

Spectrum™ Technology Platform

バージョン 2019.1.0

Analytics Scoring ガイド



目次

1 - はじめに

Analytics Scoring モジュール	4
Analytics Scoring モジュールのコンポーネント	4

2 - Analytics Scoring モジュール

Binning Lookup	7
Java Model Scoring	8
PMML Model Scoring	9
Miner データセットからの読み込み	25
Miner データセットへの書き出し	28

3 - Machine Learning モデル管理

Machine Learning モデル管理へのアクセス	35
モデル評価	36

4 - Analytics Scoring リポジトリ

Analytics Scoring リポジトリの概要	45
モデルの表示	45
モデルの詳細の表示	46
Web ブラウザ クライアントによるモデルの追加	47
Spectrum Miner Spectrum Connector によるモデルの追加	49

1 - はじめに

このセクションの構成

Analytics Scoring モジュール	4
Analytics Scoring モジュールのコンポーネント	4

Analytics Scoring モジュール

Analytics Scoring モジュールとは

- Analytics Scoring モジュールにより、QMML (Spectrum Miner の独自モデル形式) または PMML (業界標準の Predictive Model Markup Language) 形式で定義した予測モデルを使用して、データフロー内で評価を行えます。データフローにより、Miner 予測分析ソフトウェアで使用される Miner データセットにあるデータを書き込んだり取得したりできます。こうしたモデリング手法では、データ洞察チームが業界標準のデータ モデリング ツールで作成したモデルの予測またはスコア化された出力を追加して、データを充実させることができます。例として、既存顧客の離反リスクを計算するためのモデルや、消費者の信用評価を査定するための信用スコアリングモデルが挙げられます。
- また、Java スコアリングで、Machine Learning モジュールに非常に適したモデルを利用できるようになります。
- これにより、本番 IT 環境にある既存のモデルを使用して、新しいデータのスコアリングを行えます。

Analytics Scoring モジュールのコンポーネント

Analytics Scoring モジュールは、次のコンポーネントで構成されます。

- **Binning Lookup** — Machine Learning モジュールの Binning ステージによってデータフロー内に作成された既存のビンを使用して、以前に定義されたビニングを新しいデータに適用するために使用できるステージ。
- **Java Model Scoring** — 機械学習モデルのフィッティングを行ったときに作成された式を使用して、新しいデータをスコアリングするために使用できるステージ。
- **PMML Model Scoring** — Analytics Scoring リポジトリに保存された任意のモデルを、データフローのコンテキストで評価するために使用できるステージ。
- **Read from Miner Dataset** — データフロー内で使用されるフォーカス ファイルからのデータの読み取りに使用できるステージ。
- **Write to Miner Dataset** — データフローからフォーカス ファイルへのデータの書き出しに使用できるステージ。
- **Machine Learning モデル管理** — Spectrum™ Technology Platform サーバー上のすべての機械学習モデルを管理するモデル評価と、Spectrum™ Technology Platform サーバー上のすべてのビニングを管理するビニング管理が含まれているリポジトリ。

- **Analytics Scoring** リポジトリ — Analytics Scoring モジュールで使用可能なすべてのモデルが格納された中央リポジトリ。ユーザは、Webクライアントによってリポジトリを管理できます。

Java Model Scoring を使用したデータのスコアリングが含まれている、教師あり学習と教師なし学習の例については、『[Machine Learning ガイド](#)』の「データサイエンスのデモンストレーションフロー」を参照してください。

2 - Analytics Scoring モジュール

このセクションの構成

Binning Lookup	7
Java Model Scoring	8
PMML Model Scoring	9
Miner データセットからの読み込み	25
Miner データセットへの書き出し	28

Binning Lookup

Binning Lookup の概要

Binning Lookup は、**Binning** ステージを使用してデータフローで作成された既存のビンを使用して、以前に定義されたビンングを新しいデータに適用します。

ビンングのプロパティの定義

1. **[プライマリ ステージ]** > **[展開済みステージ]** > **[Machine Learning]** の下で、**[Binning Lookup]** ステージをクリックしてキャンバス上にドラッグし、データフロー内の所望の位置に配置して、他のステージに接続します。

注: 入力ステージにはビンングされるデータが含まれている必要があります。1つの出力ステージがビンングされた出力のために必要です。オプションで、ビンング サマリを取得するために第 2 の出力ステージを接続できます。

2. **[ビンング名]** ドロップダウンから適切なものを選択します。
これらは、Binning ステージを使用するデータフローによって作成された既存のビンの名前です。
[ビンング タイプ] および **[説明]** フィールドは、選択したビンング名のビンによってインポートされるため、編集できません。
3. **[入力]** グリッドには、Binning ステージでのビンングのために含まれた各フィールドがデータ タイプとともに表示されます。
4. **[OK]** をクリックして、設定を保存します。

ビンング出力

このタブには、Binning Lookup ステージによってビンングされるフィールドとデータ タイプが表示されます。ビンングステージを使用して生成される出力の詳細については、**ビンング出力** を参照してください。現在、ビンングされたフィールドをビンング出力タブで編集するオプションが用意されています。ユーザは "ビンングされた Spectrum フィールド" を使用して、ビンングされ

たフィールドに新しい名前を付けることができます。ビニングされたフィールドを含めたり除外したりするオプションも用意されています。

Java Model Scoring

Java Model Scoring の概要

Java Model Scoring では、機械学習モデルの適合を行った時に作成された式を使用して、新しいデータをスコアリングすることができます。

注：モデルを Java Model Scoring ステージで使用できるようにするにはまず、Machine Learning モデル管理を介してそれを公開する必要があります。詳細については、[モデル評価の概要](#)（36ページ）を参照してください。

データをスコアリングするには、**[Java Model Scoring オプション]** ダイアログで2つのタブの設定を終える必要があります。まずモデルとそのタイプを指定してから、モデルのフィールドが Spectrum™ Technology Platform のフィールドに正しくマッピングされていることを確認します。続いて、出力を設定します。ジョブに含めるフィールドを選択し、ジョブを実行します。**[モデル出力]** タブには、Spectrum™ Technology Platform とモデルのデータタイプのマッピングが含まれます。

ジョブに、ファイルまたはテーブルに出力を取得するステージが含まれている場合は、後続のデータフローまたは Web サービスでその出力を使用できます。

モデルのプロパティの定義

1. **[プライマリ ステージ] > [展開済みステージ] > [Advanced Analytics]** の下で、**[Java Model Scoring]** ステージをクリックしてキャンバス上にドラッグし、データフロー内の所望の位置に配置して、入出力ステージに接続します。

注：入力ステージは、モデルの目標フィールドと入力変数フィールドの両方を含むデータソースでなければなりません。ジョブをバッチモードで実行する場合は、モデルスコアを取得するための出力ステージも必要です。あるいは、Spectrum™ Technology Platform の Web サービスを使用して、リアルタイムでデータをスコアリングします。

2. **Java Model Scoring** ステージをダブルクリックして、**[Java Model Scoring オプション]** ダイアログ ボックスを表示します。
3. オプション: スコアリング対象のモデルのタイプを **[タイプ フィルタ]** ドロップダウンから選択します。
4. モデルのスコアリングに使用する **[タイプ フィルタ]** を選択します。
5. **[モデル名]** をドロップダウンから選択します。
6. スコアリング対象のモデルのタイプを、**[モデル タイプ]** フィールドに入力します。
7. オプション: モデルの **[説明]** を入力します。
8. モデルの適合に使用したデータに含まれないカテゴリ レベルの行について、**[Predicted_Value]** 列にデータを返すには、**[不明なカテゴリ レベルを無視]** をオンにします。
このボックスをオフのままにすると、これらの行の **[Predicted_Value]** 列は「Null/NA」を返します。
9. **[入力]** テーブルには、モデルの入力フィールドに関する情報が表示されます。
これらのフィールドとそのデータ タイプは、**Spectrum** のフィールドとデータ タイプに自動的にマッピングされます。
10. **[OK]** をクリックしてこれらのオプションを保存するか、次のタブで操作を続行します。

モデル出力

[出力] テーブルには、モデルの出力フィールドに関する情報が表示されます。これらのフィールドとそのデータ タイプは、**Spectrum** のフィールドとデータ タイプに自動的にマッピングされません。

1. モデルの出力に含めるデータの各フィールドに対し、**[含める]** をクリックします。
2. **[OK]** をクリックしてモデルを保存します。

PMML Model Scoring

PMML Model Scoring の概要

PMML Model Scoring ステージは、**Analytics Scoring** リポジトリに公開されている分析モデルを、データ フローのコンテキストで評価することができます。単一のデータ行を対象とし、各行の

