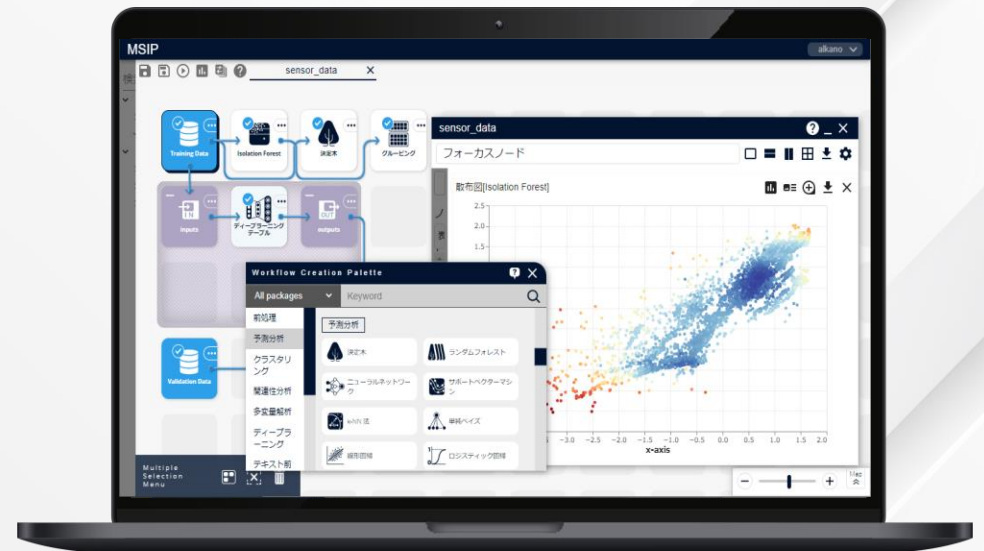


テクニカルサンプルプロジェクト 時系列データ予測分析



株式会社 NTTデータ数理システム

このプロジェクトについて

こんな方におすすめします

時系列データを用いて予測分析をしたい方

製品などの需要予測をしたい方

何をするプロジェクト？

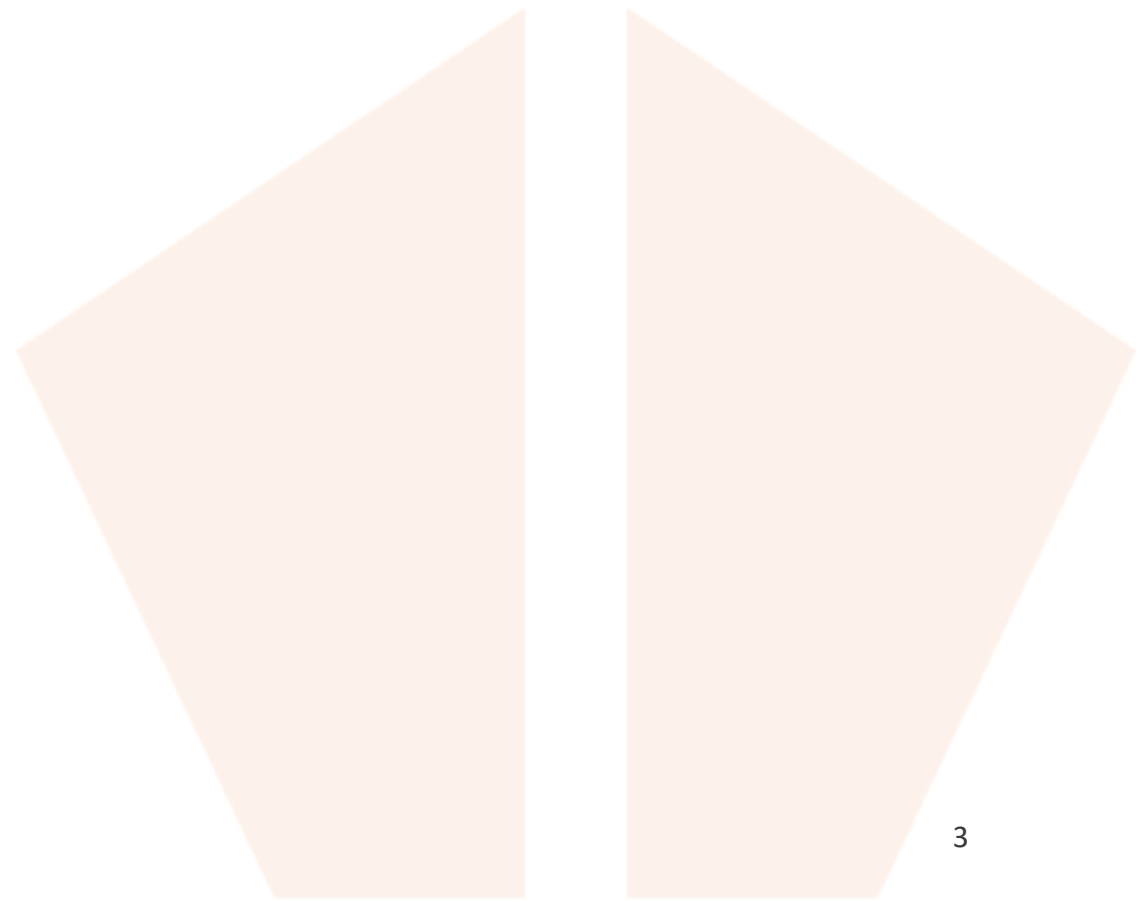
需要予測などの時系列データを利用した予測分析においては、次のような課題がよく発生します。

