierten Typen von Analysemodellen und Segmentierungen werden vom „PMML Model Scoring“-Schritt unterstützt und können innerhalb eines Datenflusses von Spectrum™ Technology Platform bereitgestellt werden.

Typ-Mapping

QMML-Modelleingaben und -ausgaben werden automatisch zu Feldern von Spectrum™ Technology Platform zugeordnet.

QMML-Feldtyp	Spectrum™ Technology Platform-Feldtyp
integer	integer
real	double

QMML-Feldtyp	Spectrum™ Technology Platform-Feldtyp
string	string
date	datetime

Unterstützte Modelle

Alle Modelle, die innerhalb von Spectrum Miner (einschließlich Entscheidungsbaum-, Scorecard, Clusteranalyse- und naiver Bayes-Modelle) erstellt wurden, werden vom Analytics Scoring-Modul als Miner-Modell interpretiert.

Miner-Modell

Alle **QMML** auf Seite 13-Modelltypen, die aus Spectrum Miner exportiert wurden, wie Modelle, die aus Entscheidungsbäumen, Scorecards, Clusteranalysen oder naiven Bayes-Modellierungstools generiert wurden, sind Miner-Modelle.

Nicht unterstützte Features

Alle kompilierten **QMML** auf Seite 13-Dateien, die aus Spectrum Miner exportiert wurden, werden unterstützt.

Modellausgaben

Feld	Beschreibung
<dynamische Felder>	Es gibt ein Ausgabefeld für jede QMML auf Seite 13-Modellausgabe.

PMML

Bei der Predictive Model Markup Language (PMML) handelt es sich um ein XML-basiertes Dateiformat, das von der Data Mining Group entwickelt wurde, um Applikationen zu ermöglichen, von Data Mining- und Machine Learning-Algorithmen erstellte Modelle zu beschreiben und auszutauschen. PMML-Dateien können in vielen kommerziellen und Open-Source-Modellierungstools erstellt und aus ihnen exportiert werden.

Typ-Mapping

PMML-Modelleingaben und -ausgaben werden automatisch zu Feldern von Spectrum™ Technology Platform zugeordnet. Die Typen aller PMML-Modelleingaben und -ausgaben müssen unterstützt werden.

PMML-Feldtyp	Spectrum™ Technology Platform-Feldtyp
string	string
integer	integer
float	float
double	double
boolean	boolean
date	date
dateDaysSince[1960]	date
dateDaysSince[1970]	date
dateDaysSince[1980]	date
dateDaysSince[0]	nicht unterstützt
time	time
timeSeconds	time
dateTime	datetime
dateTimeSecondsSince[1960]	datetime
dateTimeSecondsSince[1970]	datetime
dateTimeSecondsSince[1980]	datetime
dateTimeSecondsSince[0]	nicht unterstützt

Unterstützte Modelle

Das Analytics Scoring-Modul unterstützt derzeit die PMML-Modelltypen, die in den folgenden Abschnitten beschrieben sind.

Zuordnungsregel

Ein Zuordnungsregelmodell in **PMML** auf Seite 14 steht für Regeln, die einen Satz von Elementen einem anderen Satz von Elementen zuordnen. Eine Regel kann beispielsweise ausdrücken, dass ein bestimmtes Produkt oder ein bestimmter Produktsatz oft zusammen mit einem bestimmten Satz anderer Produkte gekauft wird, was auch als Warenkorbanalyse bekannt ist. Ein Zuordnungsregelmodell verfügt typischerweise über zwei Variablen: eine, die Datensätze in Transaktionen gruppiert, und eine weitere, die jeden Datensatz eindeutig identifiziert.

Modellelement

```
<AssociationModel functionName="associationRules" ...
```

Nicht unterstützte Features

Für das Element identifizierende Felder werden andere Feldtypen als der Feldtyp „String“ nicht unterstützt.

Gruppierung von Datensätzen nach mehr als einem Feld wird nicht unterstützt.

Zuordnungsregelmodelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Unterstützte Features für Modellausgaben	Beschreibung
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.
entityId	Dies ist die ID der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird. Wenn die ausgewählte Regel über keine ID verfügt, wird ein Wert eines bei 1 startenden Index zurückgegeben.
ruleId	Diese Option ist identisch mit „entityId“ und seit PMML 4.2 veraltet. Obwohl die Verwendung derzeit unterstützt wird, empfehlen wir, „entityId“ zu verwenden.
affinity	Dies ist die Affinität der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird.

Unterstützte Features für Modellausgaben	Beschreibung
antecedent	Dies ist die Antezedenz der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird. Diese Ausgabe wird als kommagetrennte Zeichenfolge von Werten formatiert.
consequent	Dies ist die Konsequenz der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird. Diese Ausgabe wird als kommagetrennte Zeichenfolge von Werten formatiert.
Regel	Dies ist die angewendete Regel (Standard) oder die Regel, die durch den Rangwert festgelegt wird. Dieser Ausgabe gibt eine Beschreibung der Regel zurück, die folgendermaßen formatiert ist: {<Antezedenz>-}{<Konsequenz>}.
confidence	Dies ist die Zuverlässigkeit der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird.
support	Dies ist der Support der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird.
lift	Dies ist die Anhebung der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird.
leverage	Dies ist der Einfluss der angewendeten Regel (Standard) oder der Regel, die durch den Rangwert festgelegt wird.

Clustering

Ein **PMML** auf Seite 14-Clustering-Modell bestimmt das am besten übereinstimmende Cluster für einen angegebenen Datensatz, basierend auf dem im Cluster verwendeten Maß für Entfernung oder Ähnlichkeit. Ein Cluster ist eine Teilmenge mit ähnlichen Daten. Clustering (auch bekannt als unüberwachtes Lernen) ist der Prozess, bei dem ein Dataset in Gruppen aufgeteilt wird, sodass die Mitglieder jeder Gruppe einander so ähnlich wie möglich sind und unterschiedliche Gruppen sich so viel wie möglich voneinander unterscheiden.

Modellelement

```
<ClusteringModel functionName="clustering" ...
```

Nicht unterstützte Features

Clustering-Modelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist das am besten übereinstimmende Cluster, basierend auf dem im Cluster verwendeten Maß für Entfernung oder Ähnlichkeit.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.
predictedDisplayValue	Dies ist der visuell lesbare Wert, der zur Anzeige des vom Modell vorhergesagten Wertes verwendet wird.
entityId	Dies ist der bei 1 startenden Index (implizite Identifikation) des vorhergesagten Ergebnisclusters.
affinity	Dies ist der Wert der Entfernung oder der Ähnlichkeit des angegebenen Datensatzes bezüglich des vorhergesagten Clusters, wie er im Modell definiert wird.

Klassifizierungsbaum

Ein **PMML** auf Seite 14-Klassifizierungsbaummodell sagt die Mitgliedschaft einer kategorieabhängigen Variable aus einer oder mehreren unabhängigen Variablen voraus.

Modellelement

```
<TreeModel functionName="classification" ...
```

Nicht unterstützte Features

Klassifizierungsbäume, bei denen eine Wertstrategie „aggregateNodes“ oder „weightedConfidence“ fehlt, werden nicht unterstützt.

Klassifizierungsbaummodelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die kategorieabhängige Variable, deren Mitgliedschaft vorausgesagt werden soll.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.
predictedDisplayValue	Dies ist der visuell lesbare Wert, der zur Anzeige des vom Modell vorhergesagten Wertes verwendet wird.
probability	Dies ist die statistische Wahrscheinlichkeit des vorhergesagten Wertes. Im Modell können mehrere Wahrscheinlichkeitsausgaben angegeben sein, eine für jede vorausgesagte Kategorie oder mehrere Wahrscheinlichkeitsausgaben nach Rang.
residual	Der Restwert des Ausgabewertes der Wahrscheinlichkeit (1 - Wahrscheinlichkeit) für die vorhergesagte Kategorie oder den Rang. Im Modell können mehrere Restwertausgaben angegeben sein, eine für jede vorausgesagte Kategorie oder mehrere Restwertausgaben nach Rang.
entityId	Dies ist die ID des Baumknotens des vorhergesagten Ergebnisses, falls vorhanden.