フィールドをモデルへの入力として使用します。ユーザが選択したモデル出力が、出力チャンネルに書き出されます。

注: サポートされるモデル タイプとタイプ マッピングの詳細については、[サポートされるモデル フォーマット](#) (14ページ) を参照してください。

モデルの展開

以下では、分析モデルをデータフローの一部として展開するために PMML Model Scoring ステージを設定する手順について説明します。

1. **[プライマリ ステージ]** > **[展開済みステージ]** > **[Advanced Analytics]** の下で、**[PMML Model Scoring]** ステージをクリックしてキャンバス上にドラッグし、データフロー内の所望の位置に配置して、入出力ステージに接続します。
2. PMML Model Scoring ステージをダブルクリックして、**[PMML Model Scoring オプション]** ダイアログ ボックスを表示します。

このオプション ダイアログにはデフォルトで、使用可能なモデルの一覧の中の最初のモデルの詳細が表示されます。

3. **[タイプ フィルタ]** ドロップダウンをクリックして、フィルタ対象とするモデル タイプを選択します。

Analytics Scoring リポジトリ 内にモデルが少なくとも 1 つは存在するモデル タイプのみが一覧表示されます。

4. **[モデル]** ドロップダウンをクリックして、展開するモデルを選択します。選択したモデルの詳細がダイアログに表示されます。

モデル タイプ 選択したモデルのタイプ ([サポートされるモデル フォーマット](#) (14ページ) を参照)

説明 モデルの目的に関する簡単な説明

入力 モデルの全必須入力フィールドに関する情報が記載されたテーブル。各行に 1 つの入力フィールドに関する情報が記載されています。モデルの入力フィールド名は公開時に、有効な Spectrum フィールド名に自動的にマッピングされます ([Web ブラウザクライアントによるモデルの追加](#) (47ページ) または [Spectrum Miner Spectrum Connector によるモデルの追加](#) (49ページ) を参照)。

モデル フィールド名 モデル定義で指定されているフィールドの名前

Spectrum フィールド名 Spectrum Platform で使用されるフィールドの名前

モデル フィールド タイプ モデル定義で指定されているフィールドのタイプ

Spectrum フィールド タイプ モデル フィールド タイプにマッピングされる
 プ **Spectrum** フィールド タイプ (**QMML** (14ページ)
 および**PMML** (16ページ) を参照)

5. **[環境設定]** タブを選択します。

モデルの **出力** に関する詳細情報がテーブルに表示されます。モデルの全出力フィールドに関する情報が含まれます。各行に 1 つの出力フィールドに関する情報が記載されています。モデルの出力フィールド名は公開時に、有効な **Spectrum** フィールド名に自動的にマッピングされます (**Web ブラウザクライアントによるモデルの追加** (47ページ) または**Spectrum Miner Spectrum Connector によるモデルの追加** (49ページ) を参照)。

モデル フィールド名	モデル定義で指定されているフィールドの名前
Spectrum フィールド名	Spectrum Platform で使用されるフィールドの名前
モデル フィールド タイプ	モデル定義で指定されているフィールドのタイプ
Spectrum フィールド タイプ	モデル フィールド タイプにマッピングされる Spectrum フィールド タイプ (QMML (14ページ) および PMML (16ページ) を参照)
含める	この出力を使用するかどうかを指定するチェックボックス。

6. オプション: **[出力]** テーブルの任意の行で **[含める]** 列のチェックを外すと、その出力は除外されます (出力チャンネルに書き出されません)。

少なくとも 1 つの出力は選択されたままにしておく必要があります。すべての出力を除外すると、バリデーションエラーの記号が **[出力]** テーブルの横に表示されます。これは、現在のモデル設定が無効で、モデルが展開できないことを意味します。バリデーションエラーの記号は、エラーが修正されるまで表示されたままとなります。

7. オプション: **[出力]** テーブルの任意の行で **[Spectrum フィールド名]** 列をクリックすると、必要に応じてフィールド名を変更できます。

2 つの出力で同じ **[Spectrum フィールド名]** を共有することはできず、また、**[Spectrum フィールド名]** は **Spectrum™ Technology Platform** の標準フィールド命名規則に従う必要があります。何らかのバリデーションエラーが検出された場合は、バリデーションエラーの記号が **[出力]** テーブルの横に表示されます。マウスカーソルをバリデーションエラー記号の上に合わせると、エラーの詳細が表示されます。

注: 出力の **[Spectrum フィールド名]** を変更しても、ステージの特定のインスタンスにしか影響はありません。Analytics Scoring リポジトリが更新されることはありません。

注：Spectrum™ Technology Platform フィールド名は、以下の条件を満たす必要があります。

- 一意である。
- 空ではない。
- 英数字、ピリオド、アンダースコアのみを含む。
- ピリオドで始まらない。

8. 最後に **[OK]** をクリックして、選択したモデルと設定を保存します。

PMML Model Scoring 設定の再構成

以下では、PMML Model Scoring ステージを再構成して、PMML Model Scoring ステージから生成される出力やその名前を変更したり、このステージで評価する展開モデルを変更したりする手順について説明します。

1. PMML Model Scoring ステージをダブルクリックして、**[PMML Model Scoring オプション]** ダイアログ ボックスを表示します。

このオプション ダイアログには、それまで選択、設定、展開されていたモデルが表示されます。

注：**[PMML Model Scoring オプション]** ダイアログを開く前に、選択されていたモデルが Analytics Scoring リポジトリから削除されていた場合は、**[モデル]** ドロップダウンリストの横にバリデーションエラーの記号が表示されます。これは、そのモデル設定に新たに加えられた変更は展開できないことを意味します。**[キャンセル]** をクリックすると **[PMML Model Scoring オプション]** ダイアログが終了し、削除されたモデルを以前の設定どおりにデータフロー内で使用できます。ステージ設定に新たに変更を加えるには **[モデル]** ドロップダウンリストから別の (削除されていない) モデルを選択します。モデルに変更を加えた後、このステージをさらに再構成する際には、削除されたモデルはもう **[モデル]** ドロップダウン リストに表示されません。

2. オプション:**[モデル]** タブで、**[モデル]** ドロップダウンリストから別のモデルを選択し、データフロー内で使用するモデルを変更します。

注：**[モデル]** ドロップダウン リストで選択モデルを変更すると、前に選択されていたモデルに対する設定の変更はすべて破棄されます。そのモデルを再度選択すると、モデルの出力設定はデフォルトに戻ります。**[キャンセル]** をクリックすると、**[PMML Model Scoring オプション]** ダイアログを開いた後に加えた変更がすべて取り消され、未確定の変更をすべて元に戻すことができます。

注：モデルの選択時に、選択したモデルが **[PMML Model Scoring オプション]** ダイアログを開いた後に削除されていた場合は、**[モデル]** ドロップダウンリストの横にバリデーションエラーの記号が表示されます。この場合、削除されていないモデルを選択するまで **[入力]** と **[出力]** は使用できず、また、モデル設定の再構成を適用することはできません。

3. **[環境設定]** タブを選択します。
4. 選択したモデルの出力設定に、必要な変更を加えます。
例えば、**[Spectrum フィールド名]** を変更したり、その **[含める]** チェックボックスのオン/オフを切り替えてデータフローに含めるかどうかを変更します。

出力に対する変更は必ず、次のバリデーションルールに従う必要があります。まず、少なくとも1つの出力を含める必要があります。また、出力の **[Spectrum フィールド名]** は一意で、標準の Spectrum™ Technology Platform フィールド命名規則に従う必要があります。何らかのバリデーションエラーが検出された場合は、バリデーションエラーの記号がテーブルの横に表示されます。マウスカーソルをバリデーションエラー記号の上に合わせると、エラーの詳細が表示されます。

注：Spectrum™ Technology Platform フィールド名は、以下の条件を満たす必要があります。

- 一意である。
- 空ではない。
- 英数字、ピリオド、アンダースコアのみを含む。
- ピリオドで始まらない。

5. 必要な変更をすべて加えたら、**[OK]** ボタンをクリックして変更を適用します。

出力

PMML Model Scoring ステージは、選択したモデルの出力フィールドを返します。また、PMML Model Scoring ステージは、レコードの処理に失敗すると、**Status**、**Status.Code**、および **Status.Description** の各フィールドを返します。これらのフィールドは、このステージがレコードの処理に失敗した理由に関する情報を提供します。

フィールド名	説明
ステータス	<p>評価試行の成功または失敗を報告します。</p> <p>NULL 成功</p> <p>F 失敗</p>
Status.Code	<p>失敗またはエラーの理由</p> <p>InputConversionFailed Spectrum 入力フィールドから要求されたモデル フィールドへのタイプの変換に失敗しました。</p> <p>ModelEvaluationFailed 指定されたレコードに対するモデルの評価に失敗しました。</p>
Status.Description	問題の説明

サポートされるモデル フォーマット

PMML Model Scoring は、**QMML** (14ページ) と **PMML** (16ページ) の両方のファイル形式で保存された分析モデルの展開をサポートします。

