### 機械学習学習データセット

#### 概要

機械学習学習データセット（MLデータセット）とは、機械学習モデルの学習に使用されるデータの集合です。機械学習モデルは、与えられたデータセットからパターンを学習し、類似データに対して値やパターンを予測します。そのため、学習データセットの品質はモデルの成果に大きな影響を与えます。

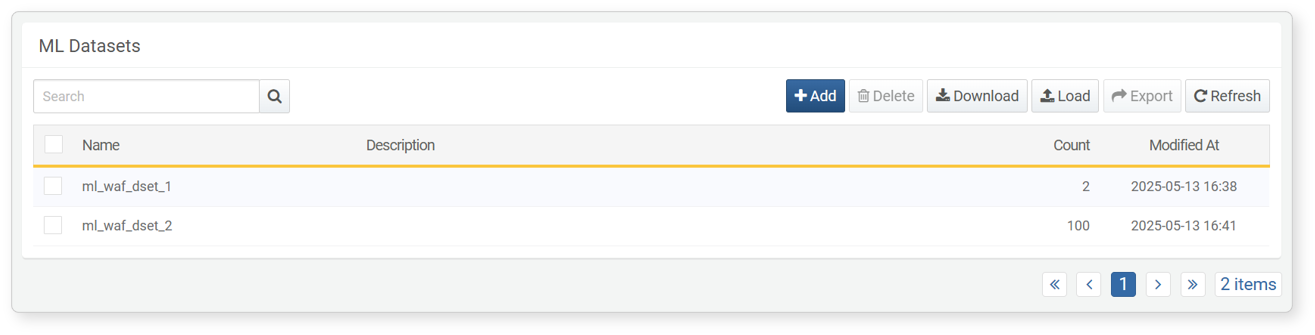
留意事項

学習データセットは、機械学習モデルを適用する環境で生成されるデータとできるだけ近いものである必要があります。データセット作成時には以下の点に注意してください。

* 多様な情報の含有：時間、場所、ユーザー情報、イベントタイプ、成否など、さまざまな特性を網羅したデータを収集・準備します。実際のデータを使用することが推奨されます。
* 教師あり学習の目的変数選定：ランダムフォレストなどの教師あり学習モデルでは、予測対象となる目的変数が必要です。学習データセット作成時には、必ず目的変数（予測対象値）を含めてください。
* データ前処理：データの誤りや極端値を除去または調整します。特徴量の正規化やスケーリングを行い、大きなばらつきを防ぎます。
* 文字列のベクトル化：機械学習モデルは生のテキストを処理できないため、[tfidf](https://docs.logpresso.com/en/query/tfidf-command)クエリコマンドなどを用いてテキストデータを数値化します。
* データの質と量：データが正確で、完全かつ一貫性があることを確認します。多様なパターンを学習するためには十分なデータ量が必要であり、特に異常イベントの事例を十分に含めることが重要です。
* 過学習の防止：クロスバリデーションや正則化などの手法を適用し、学習データへの過度な適合を防ぎます。
* パターンの変化：セキュリティ脅威のパターンは時間とともに変化するため、機械学習モデルを定期的に再学習してください。

#### MLデータセットの検索

**ポリシー > MLデータセット** で学習データセットの一覧を閲覧・検索できます。



* **名前**：学習データセット名
* **説明**：学習データセットに関する追加情報
* **件数**：学習データセット内のデータ件数
* **最終更新日時**：学習データセットの作成または最終更新日時

特定のMLデータセットを探す場合は、ツールバーの検索ツールを利用してください。検索ツールは **名前** および **説明** フィールドのキーワードでデータセットを絞り込みます。大文字・小文字は区別されません。

MLデータセット一覧のダウンロード

学習データセット一覧をファイルとしてローカルPCにダウンロードするには、ツールバーの **ダウンロード** をクリックし、希望するファイル形式を選択してください。