Regressionsbaum

Ein **PMML-Regressionsbaummodell** sagt einen Wert einer numerisch abhängigen Variable (wie der Preis eines Hauses) auf Basis einer oder mehrerer unabhängigen Variablen voraus. Dies wird durch Erstellen eines Entscheidungsbaummodells erreicht, das auf einem oder mehreren Prädiktoren basiert.

Modellelement

```
<TreeModel functionName="regression" ...
```

Nicht unterstützte Features

Regressionsbäume mit Zielfeldern vom Typ „Integer“ oder „Float“ werden nicht unterstützt, wenn ein <Targets>-Element nicht mit dem korrekten Attribut „castInteger“ festgelegt wurde.

Regressionsbaummodelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“. Zielfelder vom Typ „Float“ werden immer in Integer umgewandelt.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die numerisch abhängige Variable, die vorausgesagt werden soll.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.
entityId	Dies ist die ID des Baumknotens des vorhergesagten Ergebnisses, falls vorhanden.

Naiver Bayes

Ein naives Bayes-**PMML** auf Seite 14-Modell sagt unter Verwendung des Bayes-Theorems den Wert eines Ziels aus Beweisen voraus, die von einem oder mehreren Prädiktorfeldern stammen. Bei einem naiven Bayes-Modell muss das Zielfeld diskretisiert sind, sodass von dem Modell eine endliche Anzahl an Werten berücksichtigt wird. Prädiktorfelder können entweder diskret oder kontinuierlich sein.

Modellelement

```
<NaiveBayesModel functionName="classification" ...
```

Nicht unterstützte Features

Naive Bayes-Modelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die Kategorievariable, deren Mitgliedschaft vorausgesagt werden soll.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.
predictedDisplayValue	Dies ist der visuell lesbare Wert, der zur Anzeige des vom Modell vorhergesagten Wertes verwendet wird.
probability	Dies ist die statistische Wahrscheinlichkeit des vorhergesagten Wertes.
residual	Der Restwert des Ausgabewertes der Wahrscheinlichkeit (1 - Wahrscheinlichkeit) für den vorhergesagten Wert.

Regression

Ein **PMML** auf Seite 14-Regressionsmodell sagt den Wert einer numerisch abhängigen Variable aus einer oder mehreren unabhängigen Variablen voraus.

Modellelement

```
<RegressionModel functionName="regression" ...
```

Nicht unterstützte Features

Regressionsmodelle, bei denen das Attribut „normalizationMethod“ auf den Wert „simplemax“, „probit“, „cloglog“ oder „loglog“ festgelegt ist, werden nicht unterstützt.

Regressionsmodelle mit einem Zielfeld vom Typ „Integer“ oder „Float“ werden nicht unterstützt, wenn ein <Targets>-Element nicht mit dem korrekten Attribut „castInteger“ angegeben wurde.

Regressionsmodelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“. Zielfelder vom Typ „Float“ werden immer in Integer umgewandelt.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die numerisch abhängige Variable, die vorausgesagt werden soll.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.

Regressionsklassifikator

Ein **PMML** auf Seite 14-Regressionsklassifikator kombiniert die Ausgabe von mehreren Regressionsgleichungen, um einen Kategoriewert vorauszusagen.

Modellelement

```
<RegressionModel functionName="classification" ...
```

Nicht unterstützte Features

Regressionsmodelle mit dem Element `<MiningSchema>`, das einen Verweis auf ein Element `<DerivedField>` enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die kategorieabhängige Variable, deren Mitgliedschaft vorausgesagt werden soll.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.

Unterstützte Features für Modellausgaben	Beschreibung
predictedDisplayValue	Dies ist der visuell lesbare Wert, der zur Anzeige des vom Modell vorhergesagten Wertes verwendet wird.
probability	Dies ist die statistische Wahrscheinlichkeit des vorhergesagten Wertes. Im Modell können mehrere Wahrscheinlichkeitsausgaben angegeben sein, eine für jede vorausgesagte Kategorie oder mehrere Wahrscheinlichkeitsausgaben nach Rang.
residual	Der Restwert des Ausgabewertes der Wahrscheinlichkeit (1 - Wahrscheinlichkeit) für den vorhergesagten Wert. Im Modell können mehrere Restwertausgaben angegeben sein, eine für jede vorausgesagte Kategorie oder mehrere Restwertausgaben nach Rang.

Scorecard

Ein **PMML** auf Seite 14-Scorecard-Modell ist eine regressionsbasierte Modellierungstechnik, die hauptsächlich zur Berechnung von Risiken oder Standardwahrscheinlichkeiten verwendet wird.

Modellelement

```
<Scorecard functionName="regression" ...
```

Nicht unterstützte Features

Scorecard-Modelle mit dem Element <MiningSchema>, das einen Verweis auf ein Element <DerivedField> enthält, werden nicht unterstützt.

Modellausgaben

Standardmäßig ist das Zielfeld als Ausgabefeld verfügbar, was gleichbedeutend ist mit dem Feature „predictedValue“, und immer vom Typ „Double“.

Unterstützte Features für Modellausgaben	Beschreibung
predictedValue	Dies ist die berechnete Punktzahl.
transformedValue	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Transformationsausdruck generiert wird.
decision	Dies ist ein Wert, der über einen auf die vorhergesagte Modellausgabe angewendeten Ausdruck generiert wird, der einen kategorisierten Wert ergibt.

Unterstützte Features für Modellausgaben	Beschreibung
reasonCode	Wenn angegeben, sind dies die Grundcodes, die den Merkmalen zur Bestimmung der Punktzahl zugeordnet sind. Die Reihenfolge richtet sich dabei nach ihrem Beitrag zur endgültigen Punktzahl. Der Rang des Anforderungsgrundcodes kann festgelegt werden. Anderenfalls wird standardmäßig der Grundcode mit dem höchsten Rang zurückgegeben.

Read from Miner Dataset

Einführung in den „Read from Miner Dataset“-Schritt

Der „Read from Miner Dataset“-Schritt kann Feldinformationen und Daten aus einem Miner-Dataset extrahieren.

Miner-Dataset auslesen

Diese Schritte beschreiben, wie Sie Feldinformationen und Daten aus einem Miner-Dataset auslesen können.

1. Ziehen Sie unter **Primärschritte / Bereitgestellte Schritte / Analytics Scoring** den **Read from Miner Dataset**-Schritt auf die Arbeitsfläche, und verbinden Sie ihn mit dem Datenfluss.
2. Doppelklicken Sie auf den **Read from Miner Dataset**-Schritt, um das Dialogfeld „**Read from Miner Dataset**“-Optionen anzuzeigen.
3. Klicken Sie auf die Schaltfläche ... im Feld **Dateiname**, um das Dialogfeld **Datei öffnen** anzuzeigen.
4. Suchen Sie über das Dialogfeld **Datei öffnen** die Fokusdatei, die das Miner-Dataset enthält, und wählen Sie sie aus.
Fokusdateien haben die Erweiterung .ftr.
5. Klicken Sie im Dialogfeld „**Read from Miner Dataset**“-Optionen auf die Registerkarte **Felder**, um die Feldtabelle anzuzeigen.
Wenn Sie den **Read from Miner Dataset**-Schritt zum ersten Mal konfigurieren und jedes Mal, wenn eine neue Fokusdatei ausgewählt wird (Schritt 4), ist die Feldtabelle leer.
6. Klicken Sie auf die Schaltfläche **Erneut generieren**.

In der Feldtabelle werden alle Felder aufgelistet, die den **Read from Miner Dataset**-Schritt aus dem Miner-Dataset liest.

Anmerkung: Die Schaltfläche **Erneut generieren** ist deaktiviert, wenn keine Fokusdatei ausgewählt wurde (Schritte 3 und 4).

Anmerkung: Wenn der Schritt keine Felder vom Miner-Dataset abrufen kann, wird ein Fehlerdialogfeld angezeigt. Wählen Sie eine gültige Fokusdatei aus und versuchen Sie es erneut.

7. Ändern Sie falls erforderlich die Feldliste mithilfe der verfügbaren Optionen (siehe [Registerkarte „Felder“](#) auf Seite 25).
8. Klicken Sie schließlich auf die Schaltfläche **OK**, um die Konfiguration anzuwenden.

Wenn Überprüfungsfehler erkannt werden, wird eine Überprüfungsfehlermeldung angezeigt. Beheben Sie die Überprüfungsfehler, bevor Sie erneut auf die Schaltfläche **OK** klicken. Wenn keine Überprüfungsfehler gefunden werden, wird das Dialogfeld **„Read from Miner Dataset“-Optionen** geschlossen.