QMML (14ページ) モデルは、Spectrum Miner で作成してエクスポートできます。Spectrum Miner からエクスポートされたすべての種類の分析モデルおよびセグメンテーションが、PMML Model Scoring ステージでサポートされ、Spectrum™ Technology Platform データフロー内に展開可能です。

PMML (16ページ) モデルは、多くの商用およびオープン ソースのモデリング ツールで作成してエクスポートできます。

注：モデルは、UTF-8 のエンコーディングでのみ公開できます。

QMML

QMML は、Spectrum Miner から生成されたモデル結果を表現するために使用されるプロプライエタリな XML ベースのファイル形式です。

Spectrum Miner からエクスポートされたすべての種類の分析モデルおよびセグメンテーションが、PMML Model Scoring ステージでサポートされ、Spectrum™ Technology Platform データフロー内に展開可能です。

タイプ マッピング

QMML モデルの入力と出力は、Spectrum™ Technology Platform フィールド タイプに自動的にマッピングされます。

QMML フィールド タイプ	Spectrum™ Technology Platform フィールド タイプ
integer	integer
real	double
string	string
date	datetime

サポートされるモデル

Spectrum Miner 内で構築されたすべてのモデル タイプ (デシジョン ツリー、スコアカード、クラスタ分析、Naive Bayes モデルなど) が、Analytics Scoring モジュールによって Miner Model として解釈されます。

Miner Model

Miner Model は、Spectrum Miner からエクスポートされた任意のタイプの **QMML** (14ページ) モデルです。デシジョン ツリー、スコアカード、クラスタ分析、Naive Bayes モデリング ツールから生成されたものなどがあります。

サポートされない機能

Spectrum Miner からエクスポートされたすべてのコンパイル済み **QMML** (14ページ) がサポートされます。

モデル出力

フィールド	説明
<動的フィールド>	各 QMML (14ページ) モデル出力に対してフィールドが出力されます。

PMML

Predictive Model Markup Language (PMML) は、Data Mining Group によって開発された XML ベースのファイル形式です。データマイニングおよび機械学習アルゴリズムによって生成されたモデルを記述し、交換する手段を、アプリケーションに提供します。PMML ファイルは、多くの商用およびオープンソースのモデリングツールで作成してエクスポートできます。

タイプマッピング

PMML モデルの入力と出力は、Spectrum™ Technology Platform フィールドタイプに自動的にマッピングされます。PMML モデルの入出力はすべて、サポートされているタイプである必要があります。

PMML フィールドタイプ	Spectrum™ Technology Platform フィールドタイプ
string	string
integer	integer
float	float
double	double
boolean	boolean
date	date
dateDaysSince[1960]	date
dateDaysSince[1970]	date
dateDaysSince[1980]	date
dateDaysSince[0]	未サポート
time	time
timeSeconds	time
dateTime	datetime

PMML フィールド タイプ	Spectrum™ Technology Platform フィールド タイプ
dateTimeSecondsSince[1960]	datetime
dateTimeSecondsSince[1970]	datetime
dateTimeSecondsSince[1980]	datetime
dateTimeSecondsSince[0]	未サポート

サポートされるモデル

Analytics Scoring モジュールは現在、以下のセクションで説明されている PMML モデル タイプをサポートします。

アソシエーションルール

PMML (16ページ) アソシエーションルール モデルは、ある項目集合が別の項目集合に関連付けられる場合のルールを表します。例えば、ある製品または製品集合が、他のある製品集合と組み合わせて購入されるケースが多いことを、ルールで表すことができます。これは、マーケットバスケット分析としても知られます。アソシエーションルール モデルには通常、2つの変数があります。1つは、レコードをトランザクションにグループ化するために使用され、もう1つは各レコードを一意に識別するために使用されます。

モデル要素

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

サポートされない機能

項目を識別するフィールドに対し、文字列以外のフィールド タイプはサポートされていません。複数のフィールドによるレコードのグループ化はサポートされていません。

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つアソシエーションルール モデルはサポートされていません。

モデル出力

サポートされるモデル出力機能	説明
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
entityId	ウィニングルール (デフォルト)、またはランク値によって指定されたルールの ID。選択されたルールに ID がない場合は、1 から始まるインデックスが返されます。
ruleId	これは entityId オプションと同一で、PMML 4.2 から非推奨となっています。現時点で使用はサポートされていますが、entityId を使用することが推奨されます。
affinity	ウィニングルール (デフォルト)、またはランク値によって指定されたルールのアフィニティ。
antecedent	ウィニングルール (デフォルト)、またはランク値によって指定されたルールの前件。この出力は、値のカンマ区切り文字列として書式設定されます。
consequent	ウィニングルール (デフォルト)、またはランク値によって指定されたルールの後件。この出力は、値のカンマ区切り文字列として書式設定されます。
ルール	ウィニングルール (デフォルト)、またはランク値によって指定されたルール。この出力は、ルールの説明を、{<antecedent>}->{<consequent>} の形式で返します。
信頼性	ウィニングルール (デフォルト)、またはランク値によって指定されたルールの信頼度。
support	ウィニングルール (デフォルト)、またはランク値によって指定されたルールのサポート。
リフト	ウィニングルール (デフォルト)、またはランク値によって指定されたルールのリフト。
leverage	ウィニングルール (デフォルト)、またはランク値によって指定されたルールのレバレッジ。

クラスタ化

PMML (16ページ) クラスタ化モデルは、特定のレコードに対して最適なマッチング クラスタを決定します。この決定は、クラスタリングに使用される距離または類似性の尺度に基づいて行われます。クラスタは、類似するデータのサブセットです。クラスタリング (教師なし学習とも呼ばれます) は、データセットをグループに分割する処理です。グループのメンバーはできるだけ似た

ものになるように選択され、また異なるグループどうしはできるだけ似ていないものになるように選択されます。

モデル要素

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

サポートされない機能

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つクラスタリングモデルはサポートされていません。

モデル出力

サポートされるモデル出力機能	説明
predictedValue	クラスタリングに使用される距離または類似性の尺度に基づいて決定される最適なマッチングクラスタ。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
predictedDisplayValue	モデルからの予測値を表現するために使用される、人間が識別しやすい値。
entityId	存在する場合は、ウィニング/予測クラスタの 1 ベースのインデックス (暗黙的識別子)。
affinity	与えられたレコードからモデルに定義されている予測クラスタへの距離または類似性の値。

分類ツリー

PMML (16ページ) 分類ツリー モデルは、1つ以上の独立変数からカテゴリ依存変数のメンバーシップを予測します。

モデル要素

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

サポートされない機能

missing value strategy が "aggregateNodes" または "weightedConfidence" の分類ツリーはサポートされていません。

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つ分類ツリー モデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは predictedValue 機能と同義です。

サポートされるモデル出力機能

説明

predictedValue	メンバーシップを予測するカテゴリ依存変数。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
predictedDisplayValue	モデルからの予測値を表現するために使用される、人間が識別しやすい値。
probability	予測値の統計的確率。モデルの中で、各予測カテゴリに対して 1 つ、またはランク別に、複数の probability 出力を指定できます。
residual	予測カテゴリまたはランクに対する、出力値 probability の剰余 (1 - probability)。モデルの中で、各予測カテゴリに対して 1 つ、またはランク別に、複数の residual 出力を指定できます。
entityId	存在する場合は、予測結果のツリー ノードの ID です。

回帰ツリー

PMML 回帰ツリー モデルは、1 つ以上の変数から数値依存の変数 (家屋の価格など) の値を予測します。1 つ以上の予測因子に基づくデシジョンツリー モデルを構築することによって、これが行われます。

モデル要素

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

サポートされない機能

integer または float 型のターゲット フィールドを持つ回帰ツリーはサポートされていません(ただし、<Targets> 要素が適切な castInteger 属性とともに指定されている場合を除きます)。

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つ回帰ツリー モデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは predictedValue 機能と同義です。float 型のターゲット フィールドは、必ず integer 型に変換されます。

サポートされるモデル出力機能

説明

predictedValue	予測する数値依存変数。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
entityId	存在する場合は、予測結果のツリー ノードの ID です。

Naive Bayes

PMML (16ページ) Naive Bayes モデルはベイズの定理を用いて、1つ以上の予測フィールドによって与えられた証拠から、ターゲットの値を予測します。Naive Bayes モデルでは、モデルによって検討される値が有限個となるように、ターゲット フィールドが離散化されている必要があります。予測フィールドは、離散値でも連続値でもかまいません。

モデル要素

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

サポートされない機能

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つ Naive Bayes モデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは predictedValue 機能と同義です。