MLデータセット一覧の更新

学習データセット一覧を最新情報に更新するには、ツールバーの **更新** をクリックしてください。

MLデータセットのインポート・エクスポート

学習データセットはファイルとしてエクスポート・インポートでき、バックアップや復元に利用できます。

学習データセットをエクスポートするには：

一覧からエクスポートしたいデータセットのチェックボックスを選択します。

ツールバーの **エクスポート** をクリックします。

**学習データセットのエクスポート** ダイアログで名前を設定し、**OK** をクリックします。

学習データセットをインポートするには：

ツールバーの **インポート** をクリックします。

**学習データセットのインポート** ダイアログで **ファイル選択** をクリックし、事前にエクスポートしたデータセットファイルを選択します。

ファイル選択後、**OK** をクリックします。

#### MLデータセットの追加

学習データセットを追加するには：

**ポリシー > MLデータセット** に移動し、ツールバーの **追加** をクリックします。

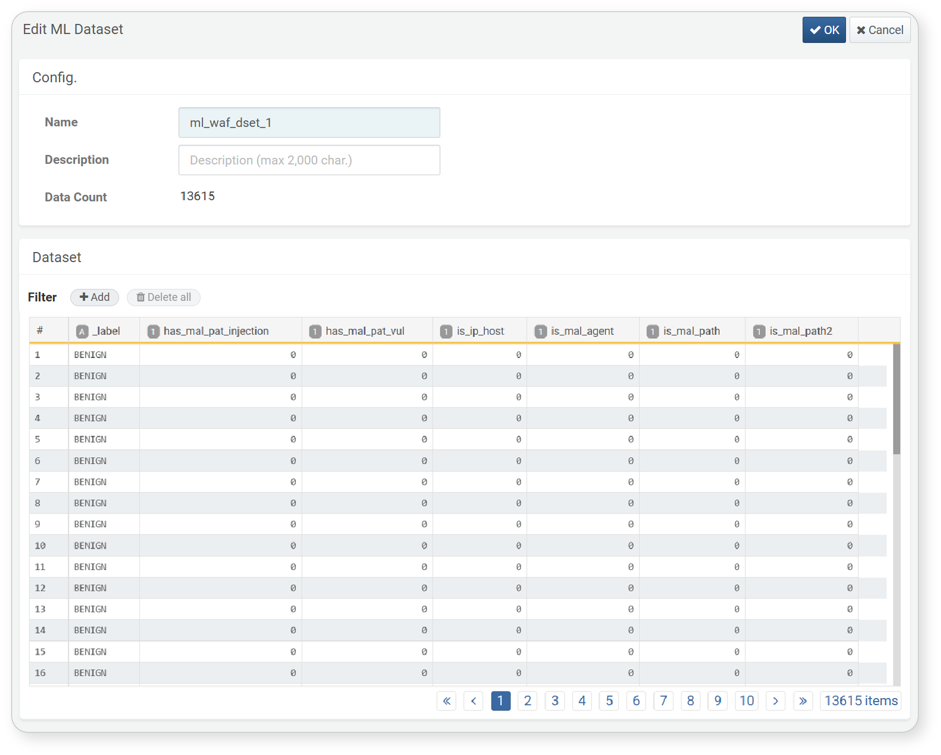
**MLデータセット追加** 画面で必要事項を入力し、**OK** をクリックします。

* **名前**：学習データセット名（最大50文字）
* **説明**：データセットの詳細説明（最大2,000文字）
* **クエリ文**：学習データセットを生成するためのクエリ（最大10,000文字）。機械学習モデルの学習時には、データセットのフィールド値が使用されます。必要に応じてフィールド名を変更できます。

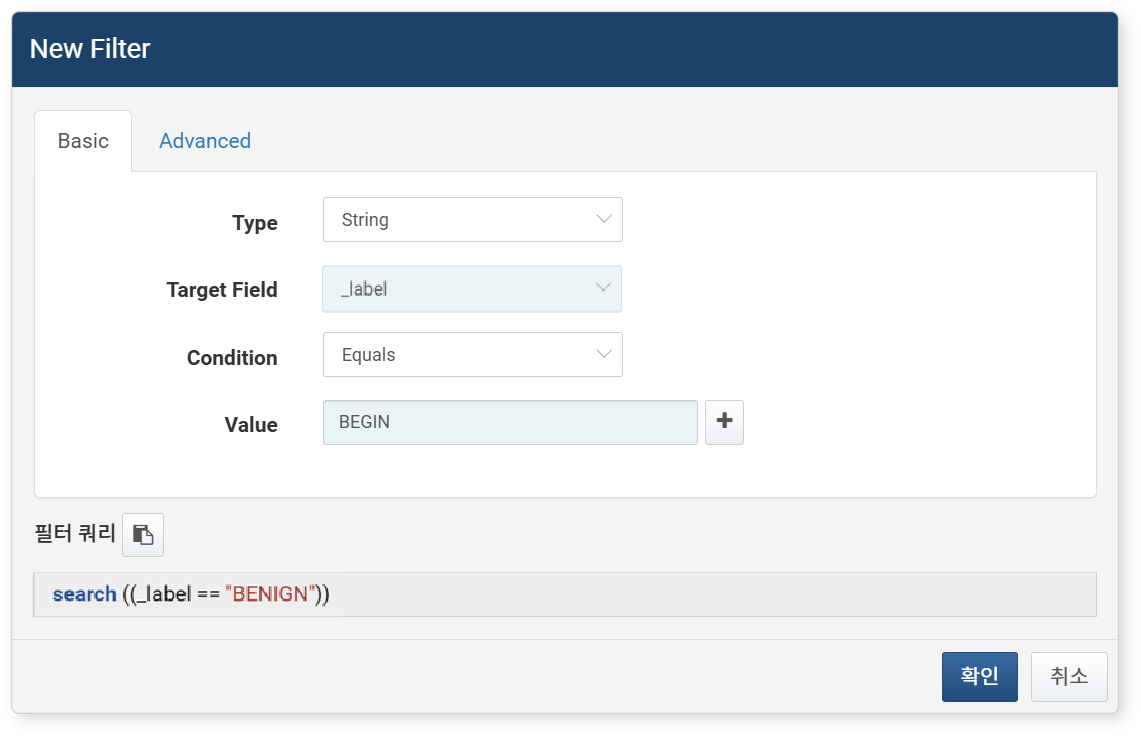
データセット名は重複できません。また、不正なクエリ文や存在しないデータテーブルを指定した場合、データセットを追加できません。

#### MLデータセットの閲覧

一覧の学習データセット名をクリックすると、そのデータを閲覧できます。



表示データを絞り込むためにフィルターを適用できます。フィルターを追加するには、**フィルター** の横の **追加** をクリックしてください。



#### MLデータセットの編集

学習データセットを編集するには：

一覧の学習データセット名をクリックします。

**MLデータセット編集** 画面で情報を更新し、**OK** をクリックします。編集可能な項目は **説明** のみです。

#### MLデータセットの削除

学習データセットを削除するには：

一覧から削除したいデータセットのチェックボックスを選択します。

ツールバーの **削除** をクリックします。

**MLデータセット削除** ダイアログで選択したデータセットを確認し、**削除** をクリックします。キャンセルする場合は **キャンセル** をクリックしてください。

機械学習モデルに紐づいている学習データセットは削除できません。削除するには、該当する機械学習モデルからデータセットを先に解除してください。