Anmerkung: Um den Schritt neu zu konfigurieren, doppelklicken Sie auf den **Read from Miner Dataset**-Schritt, um das Dialogfeld **„Read from Miner Dataset“-Optionen** erneut anzuzeigen.

Registerkarte „Felder“

Die Tabelle **Felder** enthält eine Tabelle mit allen Feldern, die den **Read from Miner Dataset**-Schritt aus dem Miner-Dataset liest. In der folgenden Tabelle werden die Feldoptionen beschrieben, die Sie zum Ändern der Tabelleninhalte verwenden können.

Name der Option	Beschreibung
Erneut generieren	Ersetzt die aktuell definierten Felder mit denen, die aus dem Miner-Dataset gelesen werden.

In der folgenden Tabelle werden die Eigenschaften aller Felder beschrieben.

Spaltenname	Beschreibung	Behebbar
Miner-Feldname	Dies ist der Name des Feldes, wie er im Miner-Dataset festgelegt wird.	Nein

Spaltenname	Beschreibung	Bearbeitbar
Spectrum-Feldname	Dies ist der Name des Feldes, das im Datenflusskanal erstellt wird. Wenn der Miner-Feldname nicht mit den Benennungsregeln von Spectrum™ Technology Platform kompatibel ist, unterscheidet er sich vom Spectrum-Feldnamen . Bearbeiten Sie den Text in der Zelle, um den gewünschten Feldnamen für den Datenflusskanal festzulegen. Feldnamen von Spectrum™ Technology Platform: <ul style="list-style-type: none"> • müssen eindeutig sein, • dürfen nicht leer sein, • dürfen nur alphanumerische Zeichen, Punkte und Unterstriche enthalten, • dürfen nicht mit einem Punkt beginnen. 	Ja
Spectrum-Feldtyp	Dies ist der Datentyp des Feldes, das im Datenflusskanal erstellt wird (siehe Ausgabe auf Seite 26).	Nein
Einschließen	Über dieses Kästchen wird angegeben, ob dieses Feld in den Datenflusskanal eingeschlossen wird.	Ja

Anmerkung: Sie können die Felder nach **Miner-Feldnamen**, **Spectrum-Feldnamen** oder **Spectrum-Feldtyp** sortieren, indem Sie auf die entsprechende Tabellenspaltenüberschrift klicken.

Ausgabe

Das folgende Typ-Mapping zwischen Miner-Feldtyp und Spectrum-Feldtyp wird automatisch durchgeführt.

Tabelle 1: Feldzuordnungen

Miner-Feldtyp	Spectrum-Feldtyp
integer	integer
string	string
real	double

Miner-Feldtyp

Spectrum-Feldtyp

date

datetime

Wenn ein Feld über einen Namen mit Symbolen verfügt, die in Spectrum ungültig sind, werden die ungültigen Symbole automatisch durch „_“ ersetzt. Wenn zwei oder mehr Felder denselben transformierten Namen haben, wird wie im folgenden Beispiel ein zusätzlicher Index an den Namen angehängt.

`{x}Feld{x}``{y}Feld{y}`

`{x}` und `{y}` sind hier Zeichen, die in einem Spectrum-Feldnamen nicht unterstützt werden. Die Feldnamen würden dann zu

`_Feld_1``_Feld_2`

im Ausgabekanal.

Write to Miner Dataset

Einführung in den „Write to Miner Dataset“-Schritt

Der „Write to Miner Dataset“-Schritt kann ein Miner-Dataset erstellen, um die angegebenen Felder und Daten aus einem Spectrum-Datenfluss aufzunehmen.

In ein Miner-Dataset schreiben

Diese Schritte beschreiben, wie Sie Feldinformationen und Daten in ein Miner-Dataset schreiben können.

1. Ziehen Sie unter **Primäre Phasen / Bereitgestellte Phasen / Advanced Analytics** die **Write to Miner Dataset**-Phase auf die Canvas und verbinden Sie sie mit dem Datenfluss.
2. Doppelklicken Sie auf den **Write to Miner Dataset**-Schritt, um das Dialogfeld „**Write to Miner Dataset**“-Optionen anzuzeigen.

3. Klicken Sie auf die Schaltfläche ... im Feld **Dateiname**, um das Dialogfeld **Datei speichern** anzuzeigen.
4. Wählen Sie aus, wo die neue Fokusdatei gespeichert werden soll, und geben Sie einen entsprechenden Dateinamen an. Eine .ftr-Erweiterung wird automatisch an den Dateinamen angehängt. Stellen Sie sicher, dass der Ordner, in dem Sie die Datei speichern, nicht bereits eine *.ftr-Datei, einen RDX-Ordner oder einen XTR-Ordner mit demselben Namen enthält.
5. Optional: Deaktivieren Sie das Kästchen **Überschreiben**.
Wenn eine Fokusdatei mit demselben Namen bereits im angegebenen Verzeichnis (Schritt 4) vorhanden ist, verhindert das Deaktivieren des Kästchens **überschreiben**, dass die vorhandene Fokusdatei jedes Mal überschrieben wird, wenn der Schritt innerhalb eines Datenflusses ausgeführt wird. (Die Ausführung des Schrittes schlägt fehl.) Das bedeutet jedoch, dass Sie den Dateipfad neu konfigurieren (Schritt 4) oder die vorhandene Fokusdatei manuell verschieben oder löschen müssen, wenn der **Write To Miner Dataset**-Schritt erfolgreich innerhalb eines Datenflusses ausgeführt wurde.
6. Optional: Wenden Sie eine Metadatendatei auf die neue Fokusdatei an (siehe [Metadaten anwenden](#) auf Seite 29).
7. Klicken Sie im Dialogfeld „**Write to Miner Dataset**“-**Optionen** auf die Registerkarte **Felder**, um die Feldrelation anzuzeigen.
In der Feldrelation werden alle Felder aufgelistet, die der **Write to Miner Dataset**-Schritt in das Miner-Dataset schreibt.
8. Optional: Klicken Sie auf die Schaltfläche **Schnell hinzufügen**.
Wenn Sie auf die Schaltfläche **Schnell hinzufügen** klicken, wird das Dialogfeld **Schnell hinzufügen** geöffnet, in dem alle Felder des Eingabekanals und die aktuelle Feldliste angezeigt wird. Aktivieren/deaktivieren Sie die Kästchen neben den Feldern, um die gewünschten Felder zur Feldliste hinzuzufügen oder sie zu entfernen.
9. Ändern Sie falls erforderlich die Feldliste mithilfe der verfügbaren Optionen (siehe [Registerkarte „Felder“](#) auf Seite 29).
Mindestens ein Feld muss angegeben werden, da ansonsten ein Überprüfungsfehlerindikator angezeigt wird.
10. Klicken Sie schließlich auf die Schaltfläche **OK**, um die Konfiguration anzuwenden.
Wenn Überprüfungsfehler erkannt werden, wird eine Überprüfungsfehlermeldung angezeigt. Beheben Sie die Überprüfungsfehler, bevor Sie erneut auf die Schaltfläche **OK** klicken. Wenn keine Überprüfungsfehler gefunden werden, wird das Dialogfeld „**Write to Miner Dataset**“-**Optionen** geschlossen.

Anmerkung: Um den Schritt neu zu konfigurieren, doppelklicken Sie auf den **Write to Miner Dataset**-Schritt, um das Dialogfeld „**Write to Miner Dataset**“-**Optionen** erneut anzuzeigen.

Metadaten anwenden

Eine .qsfm-Metadatendatei enthält zusätzliche Informationen, die auf eine Fokusdatei angewendet werden können, wie beispielsweise abgeleitete Felddefinitionen, Feldinterpretationen, Klasseneinteilungen, Datensatzauswahlen, Feld- und Fokuscommentare sowie Fokusverlauf. Die folgenden Schritte beschreiben, wie Sie diese Art von Metadaten auf eine Fokusdatei anwenden, die unter Verwendung des **Write to Miner Dataset**-Schrittes erstellt wurde.