サポートされるモデル出力機能	説明
----------------	----

predictedValue	メンバーシップを予測するカテゴリ変数。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
predictedDisplayValue	モデルからの予測値を表現するために使用される、人間が識別しやすい値。
probability	予測値の統計的確率。
residual	予測値に対する、出力値 probability の剰余 (1 - probability)。

回帰

PMML (16ページ) 回帰モデルは、1 つ以上の独立変数から依存変数の数値を予測します。

モデル要素

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

サポートされない機能

属性 "normalizationMethod" の値が "simplemax"、"probit"、"cloglog"、または "loglog" に設定されている回帰モデルはサポートされていません。

integer または float 型のターゲット フィールドを持つ回帰モデルはサポートされていません (ただし、<Targets> 要素が適切な castInteger 属性とともに指定されている場合を除きます)。

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つ回帰モデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは predictedValue 機能と同義です。float 型のターゲット フィールドは、必ず integer 型に変換されます。

サポートされるモデル出力機能 説明

predictedValue	予測する数値依存変数。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。

回帰分類

PMML (16ページ) 回帰分類は、複数の回帰式からの出力を組み合わせることでカテゴリ値を予測します。

モデル要素

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

サポートされない機能

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つ回帰モデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは predictedValue 機能と同義です。

サポートされるモデル出力機能 説明

predictedValue	メンバーシップを予測するカテゴリ依存変数。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
predictedDisplayValue	モデルからの予測値を表現するために使用される、人間が識別しやすい値。
probability	予測値の統計的確率。モデルの中で、各予測カテゴリに対して 1 つ、またはランク別に、複数の probability 出力を指定できます。

サポートされるモデル出力機能

説明

residual	予測値に対する、出力値 probability の剰余 (1 - probability)。モデルの中で、各予測カテゴリに対して 1 つ、またはランク別に、複数の residual 出力を指定できます。
----------	---------------------------------------------------------------------------------------------------------------------

スコアカード

PMML (16ページ) スコアカードモデルは、リスクまたはデフォルト確率の計算に主に使用される、回帰に基づくモデル化手法です。

モデル要素

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

サポートされない機能

<DerivedField> 要素への参照を含む <MiningSchema> 要素を持つスコアカードモデルはサポートされていません。

モデル出力

デフォルトでは、ターゲット フィールドが出力フィールドとなります。これは **predictedValue** 機能と同義で、タイプは必ず 'double' になります。

サポートされるモデル出力機能

説明

predictedValue	計算されたスコア。
transformedValue	予測モデル出力に適用された変換式によって生成された値。
decision	分類値となった予測モデル出力に適用された式によって生成された値。
reasonCode	指定された場合は、最終スコアに対する貢献度順にスコアを決定するために使用される特性に関連付けられた理由コード。要求理由コードのランクが指定できます。指定しない場合は、最高ランクの理由コードがデフォルトで返されます。

Miner データセットからの読み込み

Read from Miner Dataset の概要

Read from Miner Dataset ステージは、Miner データセットからフィールド情報とデータを抽出できます。

Miner データセットからの読み込み

以下では、Miner データセットからフィールド情報とデータを読み込む手順について説明します。

1. **[プライマリ ステージ]** / **[展開済みステージ]** / **[Advanced Analytics]** の下で、**[Read from Miner Dataset]** ステージをキャンバス上にドラッグし、データフローに接続します。
2. **[Read from Miner Dataset]** ステージをダブルクリックして、**[Read from Miner Dataset オプション]** ダイアログを表示します。
3. **[ファイル名]** フィールドの中の **[...]** ボタンをクリックして、**[ファイルを開く]** ダイアログを表示します。
4. **[ファイルを開く]** ダイアログで、Miner データセットを含むフォーカス ファイルを探して開きます。
フォーカス ファイルは .ftr という拡張子を持ちます。
5. **[Read from Miner Dataset オプション]** ダイアログの **[フィールド]** タブをクリックして、フィールド テーブルを表示します。
Read from Miner Dataset ステージを初めて設定する場合と、新しいフォーカス ファイルを選択した場合 (手順 4 (25ページ)) は、フィールド テーブルは空です。
6. **[再生成]** ボタンをクリックします。
フィールド テーブルには、**Read from Miner Dataset** ステージが Miner データセットから読み込むすべてのフィールドが列挙されます。

注：フォーカスが選択されていない場合は、**[再生成]** ボタンは無効になります (手順 3 (25ページ) と 4 (25ページ))。

注：ステージにおいて Miner データセットからのフィールドの取得に失敗すると、エラー ダイアログが表示されます。有効なフォーカス ファイルを選択してもう一度やり直してください。

7. 必要に応じて、使用可能なオプションによってフィールドリストを変更します (**[フィールド] タブ** (26ページ) を参照)。

8. 最後に、**[OK]** ボタンをクリックして設定を適用します。

バリデーションエラーが検出された場合は、バリデーションエラーメッセージが表示されます。バリデーションエラーを修正してから、もう一度 **[OK]** ボタンをクリックします。バリデーションエラーが検出されなければ、**[Read from Miner Dataset オプション]** ダイアログが閉じます。

注：ステージを再設定するには、**[Read from Miner Dataset]** ステージをダブルクリックして、**[Read from Miner Dataset オプション]** ダイアログを再表示します。

[フィールド] タブ

[フィールド] タブには、**Read from Miner Dataset** ステージが Miner データセットから読み込むすべてのフィールドが列挙されたテーブルがあります。以下の表に、テーブルの内容を変更するために使用可能な、フィールドに対するオプションを示します。

オプション名	説明
再生成	現在定義されているフィールドを、Miner データセットから読み込んだもので置き換えます。

以下の表に、全フィールドのプロパティを示します。

列名	説明	編集可能
Miner フィールド名	Miner データセットで指定されているフィールドの名	いいえ前。

列名	説明	編集可能
Spectrum フィールド名	データフロー チャンネルで作成されるフィールドの名前は、 [Miner フィールド名] が Spectrum™ Technology Platform の命名規則に適合しない場合、 [Spectrum フィールド名] は異なるものになります。セルの中のテキストを編集して、データフロー チャンネルに対する所望のフィールド名を指定します。Spectrum™ Technology Platform フィールド名は、以下の条件を満たす必要があります。 <ul style="list-style-type: none"> 一意である。 空ではない。 英数字、ピリオド、アンダースコアのみを含む。 ピリオドで始まらない。 	はい
Spectrum フィールド タイプ	データフローチャンネルで作成されるフィールドのデータタイプ (出力 (27ページ) 参照)。	いいえ
含める	データフロー チャンネルにこのフィールドを含めるかどうかを指定するチェック ボックス。	はい

注：テーブルの列ヘッダをクリックすることによって、フィールドを **[Miner フィールド名]**、**[Spectrum フィールド名]**、または **[Spectrum フィールド タイプ]** で並べ替えることができます。

出力

Miner フィールド タイプと Spectrum フィールド タイプの間の以下のタイプ マッピングは自動的に行われます

表 1 : Field Mappings

Miner タイプ	Spectrum タイプ
integer	integer
string	string

Miner タイプ	Spectrum タイプ
real	double
date	datetime

フィールド名に **Spectrum** では無効な記号が含まれる場合、その記号は自動的に `_` に置き換えられます。複数のフィールドの変換後の名前が同じになる場合は、名前にインデックスが付加されます。例えば、次のフィールド名があるとします。

```
{x}Field{x}
```

```
{y}Field{y}
```

ここで `{x}` と `{y}` は、**Spectrum** フィールド名ではサポートされていない文字です。この場合、出力チャンネルにおいて変換後のフィールド名は、

```
_Field_1
```

```
_Field_2
```

となります。

Miner データセットへの書き出し

Write to Miner Dataset の概要

Write to Miner Dataset ステージは、**Spectrum** データフローからの指定されたフィールドとデータを保持する **Miner** データセットを作成できます。

Miner データセットへの書き出し

以下では、**Miner** データセットにフィールド情報とデータを書き出す手順について説明します。

1. **[プライマリ ステージ]** > **[展開済みステージ]** > **[Advanced Analytics]** の下で、**[Write to Miner Dataset]** ステージをキャンバス上にドラッグし、データフローに接続します。

