

Vertica バージョン 8.1 の新機能: 機械学習

原文は[こちら](#)

Vertica 7.2.2 で追加された予測分析のための機械学習の機能が、Vertica 8.1 において、いくつかの重要な機能を追加して、全体的により使いやすくなっていることに気がつくでしょう。正規化のパラメーターを保存して再適用するモデル管理のアクセス制御機能の向上、値の欠落の補完、SVM(サポートベクターマシン)のサポートがあります。

拡張機能の概要

新機能	説明
IMPUTE 関数	平均値またはモードのいずれかを使用して、データ内のパーティションごとに欠損値を補完可能に
SVM 分類	トレーニングデータに基づいて、あるカテゴリまたは他のカテゴリにデータを割り当てる分類アルゴリズムを追加
モデルのセキュリティ	機械学習のモデルに対するセキュリティ設定が可能に
新しい正規化の関数	新しい正規化の関数、NORMALIZE_FIT、APPLY_NORMALIZE、RESERSE_NORMALIZE を追加し、正規化関連のパラメーターを保持可能に
新しい構文	機械学習の関数で USING PARAMETERS シンタックスを使用可能に
ALTER MODEL と DROP MODEL のサポート	これまでの RENAME_MODEL を ALTER MODEL に、DELETE_MODEL を DROP_MODEL に置き換えより直感的に操作可能に
APPROXIMATE_MEDIAN と APPROXIMATE_PERCENTILE のサポート	新しい関数を使用して、大きなデータセットの推定値を高速取得可能に

モデル管理

Vertica 8.1 では、他のデータベースオブジェクトと同様の機械学習モデルのセキュリティがサポートされています。GRANT および REVOKE 文を使用すると、モデルに対する権限を他のユーザーに付与および取り消すことができます。

たとえば、作成したモデルを同僚と共有したい場合は、そのモデルにそのパーミッションを与えることができます。詳細情報と実行例については、Vertica ドキュメントの [Managing Model Security](#) を参照してください。

新しい正規化関数

Vertica 8.1 では、新しい正規化関数を使用して、正規化パラメーターを保存して後で適用することができます。データを正規化する目的は、異なる列の数値データを等価なスケールにスケールすることです。

Vertica では 3 つの正規化の方法を提供しています。

- MinMax: 0 と 1 の分布内になるように値を正規化する。
- Zスコア: 観測値が平均からの標準偏差の数になるように値を正規化する。
- ロバスト Zスコア: この方法を使用して、外れ値が Zスコア計算に与える影響を軽減することができます。この方法では、Zスコア法で使用される平均値ではなく中央値を使用します。

以前は、NORMALIZE 関数を使用して、入力テーブルの列に直接正規化を適用することができました。新しい関数を使用すると、以下のことを行うことができます。

- NORMALIZE FIT を使用して、後で操作するための正規化パラメーターをモデルに格納します。
- APPLY NORMALIZE を使用して、モデルに保存された正規化パラメーターを使用して、入力テーブル内の特定の列のセットに正規化を適用します。
- REVERSE_NORMALIZE を使用して、正規化データセットの正規化変換を取り消します。

新しい IMPUTE 関数

IMPUTATION とは、欠損データを置換値で置き換える処理のことをさします。Vertica 8.1 では、新しい IMPUTE 関数を使用して、平均値またはモードのいずれかの方法を指定して、欠損値を置き換えることができます。

```
=> SELECT IMPUTE ('output_view', 'customers', 'name, id', 'mean');  
=> SELECT IMPUTE ('output_view', 'customers', 'name, id', 'mode');
```

SVM(サポートベクターマシン)

SVM は、画像分類またはテキスト分類でよく使用される分類アルゴリズムです。トレーニングと予測には SVM_CLASSIFIER と PREDICT_SVM_CLASSIFIER を使用できます。

Vertica で SVM アルゴリズムを使用する方法の例については、Vertica ドキュメントの [Classifying Data Using SVM](#) を参照してください。

詳細情報

Vertica の機械学習に関する全情報につきましては、「SQL Reference manual」の [Machine Learning Functions](#) と「Analyzing Data guide」の [Machine Learning for Predictive Analytics](#) を参照してください。Vertica は今後も機械学習機能を継続して拡張しています。今後のリリースでは、更なる拡張機能をご期待ください。