1. Beschaffen Sie sich eine gültige Miner-Fokusmetadatendatei (entweder von einem Drittanbieter oder durch Exportieren aus einem vorhandenen Fokus in Spectrum Miner).
Metadatendateien müssen die Erweiterung .qsfm haben.
2. Vergewissern Sie sich, dass das Dialogfeld „**Write to Miner Dataset**“-**Optionen** aktuell sichtbar ist (siehe [In ein Miner-Dataset schreiben](#) auf Seite 27).
3. Aktivieren Sie im Dialogfeld „**Write to Miner Dataset**“-**Optionen** das Kästchen **Metadaten anwenden**.
4. Klicken Sie unterhalb des Kästchens **Metadaten anwenden** auf die Schaltfläche ... im Feld **Dateiname**, um das Dialogfeld **Datei öffnen** anzuzeigen.
Die Metadatendateiauswahl ist deaktiviert, wenn das Kästchen **Metadaten anwenden** nicht aktiviert ist.
5. Suchen Sie über das Dialogfeld **Datei öffnen** die .qsfm-Datei aus Schritt 1 und öffnen Sie sie.

Anmerkung: Wenn das Kästchen **Metadaten anwenden** aktiviert ist, müssen Sie eine Metadatendatei auswählen, bevor Sie die Konfiguration des Schrittes anwenden können.

Anmerkung: Wenn Sie das Kästchen **Metadaten anwenden** deaktivieren, nachdem Sie eine Metadatendatei ausgewählt haben, wird die Metadatendateiauswahl deaktiviert und die Metadatendatei wird nicht bei Ausführung des Schrittes angewendet. Zu Ihrer Bequemlichkeit wird die ausgewählte Datei jedoch nicht aus der Dateiauswahl entfernt, bis das Dialogfeld „**Write to Miner Dataset**“-**Optionen** geschlossen wird.

6. Optional: Aktivieren Sie das Kästchen **Warnungen ignorieren**, wenn Sie nicht möchten, dass der Datenfluss abbricht, wenn ein Teil der Metadaten nicht angewendet werden kann.

Registerkarte „Felder“

Die Tabelle **Felder** enthält eine Tabelle mit allen Feldern, die den **Write to Miner Dataset**-Schritt in das Miner-Dataset schreibt. In der folgenden Tabelle werden die Feldoptionen beschrieben, die Sie zum Ändern der Tabelleninhalte verwenden können.

Name der Option	Beschreibung
Hinzufügen	Klicken Sie auf die Schaltfläche Hinzufügen , um das Dialogfeld Hinzufügen anzuzeigen, in dem Sie neue Felder zur Feldliste hinzufügen können. Im Dialogfeld Hinzufügen müssen Name, Miner-Typ und Länge (falls der Miner-Typ „String“ ist) des neuen Feldes angegeben werden.
Entfernen	Dies löscht die aktuell ausgewählten Felder aus der Feldliste.
Erneut generieren	<p>Wenn Sie auf die Schaltfläche Erneut generieren klicken, wird ein Dialogfeld angezeigt, in dem Sie die erneute Feldgeneration bestätigen müssen. Wenn Sie die Aktion bestätigen, werden alle vorhandenen Felder durch die Felder ersetzt, die in der ausgewählten Fokusdatei gefunden werden.</p> <p>Anmerkung: Die Schaltfläche Erneut generieren ist deaktiviert, wenn keine Fokusdatei ausgewählt wurde.</p> <p>Anmerkung: Bei einer erneuten Generierung aus einer ungültigen Fokusdatei wird die vorhandene Feldliste nicht gelöscht oder geändert.</p>
Schnell hinzufügen	<p>Wenn Sie auf die Schaltfläche Schnell hinzufügen klicken, wird das Dialogfeld Schnell hinzufügen geöffnet, in dem alle Felder des Eingabekanals und die aktuelle Feldliste angezeigt wird. Aktivieren/deaktivieren Sie die Kästchen neben den Feldern, um die gewünschten Felder zur Feldliste hinzuzufügen oder sie zu entfernen.</p> <p>Anmerkung: Die Schaltfläche Schnell hinzufügen ist deaktiviert, wenn der Miner-Feldname eines Feldes ungültig ist.</p>

Zudem können Sie die Feldeigenschaften aktualisieren, indem Sie die Zellen direkt in den folgenden Tabellenspalten ändern.

Spaltenname	Beschreibung
Miner-Feldname	<p>Dies ist der Name des Feldes, wie er im neuen Miner-Dataset angezeigt wird. Bearbeiten Sie den Text in der Zelle, um den gewünschten Feldnamen festzulegen. Die Feldnamen:</p> <ul style="list-style-type: none"> • müssen eindeutig sein, • dürfen nicht leer sein, • müssen mit einem Buchstaben beginnen, • dürfen nur alphanumerische Zeichen und Unterstriche enthalten, • dürfen nicht länger als 128 Zeichen sein. <p>Wenn ein Feldname in einen ungültigen Namen geändert wird, wird ein Überprüfungsfehlersymbol neben der Feldtabelle angezeigt. Beheben Sie den Überprüfungsfehler, bevor Sie auf die Schaltfläche OK klicken.</p>
Miner-Feldtyp	<p>Dies ist der Miner-Datentyp, den das Feld im neuen Miner-Dataset verwendet (siehe Ausgabe auf Seite 31). Ändern Sie den Datentyp des Feldes, indem Sie den gewünschten Typ aus der Dropdown-Liste auswählen.</p>
Zeichenfolgenlänge	<p>Wenn der ausgewählte Miner-Feldtyp in der Tabellenzeile „String“ ist, kann der Benutzer einen Integer-Wert zwischen 1 und 4000 angeben, um die Länge der Zeichenfolge des Feldes in der Miner-Fokusdatei festzulegen. Alle Zeichenfolgenwerte für das Feld, die länger als die angegebene Zeichenfolgenlänge sind, sind in der resultierenden Fokusdatei entsprechend abgeschnitten.</p>

Ausgabe

Das folgende Typ-Mapping zwischen Spectrum-Feldtyp und Miner-Feldtyp wird automatisch durchgeführt.

Tabelle 2: Feldzuordnungen

Spectrum-Feldtyp	Miner-Feldtyp
integer	integer
string	string

Spectrum-Feldtyp	Miner-Feldtyp
double	real
datetime	date

Alle anderen Spectrum-Feldtypen werden dem Miner-Feldtyp „String“ zugeordnet.

3 - Machine Learning-Modellverwaltung

In this section

Zugreifen auf die Machine Learning-Modellverwaltung	34
Modellbewertung	35

Zugreifen auf die Machine Learning-Modellverwaltung

Es gibt drei Möglichkeiten für den Zugriff auf die Machine Learning-Modellverwaltung:

- Verwenden Sie die Begrüßungsseite der Spectrum™ Technology Platform:

- Öffnen Sie einen Webbrowser und navigieren Sie zur Spectrum™ Technology Platform-Begrüßungsseite unter:

<Servername>:<Port>

Wenn Sie beispielsweise Spectrum™ Technology Platform auf einem Computer mit dem Namen „myspectrumplatform“ installiert haben und dieser den Standardport 8080 verwendet, navigieren Sie zu:

myspectrumplatform:8080

- Klicken Sie auf **Spectrum Machine Learning**.
- Klicken Sie auf **Machine Learning-Modellverwaltung öffnen**.
- Klicken Sie bei einem der Schritte zur Modellerstellung auf **Für Modelldetails hier klicken**.
- Verwenden Sie einen Webbrowser:
 - Öffnen Sie einen Webbrowser und navigieren Sie zur Seite „Machine Learning-Modellverwaltung“ der Spectrum™ Technology Platform unter:

<Servername>:<Port>/machinelearning

Wenn Sie beispielsweise Spectrum™ Technology Platform auf einem Computer mit dem Namen „myspectrumplatform“ installiert haben und dieser den Standardport 8080 verwendet, navigieren Sie zu:

myspectrumplatform:8080/machinelearning

- Geben Sie einen gültigen Benutzernamen und das zugehörige Kennwort für die Spectrum™ Technology Platform ein.

Modellbewertung





Einführung in die Modellbewertung

Auf der Registerkarte „Modellbewertung“ bei der Machine Learning-Modellverwaltung wird eine Liste mit allen Machine Learning-Modellen auf Ihrem Spectrum™ Technology Platform-Server angezeigt. Sie können diese Liste filtern, indem Sie eine Zeichenfolge in das Textfeld eingeben. Jedes Feld in der Tabelle wird dann nach dieser Zeichenfolge durchsucht.