2. **[Write to Miner Dataset]** ステージをダブルクリックして、**[Write to Miner Dataset オプション]** ダイアログを表示します。
3. **[ファイル名]** フィールドの中の **[...]** ボタンをクリックして、**[ファイルの保存]** ダイアログを表示します。
4. **[ファイルの保存]** ダイアログで、新しいフォーカス ファイルを保存する場所を選択し、適切なファイル名を指定します。
フォーカス ファイルは `.ftr` という拡張子を持つ必要があります。ファイル名に `.ftr` 拡張子が付いていない場合は、自動的にそれが付加されます。
5. オプション: **[上書き]** チェック ボックスをオフにします。
同じ名前のフォーカスファイルが指定されたディレクトリに既に存在する場合(手順 4 (29ページ))、**[上書き]** チェック ボックスをオフにしておけば、データフロー内でステージが実行される度に既存のフォーカスファイルが上書きされることを回避できます(ステージの実行は失敗します)。ただしこれは、**Write To Miner Dataset** ステージがデータフロー内で正常に実行されたときには、ファイルパスを再設定するか(手順 4 (29ページ))、既存のフォーカスファイルを手動で移動または削除しなければならないことを意味します。
6. オプション: 新しいフォーカス用のメタデータ ファイルを適用します ([メタデータの適用 \(30ページ\)](#) を参照)。
7. **[Write To Miner Dataset オプション]** ダイアログの **[フィールド]** タブをクリックして、フィールド テーブルを表示します。
フィールド テーブルには、**Write to Miner Dataset** ステージが Miner データセットに書き出すすべてのフィールドが列挙されます。
8. オプション: **[クイック追加]** ボタンをクリックします。
[クイック追加] ボタンをクリックすると、**[クイック追加]** ダイアログが開きます。このダイアログには、入力チャネルからのすべてのフィールドと、現在のフィールド リストが表示されます。各フィールドの横のチェックボックスをオン/オフして、フィールドリストにおいてそのフィールドを追加/削除します。
9. 必要に応じて、使用可能なオプションによってフィールドリストを変更します ([\[フィールド\] タブ \(30ページ\)](#) を参照)。
少なくとも 1 つのフィールドを指定しなければ、バリデーション エラーのインジケータが表示されます。
10. 最後に、**[OK]** ボタンをクリックして設定を適用します。
バリデーションエラーが検出された場合は、バリデーションエラーメッセージが表示されます。バリデーションエラーを修正してから、もう一度 **[OK]** ボタンをクリックします。バリデーションエラーが検出されなければ、**[Write to Miner Dataset オプション]** ダイアログが閉じます。

注: ステージを再設定するには、**[Write to Miner Dataset]** ステージをダブルクリックして、**[Write to Miner Dataset オプション]** ダイアログを再表示します。

メタデータの適用

.qsfm メタデータ ファイルには、派生フィールド定義、フィールド解釈、ビニング、レコード選択、フィールドとフォーカスのコメント、フォーカス履歴といった、フォーカス ファイルに適用可能な追加情報が含まれています。以下では、**Write to Miner Dataset** ステージを使用してこのような種類のメタデータを、生成されたフォーカス ファイルに適用する手順について説明します。

1. 有効な Miner フォーカス メタデータ ファイルを取得します (Spectrum Miner の既存のフォーカスからエクスポートするか、サードパーティから取得します)。
メタデータ ファイルは .qsfm という拡張子を持つ必要があります。
2. **[Write to Miner Dataset オプション]** ダイアログが表示中であることを確認します (**Miner データセットへの書き出し** (28ページ) を参照)。
3. **[Write to Miner Dataset オプション]** ダイアログで、**[メタデータを適用]** チェック ボックスをオンにします。
4. **[メタデータを適用]** チェック ボックスの下で、**[ファイル名]** フィールドの中の **[...]** ボタンをクリックして、**[ファイルを開く]** ダイアログを表示します。
[メタデータを適用] チェック ボックスがオンでない場合、メタデータ ファイルの選択は無効になります。
5. **[ファイルを開く]** ダイアログで、手順 **1** (30ページ) で取得した .qsfm ファイルを探して開きます。

注: **[メタデータを適用]** チェック ボックスがオンの場合は、メタデータ ファイルを選択しなければステージの設定が適用できません。

注: メタデータ ファイルを選択した後に **[メタデータを適用]** チェック ボックスをオフにすると、メタデータ ファイルの選択は無効となり、メタデータ ファイルはステージ実行に適用されません。ただし便宜上、**[Write to Miner Dataset オプション]** ダイアログを閉じるまでは、選択ファイルは削除されずにそのまま表示されます。

6. オプション:**[警告を無視]** チェック ボックスをオンにして、メタデータの一部が適用できなくてもデータフローが失敗しないようにします。

[フィールド] タブ

[フィールド] タブには、**Write to Miner Dataset** ステージが Miner データセットに書き出すすべてのフィールドが列挙されたテーブルがあります。以下の表に、テーブルの内容を変更するために使用可能な、フィールドに対するオプションを示します。

オプション名	説明
追加	<p>[追加] ボタンをクリックすると、フィールドリストに新しいフィールドを追加するための [追加] ダイアログが表示されます。[追加] ダイアログでは、新しいフィールドの名前、Miner タイプ、長さ (Miner タイプが 'string' の場合) を指定する必要があります。</p>
削除	<p>現在選択されているフィールドを、フィールド リストから削除します。</p>
再生成	<p>[再生成] ボタンをクリックすると、フィールド再生成の確認を求めるダイアログが表示されます。操作を確認すると、既存のすべてのフィールドが、選択されたフォーカスファイル内で見つかったフィールドに置き換えられます。</p> <p>注：フォーカス ファイルが選択されていない場合は、[再生成] ボタンは無効になります。</p> <p>注：無効なフォーカスからの再生成によって、既存のフィールドリストがクリアされたり変更されたりすることはありません。</p>
クイック追加	<p>[クイック追加] ボタンをクリックすると、[クイック追加] ダイアログが開きます。このダイアログには、入力チャネルからのすべてのフィールドと、現在のフィールド リストが表示されます。各フィールドの横のチェック ボックスをオン/オフして、フィールドリストにおいてそのフィールドを追加/削除します。</p> <p>注：フィールドの Miner フィールド名が無効である場合は、[クイック追加] ボタンは無効になります。</p>

以下のテーブル列の中で直接セルを変更することによって、フィールドのプロパティも更新することができます。

列名	説明
Miner フィールド名	<p>新しい Miner データセットに表示されるフィールドの名前。セルの中のテキストを編集して、所望のフィールド名を指定します。フィールド名は、以下の条件を満たす必要があります。</p> <ul style="list-style-type: none"> 一意である。 空ではない。 文字で始まる。 英数字またはアンダースコアのみを含む。 128 文字以下。 <p>フィールド名が無効な名前に変更された場合は、バリデーションエラーの記号がフィールド テーブルの横に表示されます。バリデーションエラーを修正してから、[OK] ボタンをクリックします。</p>
Miner フィールド タイプ	<p>新しい Miner データセットでフィールドが使用する Miner データ タイプ (出力 (32ページ) を参照)。ドロップダウンリストから選択することによって、フィールドのデータ タイプを変更します。</p>
文字列長	<p>テーブル行で選択された Miner フィールド タイプが 'string' の場合、ユーザは 1 ~ 4000 の整数値を指定することによって、Miner フォーカスにおけるこの文字列フィールドの長さを指定できます。フィールドの文字列値が指定された文字列長よりも長い場合は、結果のフォーカスにおいてこの長さまでに切り捨てられます。</p>

出力

Spectrum フィールド タイプと Miner フィールド タイプの間の以下のタイプ マッピングは自動的に行われます

表 2 : Field Mappings

Spectrum タイプ	Miner タイプ
integer	integer

Spectrum タイプ	Miner タイプ
string	string
double	real
datetime	date

その他の Spectrum タイプはすべて、Miner タイプ 'String' にマッピングされます。

3 - Machine Learning

モデル管理

このセクションの構成

Machine Learning モデル管理へのアクセス	35
モデル評価	36

Machine Learning モデル管理へのアクセス

Machine Learning モデル管理には、次の 3 つの方法でアクセスできます。

- Spectrum™ Technology Platform ようこそページを使用します。
 - Web ブラウザを起動し、次の Spectrum™ Technology Platform の Welcome ページを開きます。

サーバー名:ポート

例えば、Spectrum™ Technology Platform が "myspectrumplatform" という名前のコンピュータにインストールされており、デフォルトポート 8080 を使用している場合は、次のアドレスに移動します。

myspectrumplatform:8080

- **[Spectrum Machine Learning]** をクリックします。
- **[Machine Learning モデル管理を開く]** をクリックします。
- いずれかのモデル構築ステージから **[モデルの詳細についてはここをクリック]** をクリックします。
- Web ブラウザを使用して次の手順を実行します。

- Web ブラウザを起動し、以下の Spectrum™ Technology Platform の Machine Learning モデル管理ページを開きます。

サーバー名:ポート/machinelearning

例えば、Spectrum™ Technology Platform が "myspectrumplatform" という名前のコンピュータにインストールされており、デフォルトポート 8080 を使用している場合は、次のアドレスに移動します。

myspectrumplatform:8080/machinelearning

- 有効な Spectrum™ Technology Platform ユーザー名とパスワードを入力します。

モデル評価






モデル評価の概要

Machine Learning モデル管理の [モデル評価] タブには、Spectrum™ Technology Platform サーバー上にある機械学習モデルの全一覧が表示されます。テキスト ボックスに文字列を入力することによって、この一覧にフィルタを適用することができます。その文字列によって、テーブルのすべてのフィールドが検索されます。

これらのモデルに対して、複数の操作が実行できます。モデルのインポート、エクスポート、エクスポート、アンエクスポート、削除が可能です。エクスポートされたモデルは **Java Model Scoring** ステージで、機械学習モデルの適合を行った時に作成された式を使用して、新しいデータをスコアリングするために使用されます。また、各モデルの詳細情報を表示できます。詳細情報は、データを表示するモデルのタイプによって異なります。最後に、同じタイプの任意の 2 つのモデルを比較できます。比較を実行すると、比較する各モデルに対して [モデルの詳細] タブに表示されるのと同じ情報が、左右に並んで表示されます。

モデル評価の操作

モデルを選択して該当するボタンをクリックすることによって、以下の操作を実行します。

	モデル出力の詳細を表示します。K-Means Clustering ステージと Logistic Regression ステージからのこの情報は、[モデル出力] タブの [モデルの詳細についてはここをクリック] をクリックすることによっても参照できます。
	モデルを比較します。
	モデルを特定のパスからインポートします。必要に応じて、同じ名前の既存のモデルを上書きするかどうかを選択します。
	モデルを特定のパスにエクスポートします。必要に応じて、同じ名前の既存のモデルを上書きするかどうかを選択します。
	モデルをエクスポートして、Java Model Scoring ステージで使用できるようにします。エクスポートされていないモデルを、スコアリングに使用することはできません。



モデルをアンエクスポートします。



モデルを削除します。

注：エクスポートされているモデルを削除することはできません。ただし現時点では、他のユーザのモデルを削除できないようにするためのセキュリティ機能は装備されていません。

[モデルの詳細] タブ

[モデルの詳細] 画面には、すべてのモデルに関する以下の情報が表示されます。

モデル名	モデル名
モデル タイプ	機械学習モデルのタイプ
ユーザ	モデルを作成したユーザのユーザ名
説明	モデルの説明 (作成時に記述された場合)
ステータス	モデルがエクスポートされているかどうか
データフロー名	モデルを生成したデータフローの名前
作成時間	モデルが作成された日時

モデル タイプに応じて、その他の詳細情報が表示されます。

K-Means Clustering の詳細情報

[モデルの詳細] 画面には、K-Means Clustering モデルに関する以下の情報が表示されます。

モデル サマリ

以下の項目に対するトレーニング データを提供します。

- 行数
- クラスタ数
- カテゴリ列数
- 反復回数
- クラスタ内平方和
- 総平方和
- クラスタ間平方和

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 総クラスタ内平方和
- 総平方和
- クラスタ間平方和

セントロイド統計

各中心点 (セントロイド) に対する以下のトレーニング、テスト、N フォールド データを提供します。

- サイズ
- クラスタ内平方和

クラスタ平均

各中心点の詳細情報を提供します。内容は入力データによって異なります。クラスタとは、特定のクラスタリング アルゴリズムに基づいて類似と識別された、データ セットからのオブザベーションのグループです。

標準化されたクラスタ平均

各中心点の正規化情報を提供します。内容は入力データによって異なります。

Logistic Regression の詳細情報

[モデルの詳細] 画面には、Logistic Regression モデルに関する以下の情報が表示されます。

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 平均二乗誤差 (MSE)
- 平方根平均二乗誤差 (RMSE)
- オブザベーション数
- R の二乗 (R²)
- 対数損失 (Logloss)
- 曲線下面積 (AUC)
- 適合率-再現率曲線下面積 (PR AUC)
- ジニ係数
- クラスあたり平均誤差
- 赤池情報量基準 (AIC)
- ラムダ

- 残差逸脱度
- Null 逸脱度
- Null 自由度
- 残差自由度

最大メトリクスしきい値

以下のメトリクスを使用するトレーニング、テスト、N フォールド データに対する、トレーニング最大メトリクスしきい値を提供します。

- f1 最大値
- f2 最大値
- f0point5 最大値
- 最大正確度
- 最大適合率
- 最大再現率
- 最大特異度
- absolute_mcc 最大値
- min_per_class_accuracy 最大値
- mean_per_class_accuracy 最大値

混同行列

真 (true) の値が既知の一連のトレーニング、テスト、N フォールド データに対するモデルのパフォーマンスを表します。

標準化係数チャート

入力がどれだけ変化すると目標が変化するかを表す、係数の相対値を提供することによって、最も重要な予測因子を示します。

GLM 係数

指数分布に従う結果の回帰モデルを推定する、一般化線形モデル (GLM: Generalized Linear Model) の係数を示します。

AUC 曲線

曲線下面積 (AUC)。使用モデルの中で、トレーニング、テスト、N フォールド データを使用して最も正確にクラスを予測するものを判定します。

リフト/ゲイン曲線

トレーニング、テスト、N フォールド データを使用してバイナリ分類モデルの予測能力を評価します。

Logistic Regression の詳細情報

[モデルの詳細] 画面には、Linear Regression モデルに関する以下の情報が表示されます。

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 平均二乗誤差 (MSE)
- 平方根平均二乗誤差 (RMSE)
- オブザベーション数
- R の二乗 (R2)
- 平均残差逸脱度
- 平均絶対誤差 (MAE)
- 二乗対数平均平方根誤差 (RMSLE)
- 赤池情報量基準 (AIC)
- ラムダ
- 残差逸脱度
- Null 逸脱度
- Null 自由度
- 残差自由度

標準化係数チャート

特定の予測係数値の変化により目標値がどれだけ変化 (正または負の変化) するかを表す係数の相対値を提供することによって、最も重要な予測因子を示します。さらに、モデルの上位 25 の係数をグラフで示します。

GLM 係数

指数分布に従う結果の回帰モデルを推定する、一般化線形モデル (GLM: Generalized Linear Model) の係数を示します。

Random Forest Regression の詳細情報

[モデルの詳細] 画面には、Random Forest Regression モデルに関する以下の情報が表示されます。

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 平均二乗誤差 (MSE)
- 平方根平均二乗誤差 (RMSE)
- オブザベーション数
- R の二乗 (R²)
- 平均残差逸脱度
- 平均絶対誤差 (MAE)
- 二乗対数平均平方根誤差 (RMSLE)

変数重要度

以下のメトリクスを使用して、各変数の重要度の値を提供します。

- 相対的重要度
- 小数点以下桁数を含む重要度
- 比率

さらに、モデルの上位 25 の変数をグラフで示します。

Random Forest Classification の詳細情報 — 二項

[モデルの詳細] 画面には、Random Forest Classification の二項モデルに関する以下の情報が表示されます。

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 平均二乗誤差 (MSE)
- 平方根平均二乗誤差 (RMSE)
- オブザベーション数
- R の二乗 (R²)
- Logloss
- 曲線下面積 (AUC)
- 適合率-再現率曲線下面積 (PR AUC)
- ジニ
- クラスあたり平均誤差

最大メトリクスしきい値

以下のメトリクスを使用するトレーニング、テスト、N フォールド データに対する、トレーニング最大メトリクスしきい値を提供します。

- f1 最大値
- f2 最大値
- f0point5 最大値
- 最大正確度
- 最大適合率
- 最大再現率
- 最大特異度
- absolute_mcc 最大値
- min_per_class_accuracy 最大値
- mean_per_class_accuracy 最大値

混同行列

真 (true) の値が既知の一連のトレーニング、テスト、N フォールド データに対するモデルのパフォーマンスを表します。

変数重要度

以下のメトリクスを使用して、各変数の重要度の値を提供します。

- 相対的重要度
- 小数点以下桁数を含む重要度
- 比率

さらに、モデルの上位 25 の変数をグラフで示します。

AUC 曲線

曲線下面積 (AUC)。使用モデルの中で、トレーニング、テスト、N フォールド データを使用して最も正確にクラスを予測するものを判定します。

リフト/ゲイン曲線

トレーニング、テスト、N フォールド データを使用してバイナリ分類モデルの予測能力を評価します。

Random Forest Classification の詳細情報 — 多項

[モデルの詳細] 画面には、Random Forest Classification の多項モデルに関する以下の情報が表示されます。

メトリクス

以下の項目に対するトレーニング、テスト、N フォールド データを提供します。

- 平均二乗誤差 (MSE)
- 平方根平均二乗誤差 (RMSE)
- オブザベーション数
- R の二乗 (R2)
- Logloss
- クラスあたり平均誤差