In diesen Modellen können mehrere Vorgänge durchgeführt werden. Sie können Modelle verfügbar machen, die Verfügbarkeit von Modellen aufheben oder Modelle löschen. Mit den verfügbar gemachten Modellen werden im „Java Model Scoring“-Schritt mithilfe der beim Anpassen der Machine Learning-Modelle erstellten Formeln neue Daten bewertet. Zusätzlich können Sie zu jedem Modell detaillierte Informationen anzeigen. Welche Details zurückgegeben werden, hängt von dem Modelltyp ab, dessen Daten Sie anzeigen. Schließlich können Sie zwei beliebige Modelle des gleichen Typs miteinander vergleichen. Bei diesem Vergleich werden für alle Modelle, die Sie miteinander vergleichen, die gleichen Informationen nebeneinander angezeigt, die auf der Registerkarte „Modelldetail“ enthalten sind.

Modellbewertungsvorgänge

Führen Sie die folgenden Vorgänge durch, indem Sie ein Modell auswählen und auf die entsprechende Schaltfläche klicken:

	Zeigen Sie Details zur Modellausgabe an. Sie können auch über die „K-Means Clustering“- und „Logistic Regression“-Schritte auf diese Informationen zugreifen, indem Sie auf der Registerkarte „Modellausgabe“ auf „Für Modelldetails hier klicken“ klicken.
	Vergleichen Sie die Modelle miteinander.
	Importieren Sie ein Modell aus einem bestimmten Pfad. Wählen Sie aus, ob ein vorhandenes Modell mit demselben Namen überschrieben werden soll.
	Exportieren Sie ein Modell in einen bestimmten Pfad. Wählen Sie aus, ob ein vorhandenes Modell mit demselben Namen überschrieben werden soll.

	Machen Sie das Modell verfügbar, damit es für den „Java Model Scoring“-Schritt verfügbar ist. Wenn ein Modell nicht verfügbar gemacht wird, kann es nicht für Bewertungen verwendet werden.
	Heben Sie die Verfügbarkeit des Modells auf.
	<p>Löschen Sie das Modell.</p> <p>Anmerkung: Ein verfügbar gemachtes Modell kann nicht gelöscht werden. Zu diesem Zeitpunkt ist jedoch keine inhärente Sicherheit vorhanden, durch die verhindert wird, dass ein Benutzer die Modelle eines anderen Benutzers löscht.</p>

Die Registerkarte „Modelldetail“

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für ein Modell angezeigt:

- **Modellname:** Der Name des Modells
- **Modelltyp:** Der Typ des Machine Learning-Modells
- **Benutzer:** Der Benutzername der Person, die das Modell erstellt hat
- **Beschreibung:** Die Beschreibung des Modells, wenn bei der Erstellung des Modells eine Beschreibung angegeben wurde
- **Status:** Gibt an, ob das Modell verfügbar gemacht wurde oder ob die Verfügbarkeit aufgehoben wurde
- **Datenflussname:** Der Name des Datenflusses, der das Modell erzeugt hat
- **Erstellungszeit:** Das Datum und die Uhrzeit der Modellerstellung

Zusätzliche Details werden auf Grundlage des Modelltyps bereitgestellt.

„K-Means Clustering“-Details

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für „K-Means Clustering“-Modelle angezeigt:

Modellübersicht

Bietet Trainingsdaten für Folgendes:

- Anzahl der Zeilen
- Anzahl der Cluster
- Anzahl der kategorischen Spalten
- Anzahl der Iterationen
- Innerhalb der Cluster-Summe von Quadraten
- Gesamtsumme von Quadraten

- Zwischen der Cluster-Summe von Quadraten

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Innerhalb der Cluster-Gesamtsumme von Quadraten
- Gesamtsumme von Quadraten
- Zwischen der Cluster-Summe von Quadraten

Zentroidstatistik

Stellt für die einzelnen Zentroide folgende Trainings-, Test- und „N-fach“-Daten bereit:

- Größe
- Innerhalb der Cluster-Summe von Quadraten

Cluster-Mittelwerte

Stellt für jeden Zentroid detaillierte Informationen bereit. Der Inhalt variiert je nach Eingabedaten. Ein Cluster stellt eine Gruppe von Beobachtungen aus einem Dataset dar, die gemäß eines bestimmten Clustering-Algorithmus als ähnlich identifiziert wurden

Standardisierte Cluster-Mittelwerte

Stellt für jeden Zentroid standardisierte Informationen bereit. Der Inhalt variiert je nach Eingabedaten.

Details zu „Logistic Regression“

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für „Logistic Regression“-Modelle angezeigt:

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Anzahl der Beobachtungen
- R-Quadrat (R²)
- Logarithmischer Abfall (Logloss)
- Area under the curve (AUC)
- Precision-Recall-Bereich unter der Kurve (PR AUC)
- Gini-Koeffizient
- Mean Per Class Error
- Akaike-Informationskriterium (AIC)
- Lambda
- Restabweichung
- Abweichung von Null
- Freiheitsgrad von Null

- Restfreiheitsgrad

Maximum Metrics Threshold

Gibt den Training Maximum Metrics Threshold für Trainings-, Test- und „N-fach“-Daten mithilfe der folgenden Metriken an:

- max f1
- max f2
- max f0point5
- max accuracy
- max precision
- max recall
- max specificity
- max absolute_mcc
- max min_per_class_accuracy
- max mean_per_class_accuracy

Konfusionsmatrix

Stellt die Leistung eines Modells in einer Reihe von Trainings-, Test- und „N-fach“-Daten bereit, bei denen die tatsächlichen Werte bekannt sind.

Standardisiertes Koeffizientendiagramm

Zeigt die wichtigsten Prädiktoren an, indem der relative Wert der Koeffizienten angegeben wird. Dieser gibt an, wie stark sich das Ziel durch eine Änderung der Eingabe verändert.

GLM-Koeffizienten

Zeigt Koeffizienten für ein Generalized Linear-Modell, das Regressionsmodelle für Ergebnisse nach Exponentialverteilungen schätzt.

AUC-Kurven

Area under the curve (Fläche unter der Kurve); bestimmt, welches der Modelle die Klassen mithilfe der Trainings-, Test- und „N-fach“-Daten am besten vorhersagt.

Anhebungs-/Verstärkungskurven

Wertet die Fähigkeit des binären Klassifizierungsmodells aus, mithilfe der Trainings-, Test- und „N-Fold“-Daten Vorhersagen zu treffen.

Details zu Linear Regression

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für Linear Regression-Modelle angezeigt:

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)
- Anzahl der Beobachtungen
- R-Quadrat (R²)
- Mittlere Restabweichung
- Mittlerer absoluter Fehler (MAE)
- Wurzel aus dem mittleren quadratischen Fehler (RMSE)
- Akaike-Informationskriterium (AIC)
- Lambda
- Restabweichung
- Abweichung von Null
- Freiheitsgrad von Null
- Restfreiheitsgrad

Standardisiertes Koeffizientendiagramm

Zeigt die wichtigsten Prädiktoren an, indem der relative Wert der Koeffizienten angegeben wird. Dieser gibt an, wie stark eine Änderung des jeweiligen Prädiktorkoeffizientenwerts den Zielwert positiv oder negativ verändert. Zudem werden die besten 25 Koeffizienten im Modell dargestellt.

GLM-Koeffizienten

Zeigt Koeffizienten für ein Generalized Linear-Modell, das Regressionsmodelle für Ergebnisse nach Exponentialverteilungen schätzt.

Details zu Random Forest Regression

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für Random Forest Regression-Modelle angezeigt:

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Anzahl der Beobachtungen
- R-Quadrat (R²)
- Mittlere Restabweichung
- Mittlerer absoluter Fehler (MAE)
- Wurzel aus dem mittleren quadratischen Fehler (RMSE)

Variablenwichtigkeiten

Stellt Wichtigkeitswerte für jede Variable unter Verwendung der folgenden Metriken bereit:

- Relative Wichtigkeit
- Skalierte Wichtigkeit
- Prozentsatz

Zudem werden die besten 25 Variablen im Modell dargestellt.

Details zu Random Forest Classification – Binomial

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für **binomiale** Random Forest Classification-Modelle angezeigt:

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Anzahl der Beobachtungen
- R-Quadrat (R²)
- Logloss
- Area under the curve (AUC)
- Precision-Recall-Bereich unter der Kurve (PR AUC)
- Gini
- Mean Per Class Error

Maximum Metrics Threshold

Gibt den Training Maximum Metrics Threshold für Trainings-, Test- und „N-fach“-Daten mithilfe der folgenden Metriken an:

- max f1
- max f2
- max f0point5
- max accuracy
- max precision
- max recall
- max specificity
- max absolute_mcc
- max min_per_class_accuracy
- max mean_per_class_accuracy

Konfusionsmatrix

Stellt die Leistung eines Modells in einer Reihe von Trainings-, Test- und „N-fach“-Daten bereit, bei denen die tatsächlichen Werte bekannt sind.

Variablenwichtigkeiten

Stellt Wichtigkeitswerte für jede Variable unter Verwendung der folgenden Metriken bereit:

- Relative Wichtigkeit
- Skalierte Wichtigkeit
- Prozentsatz

Zudem werden die besten 25 Variablen im Modell dargestellt.

AUC-Kurven

Area under the curve (Fläche unter der Kurve); bestimmt, welches der Modelle die Klassen mithilfe der Trainings-, Test- und „N-fach“-Daten am besten vorhersagt.

Anhebungs-/Verstärkungskurven

Wertet die Fähigkeit des binären Klassifizierungsmodells aus, mithilfe der Trainings-, Test- und „N-Fold“-Daten Vorhersagen zu treffen.

Details zu Random Forest Classification – Multinomial

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für **multinomiale** Random Forest Classification-Modelle angezeigt:

Metriken

Stellt für folgende Informationen Trainings-, Test- und „N-fach“-Daten bereit:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Anzahl der Beobachtungen
- R-Quadrat (R²)
- Logloss
- Mean Per Class Error

Konfusionsmatrix

Stellt die Leistung eines Modells in einer Reihe von Trainings-, Test- und „N-fach“-Daten bereit, bei denen die tatsächlichen Werte bekannt sind.

Variablenwichtigkeiten

Stellt Wichtigkeitswerte für jede Variable unter Verwendung der folgenden Metriken bereit:

- Relative Wichtigkeit
- Skalierte Wichtigkeit
- Prozentsatz

Zudem werden die besten 25 Variablen im Modell dargestellt.

Details zu Principal Component Analysis

Auf dem Bildschirm „Modelldetail“ werden folgende Informationen für PCA-Modelle angezeigt:

Bedeutung der Komponenten

Zeigt die prinzipiellen Komponenten in der Reihenfolge der Wichtigkeit an, die auf folgenden Metriken basieren:

- Standardabweichung

- Streuungsverhältnis
- Kumulatives Verhältnis

Rotation

Stellt die Matrix der Variablenbelastungen dar, die Gewichtung mit der jede standardisierte Originalvariable multipliziert werden sollte, um die Komponentenpunktzahl zu erhalten.

4 - Analytics Scoring-Repository

In this section

Einführung in das Analytics Scoring-Repository	44
Modelle anzeigen	44
Modelldetails anzeigen	45
Model über Webbrowser-Client hinzufügen	46
Hinzufügen eines Modells über Spectrum Miner Spectrum Connector	48
Modell löschen	50

Einführung in das Analytics Scoring-Repository

Im Analytics Scoring-Repository werden die Analysemodelle zur Verwendung mit dem „PMML Model Scoring“-Schritt gespeichert. Zur Verwaltung der im Repository gespeicherten Modelle ist ein Webbrowser-Client vorhanden.

Anmerkung: Details zu unterstützten Modelltypen und Typzuordnungen finden Sie unter [Unterstützte Modellformate](#) auf Seite 13.

Auf den Webbrowser-Client des Analytics Scoring-Repository zugreifen

Geben Sie in einem Webbrowser die URL `http://<Servername>:<Port>/analyticsscoring` (z. B. `http://localhost:8080/analyticsscoring`) ein, um die Seite **Anmelden** anzuzeigen. Geben Sie im angezeigten Dialogfeld einen gültigen Benutzernamen und ein Kennwort für Spectrum™ Technology Platform ein. Im Browser wird dann eine Seite mit den im Repository vorhandenen Analysemodellen angezeigt.

Modelle anzeigen

Diese Schritte beschreiben, wie Sie die Liste der Modelle anzeigen können, die sich derzeit in der Analytics Scoring-Datenbank befinden.

1. Melden Sie sich mit dem Webbrowser-Client der Analytics Scoring-Datenbank an, indem Sie auf der Seite **Anmelden** einen gültigen Benutzernamen und ein Kennwort für Spectrum™ Technology Platform eingeben. Nach erfolgreicher Anmeldung wird die Seite **Datenbank** angezeigt, auf der sich eine Tabelle mit den Modellen in der Analytics Scoring-Datenbank befindet, die zur Verwendung mit dem „PMML Model Scoring“-Schritt verfügbar sind.
Details zum Anmeldevorgang über die Seite **Anmelden** finden Sie unter [Einführung in das Analytics Scoring-Repository](#) auf Seite 44.
2. Wählen Sie aus der Dropdown-Liste **Modelle pro Seite** unter der Modelltabelle die Anzahl von Modellen aus, die auf einer Seite angezeigt werden sollen. Es können 10, 20 oder 50 Modelle pro Tabellenseite angezeigt werden.
3. Sie können die Seitensteuerung neben der Dropdown-Liste **Modelle pro Seite** verwenden, um zwischen den verschiedenen Seiten zu navigieren.
Seitenwechsel über die Seitensteuerung sind nur möglich, wenn der in Schritt 2 ausgewählte Wert für die Anzahl von Modellen pro Seite kleiner ist als die Gesamtzahl der Modelle in der Analytics Scoring-Datenbank.

4. Klicken Sie auf die Spaltenüberschrift **Name**, um die Sortierung der aufgelisteten Modelle nach dem Wert **Name** zwischen absteigend und aufsteigend umzuschalten. Klicken Sie alternativ auf die Spaltenüberschrift **Typ**, um die Sortierung der aufgelisteten Modelle nach dem Wert **Typ** zwischen absteigend und aufsteigend umzuschalten.

Standardmäßig sind die Modelle nach dem Wert **Name** in aufsteigender Reihenfolge sortiert, bevor eine Spaltenüberschrift angeklickt wird.

Modelldetails anzeigen

Diese Schritte beschreiben, wie Sie Details zu einem Modell anzeigen können, das in der Analytics Scoring-Datenbank aufgelistet ist. Sie erhalten Beschreibungen und Details zu den Eingaben und Ausgaben des Modells.

1. Suchen Sie nach dem gewünschten Modell auf der Seite **Repository**.
Details zum Durchsuchen der Analytics Scoring-Datenbank finden Sie unter [Modelle anzeigen](#) auf Seite 44
2. Navigieren Sie zur Detailseite eines Modells, indem Sie entweder in der Spalte **Name** der Tabelle auf den Modellnamen klicken oder indem Sie das Kästchen des Modells aktivieren und dann auf die Schaltfläche **Modelldetails** über der Tabelle klicken.

Auf der Detailseite werden detaillierte Informationen zum Modell angezeigt.

Bezeichnung	Dies ist der Name des Modells.
Typ	Dies ist der Typ des Modells, wie unter Unterstützte Modellformate auf Seite 13 beschrieben.
Beschreibung	Dies ist ein kurzer Text zum Zweck des Modells.
Eingaben	Diese Tabelle enthält Informationen zu allen erforderlichen Feldern des Modells. Jede Zeile enthält einen Feldnamen und einen Typ, wie unter Unterstützte Modellformate auf Seite 13 beschrieben.
Ausgaben	Diese Tabelle enthält Informationen zu allen Ausgabefeldern, die vom Modell erstellt werden. Jede Zeile enthält einen Feldnamen und einen Typ, wie unter Unterstützte Modellformate auf Seite 13 beschrieben.

3. Um zurück zur Seite **Repository** zu navigieren, klicken Sie oben rechts auf die Schaltfläche **Schließen**.

Modell über Webbrowser-Client hinzufügen

Diese Schritte beschreiben, wie Sie ein neues Modell über den Webbrowser-Client des Analytics Scoring-Repositorys zum Analytics Scoring-Repository hinzufügen, um das Modell zur Bereitstellung in einem Datenfluss von Spectrum™ Technology Platform verfügbar zu machen.