混同行列

真 (true) の値が既知の一連のトレーニング、テスト、N フォールド データに対するモデルのパフォーマンスを表します。

変数重要度

以下のメトリクスを使用して、各変数の重要度の値を提供します。

- 相対的重要度
- 小数点以下桁数を含む重要度
- 比率

さらに、モデルの上位 25 の変数をグラフで示します。

主成分分析の詳細情報

[モデルの詳細] 画面には、主成分分析 (PCA) モデルに関する以下の情報が表示されます。

コンポーネントの重要度

主要コンポーネントを、以下のメトリクスに基づき重要度順に表示します。

- 標準偏差
- 寄与率
- 累積寄与率

回転

変数負荷量の行列をグラフで示します。コンポーネントのスコアを算出するには、正規化された元の各変数にこの重みを掛ける必要があります。

4 - Analytics Scoring リポジトリ

このセクションの構成

Analytics Scoring リポジトリの概要	45
モデルの表示	45
モデルの詳細の表示	46
Web ブラウザ クライアントによるモデルの追加	47
Spectrum Miner Spectrum Connector によるモデルの追加	49
モデルの削除	51

Analytics Scoring リポジトリの概要

Analytics Scoring リポジトリは、PMML Model Scoring ステージで使用する分析モデルが保存される場所です。リポジトリ内に保存されているモデルを管理するための Web ブラウザ クライアントが提供されています。

注：サポートされるモデルタイプとタイプマッピングの詳細については、[サポートされるモデルフォーマット](#)（14ページ）を参照してください。

Analytics Scoring リポジトリ Web ブラウザ クライアントへのアクセス

Web ブラウザに URL `http://servername:port/analytics scoring` (`http://localhost:8080/analytics scoring` など)を入力して、**【ログイン】**ページを表示します。表示されたブラウザ ダイアログ ボックスに、有効な Spectrum™ Technology Platform ユーザ名とパスワードを入力します。リポジトリに保持されている既存の Analytics モデルの一覧を示すページが、ブラウザに表示されます。

モデルの表示

以下では、Analytics Scoring リポジトリに現在含まれているモデルのリストを表示する手順について説明します。

1. 有効な Spectrum™ Technology Platform ユーザ名とパスワードを **【ログイン】**ページに入力することによって、Analytics Scoring リポジトリ Web ブラウザ クライアントにログインします。
【ログイン】ページからログインする方法の詳細については、[Analytics Scoring リポジトリの概要](#)（45ページ）を参照してください。
ログインに成功すると、**【リポジトリ】**ページが開き、PMML Model Scoring ステージで使用可能な Analytics Scoring リポジトリ内のモデルの一覧テーブルが表示されます。
2. **【1 ページあたりのモデル数】**では、テーブル 1 ページあたり 10 個、20 個、または 50 個のモデルを表示するように選択できます。
3. ページ間の移動には、**【1 ページあたりのモデル数】**ドロップダウンの横にあるページ コントロールを使用します。

ページコントロールでページを変更できるのは、ステップ 2 (45ページ) で1ページあたりのモデル数として選択した値が Analytics Scoring リポジトリ内のモデル総数よりも小さい場合のみです。

4. **[名前]** 列ヘッダをクリックすると、表示されているモデルの順序が **[名前]** の値の昇順/降順で切り替えられます。同様に、**[タイプ]** 列ヘッダをクリックすると、表示されているモデルの順序が **[タイプ]** の値の昇順/降順で切り替えられます。
いずれかの列ヘッダをクリックするまでは、モデルはデフォルトで **[名前]** の値の昇順で表示されます。

モデルの詳細の表示

以下では、Analytics Scoring リポジトリ内のモデルに関する詳細情報を表示して、その説明や、モデル入出力の詳細を確認する手順について説明します。

1. 対象のモデルを **[リポジトリ]** ページで探します。
Analytics Scoring リポジトリを参照する方法の詳細については、[モデルの表示 \(45ページ\)](#) を参照してください。
2. テーブルの **[名前]** 列でモデルの名前をクリックするか、モデルのチェック ボックスをオンにしてテーブルの上にある **[モデルの詳細]** ボタンをクリックすることにより、そのモデルの詳細ページに移動します。
詳細ページには、そのモデルの詳細情報が表示されます。

名前	モデル名
タイプ	モデルのタイプ (サポートされるモデル フォーマット (14ページ) を参照)
説明	モデルの目的に関する簡単な説明
入力	モデルの全必須フィールドに関する情報が記載されたテーブル各行に1つのフィールドの名前とタイプが記載されています (サポートされるモデル フォーマット (14ページ) を参照)。
出力	モデルが生成する全出力フィールドに関する情報が記載されたテーブル各行に1つのフィールドの名前とタイプが記載されています (サポートされるモデル フォーマット (14ページ) を参照)。
3. **[リポジトリ]** ページに戻るには、右上隅にある **[閉じる]** ボタンをクリックします。

Web ブラウザ クライアントによるモデルの追加

以下では、Analytics Scoring リポジトリの Web ブラウザ クライアントによって新しいモデルを Analytics Scoring リポジトリに追加し、そのモデルを Spectrum™ Technology Platform データフローに展開できるようにする手順について説明します。

1. **モデルの表示** (45ページ) の手順に従って、**[リポジトリ]** ページに移動します。
2. (モデル テーブルのすぐ上にある) **[モデルを公開]** ボタンをクリックして、**[新規モデルの追加]** ページに移動します。

モデルテーブルの現在のページでいずれかのモデルが選択されている場合は、**[モデルを公開]** ボタンは無効になります。これらのモデルを非選択にして、**[モデルを公開]** ボタンを有効にします。

3. **[新規モデルの追加]** ページには、**[名前]**、**[説明]**、**[モデルのアップロード]** という 3 つの入力フィールドがあります。**[名前]** というラベルが付いたテキスト ボックスに新しいモデルの名前を入力します。

モデル名は一意でなければなりません。モデル名が、Analytics Scoring リポジトリに保存されている、削除されていない別のモデルによって既に使用されている場合は、バリデーション エラーとともにテキスト ボックスがハイライト表示されます。

注：モデル名の先頭または末尾にある空白は必ず削除されます。つまり、入力した名前が既に使用されている場合に、先頭または末尾に空白を加えてもエラーは解決されません。

注：モデル名は大文字と小文字が区別されます。つまり、大文字と小文字の違いがあれば、複数のモデルで同じ名前を共有できます。例えば、"my Model"、"my model"、"mY ModEl" という名前のモデルは、リポジトリ内に同時に共存できます。

注：名前のバリデーションによって Analytics Scoring リポジトリのコンテンツに対する競合が検出されるのは、ページを開いたときだけです。つまり、**[新規モデルの追加]** ページに移動した後にリポジトリが変更されても (他のユーザによってモデルが追加または削除されたなど)、その変更はページバリデーションの対象にはなりません。ただし、**[公開]** ボタンをクリックした後、最新版の Analytics Scoring リポジトリに対して新しいモデルのバリデーションが再度行われます。

4. **[説明]** ボックスに、新しいモデルの説明を記述します。

このフィールドは必須です。モデルには説明を入力する必要があります。

5. **[ファイルを選択]** ボタンをクリックし、**QMML** (14ページ) または **PMML** (16ページ) のいずれかのファイル形式で保存されているモデルを選択することによって、モデル定義を追加します。

注：選択するモデルのサイズは、50MB 未満でなければなりません。

注：ファイル選択をサポートしていない古いブラウザを使用している場合は、**[モデルのアップロード]** というラベルが付いたテキスト エリアに XML モデル定義を入力する必要があります。

注：選択されたモデルが **QMML** (14ページ) モデルの場合、モデルを正しく公開するには、**QMML** (14ページ) に少なくとも1つのコンパイル済みのライブルールセットが必要です。コンパイル済みのライブルールセットが複数存在する場合は、**QMML** (14ページ) 内で見つかったそのうちの最初のルールセットがモデル定義に使用されます。

注：**QMML** (14ページ) 内のルールセットのコンパイルや、ルールセットのライブステータスの変更を行う(あるいは、ルールセットを実際に追加または削除する)場合は、インストールされている Spectrum Miner に含まれる Spectrum Miner qsqmmledit コマンドラインユーティリティが使用できます(コマンドライン ウィンドウで "qsqmmledit.exe -help" と入力することによって、このユーティリティのヘルプを表示できます)。

注：**[公開]** ボタンをクリックするまで、モデル定義のバリデーションは行われません。

6. **[公開]** ボタンをクリックして、Analytics Scoring リポジトリへのモデルのアップロードを開始します。

[公開] ボタンは、新しいモデルに対する有効な名前、説明、モデル定義が指定されるまでは無効になっています。

7. **[公開]** ボタンにビジー アイコンが表示されている間、処理の完了を待ちます。モデルの公開中は、新しいモデルの名前、説明、またはモデル定義を変更することはできません。Analytics Scoring リポジトリにモデルが正しく公開されると、新しいモデルの **[詳細]** ページが表示されます。