1. Verwenden Sie die Schritte unter **Modelle anzeigen** auf Seite 44, um zur Seite **Repository** zu navigieren.
2. Klicken Sie auf die Schaltfläche **Modell veröffentlichen** (direkt über der Modelltabelle), um zur Seite **Neues Modell hinzufügen** zu navigieren.

Die Schaltfläche **Modell veröffentlichen** wird deaktiviert, wenn Modelle auf der aktuellen Seite der Modelltabelle ausgewählt wurden. Heben Sie die Auswahl dieser Modelle auf, um die Schaltfläche **Modell veröffentlichen** wieder zu aktivieren.

3. Auf der Seite **Neues Modell hinzufügen** befinden sich drei Eingabefelder: **Name**, **Beschreibung** und **Modell hochladen**. Geben Sie den Namen des neuen Modells in das Textfeld namens **Name** ein.

Modellnamen müssen eindeutig sein. Wenn der eingegebene Modellname bereits von einem anderen, nicht gelöschten Modell verwendet wird, das im Analytics Scoring-Repository gespeichert ist, wird das Textfeld mit einem Überprüfungsfehler gekennzeichnet.

Anmerkung: Leerzeichen am Anfang oder am Ende eines Modellnamens werden immer entfernt. Das bedeutet, dass der Fehler nicht behoben ist, wenn der eingegebene Namen bereits verwendet wird und Leerzeichen zum Anfang oder Ende des Namens hinzugefügt werden.

Anmerkung: Bei Modellnamen wird die Groß-/Kleinschreibung beachtet. Das bedeutet, dass Modelle den gleichen Namen haben können, solange sich die Groß-/Kleinschreibung der Namen unterscheidet. Beispielsweise können Modelle mit den Namen „mein Modell“, „Mein Modell“ und „mEin MODell“ gleichzeitig im Repository vorhanden sein.

Anmerkung: Die Namensüberprüfung erkennt Konflikte mit den Inhalten des Analytics Scoring-Repositorys nur zu dem Zeitpunkt, an dem die Seite geöffnet wird. Das bedeutet, dass die Seitenüberprüfung Änderungen nicht berücksichtigt, wenn das Repository sich seit dem Aufrufen der Seite **Neues Modell hinzufügen** z. B. durch Hinzufügen oder Löschen von Modellen durch andere Benutzer geändert hat. Wenn Sie allerdings auf die Schaltfläche **Veröffentlichen** klicken, wird das neue Modell erneut anhand der aktuellen Version des Analytics Scoring-Repositorys überprüft.

4. Geben Sie eine Beschreibung des neuen Modells in den Textbereich namens **Beschreibung** ein.

Modelle müssen eine Beschreibung aufweisen.

5. Fügen Sie eine Modelldefinition hinzu, indem Sie auf die Schaltfläche **Datei auswählen** klicken und ein Modell auswählen, das entweder im Dateiformat **QMML** auf Seite 13 oder **PMML** auf Seite 14 gespeichert ist.

Anmerkung: Das ausgewählte Modell muss kleiner als 50 MB sein.

Anmerkung: Wenn Sie einen älteren Browser verwenden, der die Dateiauswahl nicht unterstützt, müssen Sie die XML-Modelldefinition in den Textbereich mit Namen **Modell hochladen** eingeben.

Anmerkung: Wenn es sich bei dem ausgewählten Modell um ein **QMML** auf Seite 13-Modell handelt, muss die **QMML** auf Seite 13 mindestens einen kompilierten aktiven Regelsatz enthalten, damit das Modell erfolgreich veröffentlicht werden kann. Wenn mehrere kompilierte aktive Regelsätze vorhanden sind, wird der erste in der **QMML** auf Seite 13 gefundene Regelsatz für die Modelldefinition verwendet.

Anmerkung: Wenn Sie Regelsätze in Ihrer **QMML** auf Seite 13 kompilieren und/oder deren Aktivitätsstatus ändern möchten (oder Regelsätze hinzufügen oder löschen möchten), können Sie das Befehlszeilendienstprogramm „qsqmmledit“ von Spectrum Miner Ihrer Spectrum Miner-Installation verwenden. (Sie können die Hilfe zu dem Dienstprogramm in Ihrem Befehlszeilenfenster anzeigen, indem Sie „qsqmmledit.exe -help“ eingeben.)

Anmerkung: Die Modelldefinition wird nicht überprüft, bis Sie auf die Schaltfläche **Veröffentlichen** klicken.

6. Klicken Sie auf die Schaltfläche **Veröffentlichen**, um mit dem Hochladen des Modells in das Analytics Scoring-Repository zu beginnen.

Die Schaltfläche **Veröffentlichen** ist deaktiviert, bis Sie dem neuen Modell einen gültigen Namen gegeben sowie eine Beschreibung und eine Modelldefinition angegeben haben.

7. Warten Sie, bis das Aktivitätssymbol der Schaltfläche **Veröffentlichen** nicht mehr angezeigt wird. Während das Modell veröffentlicht wird, können Sie den Namen, die Beschreibung oder die Modelldefinition des neuen Modells nicht ändern. Sobald das Modell erfolgreich im Analytics Scoring-Repository veröffentlicht wurde, wird die Seite **Details** des neuen Modells angezeigt. Während der Veröffentlichung des Modells ist die Schaltfläche **Abbrechen** deaktiviert.

Anmerkung: Wenn bei der Veröffentlichung des Modells ein Fehler auftritt, beispielsweise die XML-Modelldefinition ungültig ist oder der Modellname bereits verwendet wird, wird das Modell nicht veröffentlicht und eine Warnmeldung mit Details zum Problem angezeigt. Klicken Sie auf **OK**, um die Warnmeldung zu schließen, und beheben Sie das Problem, bevor Sie versuchen, das Modell erneut zu veröffentlichen.

Hinzufügen eines Modells über Spectrum Miner Spectrum Connector

Die Integrationskomponente Spectrum Connector ermöglicht das direkte Veröffentlichen von **QMML** auf Seite 13-Modellen im Spectrum Analytics Scoring-Repository aus Spectrum Miner.

Die Integrationskomponente Spectrum Connector kann als Zusatz zu jeder vorhandenen Installation von Spectrum Miner Version 7.0B (oder höher) einschließlich Desktop-Installationen installiert werden.

Es gibt drei verschiedene Methoden, Modelle über die Integrationskomponente Spectrum Connector im Analytics Scoring-Repository zu veröffentlichen:

QMML-Modelle aus Spectrum Miner veröffentlichen auf Seite 48

QMML-Modelle aus Modellerstellungstools in Decision Studio veröffentlichen auf Seite 49

QMML-Modell aus dem Befehlszeilentool von Spectrum Miner veröffentlichen auf Seite 49

QMML-Modelle aus Spectrum Miner veröffentlichen

Diese Schritte beschreiben, wie Sie unter Verwendung von Spectrum Miner ein neues Modell zum Analytics Scoring-Repository hinzufügen können.

1. Öffnen Sie Spectrum Miner und navigieren Sie auf der Dateibrowser-Registerkarte zum Verzeichnis, das die zu veröffentlichende .qmml-Datei enthält.
2. Klicken Sie mit der rechten Maustaste auf die .qmml-Datei und wählen Sie „Veröffentlichen -> im Spectrum Analytics Scoring-Repository“ aus.
3. Geben Sie nach Aufforderung einen gültigen Spectrum-Benutzernamen und das Kennwort ein, um sich bei der Spectrum-Plattform anzumelden. Die Anmeldeaufforderung wird einmal pro Sitzung in Spectrum Miner angezeigt. Nachfolgende Veröffentlichungen von Modellen verwenden die zu Anfang eingegebenen Anmeldeinformationen.
4. Geben Sie auf der Registerkarte mit den Modelldetails den Modellnamen, die Modellbeschreibung und den **QMML** auf Seite 13-Regelsatz des Modells an, das im Spectrum Analytics Scoring-Repository veröffentlicht werden soll.
Der Name des Modells muss im Modell-Repository von Spectrum Advanced Analytics eindeutig sein.
5. Klicken Sie auf die Schaltfläche „Los“.

QMML-Modelle aus Modellerstellungstools in Decision Studio veröffentlichen

Diese Schritte beschreiben, wie Sie ein neues Modell direkt aus einem der Modellerstellungstools von Decision Studio zum Analytics Scoring-Repository hinzufügen können.