モデルの公開中は、**[キャンセル]** ボタンも無効になります。

注：モデルの公開中に、XML モデル定義が無効であるとか、モデル名が既に使用されているといった問題が見つかった場合、モデルは公開されず、問題の詳細を示す警告エラーが表示されます。**[OK]** をクリックして警告エラーを閉じ、問題を修正してからモデルの公開を再度試みます。

Spectrum Miner Spectrum Connector によるモデルの追加

Spectrum Connector 統合コンポーネントを使用すると、**QMML** (14ページ) モデルを Spectrum Miner から直接 Spectrum Analytics Scoring リポジトリに公開できます。

Spectrum Connector 統合コンポーネントは、'デスクトップ' インストールを含む、インストール済みの任意の既存の Spectrum Miner バージョン 7.0B (またはそれ以降) に追加する形でインストール可能です。

以下に示す 3 つの異なる方法で、Spectrum Connector 統合コンポーネントを使用して Analytics Scoring リポジトリにモデルを公開することができます。

- [Spectrum Miner からの QMML モデルの公開](#) (49ページ)
- [Decision Studio Model Builder ツールからの QMML モデルの公開](#) (50ページ)
- [Spectrum Miner コマンド ライン ツールからの QMML モデルの公開](#) (50ページ)

Spectrum Miner からの QMML モデルの公開

以下では、Spectrum Miner を使用して Analytics Scoring リポジトリに新しいモデルを追加する手順を示します。

1. Spectrum Miner を起動し、**[ファイルブラウザ]** タブで、公開する .qmm1 ファイルを含むディレクトリに移動します。
2. .qmm1 ファイルを右クリックして、**[公開] > [Spectrum Analytics Scoring リポジトリへ]** を選択します。
3. 入力を求められたら、Spectrum プラットフォームにサインインするための有効な Spectrum ユーザー名とパスワードを入力します。
サインイン プロンプトは、Spectrum Miner セッションごとに一回表示され、その後のモデル公開には入力済みの資格情報が使用されます。
4. **[モデルの詳細]** タブで、Spectrum Analytics Scoring リポジトリに公開するモデルのモデル名、モデルの説明、**QMML** (14ページ) ルールセットを指定します。
モデルに付ける名前は、Spectrum Advanced Analytics モデル リポジトリの中で一意である必要があります。
5. **[Go]** ボタンをクリックします。

Decision Studio Model Builder ツールからの QMML モデルの公開

以下では、Decision House モデル構築ツールの 1 つから直接 Analytics Scoring リポジトリに新しいモデルを追加する手順を示します。

1. 選択した Decision House Model Builder ツールで、モデルを構築します。
2. **【モデルの種類】を Spectrum に公開** ボタンを押します。このボタンは、1 つ前のステップでモデルの構築を終えると有効になります。
3. 入力を求められたら、Spectrum プラットフォームにサインインするための有効な Spectrum ユーザ名とパスワードを入力します。サインインプロンプトは、Spectrum Miner セッションごとに一回表示され、その後のモデル公開には入力済みの資格情報が使用されます。
4. **【モデルの詳細】** タブで、Spectrum Analytics Scoring リポジトリに公開するモデルのモデル名、モデルの説明、出力フィールドを指定します。

モデルに付ける名前は、Spectrum Advanced Analytics モデル リポジトリの中で一意である必要があります。

注：対応するチェック ボックスがオンになっているフィールドのみが、モデルの出力フィールドとして含まれます。少なくとも 1 つのフィールドを選択する必要があります。

5. **【公開】** ボタンをクリックします。

Spectrum Miner コマンド ライン ツールからの QMML モデルの公開

以下では、Spectrum Miner の Spectrum Connector 統合コンポーネントで提供されているデータ構築コマンドラインツール `spectrumpublish` を使用して Analytics Scoring リポジトリに新しいモデルを追加する手順を示します。

コマンド プロンプトで、`spectrumpublish` ユーティリティを実行します。

```
例: $QSHOME/server/qs7.0B/win32/bin/spectrumpublish -input qmml file
path -name model name -description model description -ruleset ruleset
name -username Spectrum username -password password
```

注：-input、-name、-username、-password、および -description 引数は必須です。-ruleset 引数は省略可能です。

注：指定された **QMML** (14 ページ) ファイルに複数のルールセットが含まれているが -ruleset 引数は指定されていないという場合、**QMML** (14 ページ) ドキュメント内の最初のルールセットのみが Spectrum Analytics Scoring リポジトリに公開されます。

モデルの削除

以下では、Analytics Scoring リポジトリからモデルを削除する手順について説明します。リポジトリからモデルを削除すれば、そのモデルが新しい Spectrum™ Technology Platform データフローに展開されるのを防ぐことができます。モデルを削除しても、そのモデルを使用するように設定されている既存のデータフローの実行時動作に影響はありませんが、そのモデルを使用するように設定されているデータフロー内の PMML Model Scoring ステージの設定に変更を加えることはできません。

1. **モデルの表示** (45ページ) の手順に従って、**[リポジトリ]** ページに移動します。

[リポジトリ] ページでは、Analytics Scoring リポジトリ内に保存されたモデルが、モデルテーブル上に表示されます。Analytics Scoring リポジトリに多数のモデルが保存されている場合、モデルテーブルは複数のページにわたる可能性があります。

2. モデルテーブルを使用して、Analytics Scoring リポジトリから削除するモデルを選択します。

注：モデルテーブルの現在のページで選択されているモデルだけが削除されます。例えば、モデルテーブルが 2 ページにわたっており、両方のページでモデルが選択されている場合に **[削除]** ボタンを押すと、現在のページ上のモデルだけが削除され、もう一方のページで選択されているモデルは無視されます。

3. モデルテーブルのすぐ上にある **[モデルを削除]** ボタンをクリックします。

モデルテーブルの現在のページでモデルが選択されていない場合は、**[モデルを削除]** ボタンは無効になります。

選択されている 1 つまたは複数のモデルを Analytics Scoring リポジトリから削除することを確認するポップアップメッセージが表示されます。

4. **[はい]** ボタンをクリックして、選択されたモデルを Analytics Scoring リポジトリから削除します。

選択されたモデルが Analytics Scoring リポジトリから正しく削除されると、**[リポジトリ]** ページのモデルテーブルが更新されて再表示されます。

著作権に関する通知

© 2019 Pitney Bowes. All rights reserved. MapInfo および Group 1 Software は Pitney Bowes Software Inc. の商標です。その他のマークおよび商標はすべて、それぞれの所有者の資産です。

USPS® 情報

Pitney Bowes Inc. は、ZIP + 4® データベースを光学および磁気媒体に発行および販売する非独占的ライセンスを所有しています。CASS、CASS 認定、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、および ZIP + 4 の各商標は United States Postal Service が所有します。United States Postal Service に帰属する商標はこれに限りません。

Pitney Bowes Inc. は、NCOA^{Link}® 処理に対する USPS® の非独占的ライセンスを所有しています。

Pitney Bowes Software の製品、オプション、およびサービスの価格は、USPS® または米国政府によって規定、制御、または承認されるものではありません。RDI™ データを利用して郵便送料を判定する場合に、使用する郵便配送業者の選定に関するビジネス上の意思決定が USPS® または米国政府によって行われることはありません。

データ プロバイダおよび関連情報

このメディアに含まれて、Pitney Bowes Software アプリケーション内で使用されるデータ製品は、各種商標によって、および次の 1 つ以上の著作権によって保護されています。

© Copyright United States Postal Service. All rights reserved.

© 2014 TomTom. All rights reserved. TomTom および TomTom ロゴは TomTom N.V の登録商標です。

© 2016 HERE

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

電子データに基づいています。© National Land Survey Sweden.

© Copyright United States Census Bureau

© Copyright Nova Marketing Group, Inc.

このプログラムの一部は著作権で保護されています。© Copyright 1993-2007 by Nova Marketing Group Inc. All Rights Reserved

© Copyright Second Decimal, LLC

© Copyright Canada Post Corporation

この CD-ROM には、Canada Post Corporation が著作権を所有している編集物からのデータが収録されています。

© 2007 Claritas, Inc.

Geocode Address World データ セットには、
<http://creativecommons.org/licenses/by/3.0/legalcode> に存在するクリエイティブ コモンズ アトリビューション ライセンス (「アトリビューション ライセンス」) の下に提供されている GeoNames Project (www.geonames.org) からライセンス供与されたデータが含まれています。お客様による GeoNames データ (Spectrum™ Technology Platform ユーザ マニュアルに記載) の使用は、アトリビューションライセンスの条件に従う必要があります。お客様と Pitney Bowes Software, Inc. との契約と、アトリビューション ライセンスの間に矛盾が生じる場合は、アトリビューションライセンスのみに基づいてそれを解決する必要があります。お客様による GeoNames データの使用に関しては、アトリビューション ライセンスが適用されるためです。



3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com