1. Erstellen Sie Ihr Modell im gewünschten Modellerstellungstool von Decision House.
2. Klicken Sie auf die Schaltfläche „[Modelltyp] in Spectrum veröffentlichen“, die nach der im vorherigen Schritt abgeschlossenen Modellerstellung aktiviert wird.
3. Geben Sie nach Aufforderung einen gültigen Spectrum-Benutzernamen und das Kennwort ein, um sich bei der Spectrum-Plattform anzumelden. Die Anmeldeaufforderung wird einmal pro Sitzung in Spectrum Miner angezeigt. Nachfolgende Veröffentlichungen von Modellen verwenden die zu Anfang eingegebenen Anmeldeinformationen.
4. Geben Sie auf der Registerkarte mit den Modelldetails den Modellnamen, die Modellbeschreibung und die Ausgabefelder des Modells an, das im Spectrum Analytics Scoring-Repository veröffentlicht werden soll.

Der Name des Modells muss im Modell-Repository von Spectrum Advanced Analytics eindeutig sein.

Anmerkung: Nur Felder, bei denen die zugehörigen Kästchen aktiviert sind, werden im Modell als Ausgabefelder eingeschlossen. Mindestens ein Feld muss ausgewählt werden.

5. Klicken Sie auf die Schaltfläche „Veröffentlichen“.

QMML-Modell aus dem Befehlszeilentool von Spectrum Miner veröffentlichen

Diese Schritte beschreiben, wie Sie ein neues Modell über das Befehlszeilentool zur Datenerstellung „spectrumpublish“, das von der Integrationskomponente Spectrum Connector von Spectrum Miner zur Verfügung gestellt wird, zum Analytics Scoring-Repository hinzufügen können.

Führen Sie in einer Eingabeaufforderung das Dienstprogramm „spectrumpublish“ aus, beispielsweise „\$QSHOME/server/qs7.0B/win32/bin/spectrumpublish -input [QMML-Dateipfad] -name [Modellname] -description [Modellbeschreibung] -ruleset [Regelsatzname] -username [Spectrum-Benutzername] -password [Kennwort]“

Anmerkung: Die Argumente „-input“, „-name“, „-username“, „-password“ und „-description“ müssen angegeben werden. Das Argument „-ruleset“ ist optional.

Anmerkung: Wenn die angegebene **QMML** auf Seite 13-Datei mehrere Regelsätze enthält, aber das Argument „-ruleset“ nicht angegeben wurde, wird nur der erste Regelsatz im **QMML** auf Seite 13-Dokument im Analytics Scoring-Repository veröffentlicht.

Modell löschen

Diese Schritte beschreiben, wie Sie Modelle aus der Analytics Scoring-Datenbank löschen können. Wenn Sie ein Modell aus der Datenbank löschen, wird es nicht mehr in neuen Datenflüssen von Spectrum™ Technology Platform bereitgestellt. Das Löschen eines Modells wirkt sich nicht auf Laufzeitvorgänge von vorhandenen Datenflüssen aus, die zur Verwendung des Modells konfiguriert sind. Sie können jedoch keine Änderungen mehr an den Einstellungen eines „PMML Model Scoring“-Schrittes in einem Datenfluss vornehmen, der zur Verwendung des Modells konfiguriert ist.

1. Verwenden Sie die Schritte unter **Modelle anzeigen** auf Seite 44, um zur Seite **Repository** zu navigieren.

Auf der Seite **Datenbank** werden in einer Modelltabelle Modelle angezeigt, die in der Analytics Scoring-Datenbank gespeichert sind. Wenn eine große Anzahl von Modellen in der Analytics Scoring-Datenbank gespeichert sind, erstreckt sich die Modelltabelle möglicherweise über mehrere Seiten.

2. Wählen Sie in der Modelltabelle Modelle aus, die Sie aus der Analytics Scoring-Datenbank löschen möchten.

Anmerkung: Nur Modelle, die auf der aktuellen Seite der Modelltabelle ausgewählt sind, werden gelöscht. Wenn die Modelltabelle beispielsweise über zwei Seiten mit Modellen verfügt und auf beiden Seiten Modelle ausgewählt wurden, werden beim Anklicken der Schaltfläche „Modell löschen“ nur die Modelle auf der aktuellen Seite gelöscht und alle auf der anderen Seite ausgewählten Modelle ignoriert.

3. Klicken Sie auf die Schaltfläche **Modell löschen** direkt über der Modelltabelle.

Die Schaltfläche **Modell löschen** wird deaktiviert, wenn keine Modelle auf der aktuellen Seite der Modelltabelle ausgewählt wurden.

4. In einem Dialogfeld wird um eine Bestätigung gebeten, dass die ausgewählten Modelle aus der Analytics Scoring-Datenbank gelöscht werden sollen. Klicken Sie auf die Schaltfläche **Ja**, um dies zu bestätigen und die ausgewählten Modelle aus der Analytics Scoring-Datenbank zu löschen. Klicken Sie auf die Schaltfläche **Nein**, um den Löschvorgang abzubrechen.

Nachdem die ausgewählten Modelle erfolgreich aus der Analytics Scoring-Datenbank gelöscht wurden, wird die Seite **Datenbank** mit der Modelltabelle aktualisiert.

Notices

© 2018 Pitney Bowes. Alle Rechte vorbehalten. MapInfo und Group 1 Software sind Marken von Pitney Bowes Software Inc. Alle anderen Marken und Markenzeichen sind Eigentum ihrer jeweiligen Besitzer.

USPS® Urheberrechtshinweise

Pitney Bowes Inc. wurde eine nicht-ausschließliche Lizenz erteilt, die die Veröffentlichung und den Verkauf von ZIP + 4® Postleitzahl-Datenbanken auf optischen und magnetischen Medien genehmigt. Folgende Marken sind Markenzeichen des 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. Hierbei handelt es sich jedoch nicht um eine vollständige Liste der Marken, die zum United States Postal Service gehören.

Pitney Bowes Inc. ist nicht-exklusiver Lizenznehmer von USPS® für die Verarbeitungsprozesse von NCOA^{Link}®.

Die Preisgestaltung jeglicher Pitney Bowes Softwareprodukte, -optionen und -dienstleistungen erfolgt nicht durch USPS® oder die Regierung der Vereinigten Staaten. Es wird auch keine Regulierung oder Genehmigung der Preise durch USPS® oder die US-Regierung durchgeführt. Bei der Verwendung von RDI™-Daten zur Berechnung von Paketversandkosten wird die Entscheidung, welcher Paketlieferdienst genutzt wird, nicht von USPS® oder der Regierung der Vereinigten Staaten getroffen.

Datenbereitstellung und Hinweise

Hier verwendete Datenprodukte und Datenprodukte, die in Software-Anwendungen von Pitney Bowes verwendet werden, sind durch verschiedene Markenzeichen und mindestens eines der folgenden Urheberrechte geschützt:

- © Copyright United States Postal Service. Alle Rechte vorbehalten.
- © 2014 TomTom. Alle Rechte vorbehalten. TomTom und das TomTom Logo sind eingetragene Marken von TomTom N.V.
- © 2016 HERE
- Fuente: INEGI (Instituto Nacional de Estadística y Geografía)
- Basierend auf elektronischen Daten © National Land Survey Sweden.
- © Copyright United States Census Bureau
- © Copyright Nova Marketing Group, Inc.
- Teile dieses Programms sind urheberrechtlich geschützt durch © Copyright 1993-2007 Nova Marketing Group Inc. Alle Rechte vorbehalten.
- © Copyright Second Decimal, LLC
- © Copyright Canada Post Corporation

Diese CD-ROM enthält Daten einer urheberrechtlich geschützten Datenerfassung der Canada Post Corporation.

© 2007 Claritas, Inc.

Das Geocode Address World Dataset enthält lizenzierte Daten des GeoNames-Projekts (www.geonames.org), die unter den Bedingungen der Creative Commons Attribution License ("Attribution License") bereitgestellt werden. Die Attribution License können Sie unter <http://creativecommons.org/licenses/by/3.0/legalcode> einsehen. Ihre Nutzung der GeoNames-Daten (wie im Spectrum™ Technology Platform Nutzerhandbuch beschrieben) unterliegt den Bedingungen der Attribution License. Bei Konflikten zwischen Ihrer Vereinbarung mit Pitney Bowes Software, Inc. und der Attribution License hat die Attribution License lediglich bezüglich der Nutzung von GeoNames-Daten Vorrang.



3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com