- ・ **需要予測の対象が膨大で、人手で予測を行うには限界がある。**
- ・ **ユーザーのニーズの頻繁な変化**に追従する必要がある。
- ・ 過去の経験や直感、不確実なトレンドの変化など数値化しづらい面を考慮して予測するため、**属人性が高い。**

そこで、本プロジェクトでは「機械的に」「時系列データに合わせた方法で」**商品の販売履歴データから翌週の商品の販売件数（需要）を予測するモデルを学習し、需要予測を行います。**



プロジェクト 解説



プロジェクト 解説

1. 対象データ

1.

POS データに顧客コードが付与された **ID-POS 形式のデータ** です。
 各行に顧客コードに紐づいた商品の購買履歴が時系列順に記録されています。



可視化画面: [ID-POS.dft] ノード

No.	商品コード CATEGORY	顧客コード CATEGORY	購入日 DATETIME	価格 INT	件数 INT	商品カテゴリコード CATEGORY	発売月 DATETIME
1	40000001	008304	2021/09/10 00:00:00	5,250	1	400	2013/11/28 00:00:00
2	40000001	007113	2021/08/13 00:00:00	5,250	1	400	2013/11/28 00:00:00
3	40000001	009612	2021/07/03 00:00:00	5,250	1	400	2013/11/28 00:00:00
4	40000001	002083	2021/06/21 00:00:00	5,250	1	400	2013/11/28 00:00:00
5	40000001	007859	2021/05/26 00:00:00	5,250	1	400	2013/11/28 00:00:00
6	40000001	007372	2021/03/16 00:00:00	5,250	1	400	2013/11/28 00:00:00
7	40000001	009226	2021/02/26 00:00:00	5,250	1	400	2013/11/28 00:00:00
8	40000001	002410	2021/02/05 00:00:00	5,250	1	400	2013/11/28 00:00:00
9	40000001	002746	2020/12/07 00:00:00	5,250	1	400	2013/11/28 00:00:00
10	20000002	006270	2021/05/02 00:00:00	1,995	1	200	2013/11/28 00:00:00

プロジェクト 解説

2. 日時情報の操作

時系列リサンプリングアイコンを用いて、時系列データを再標本し、**等間隔なデータに変換**します。

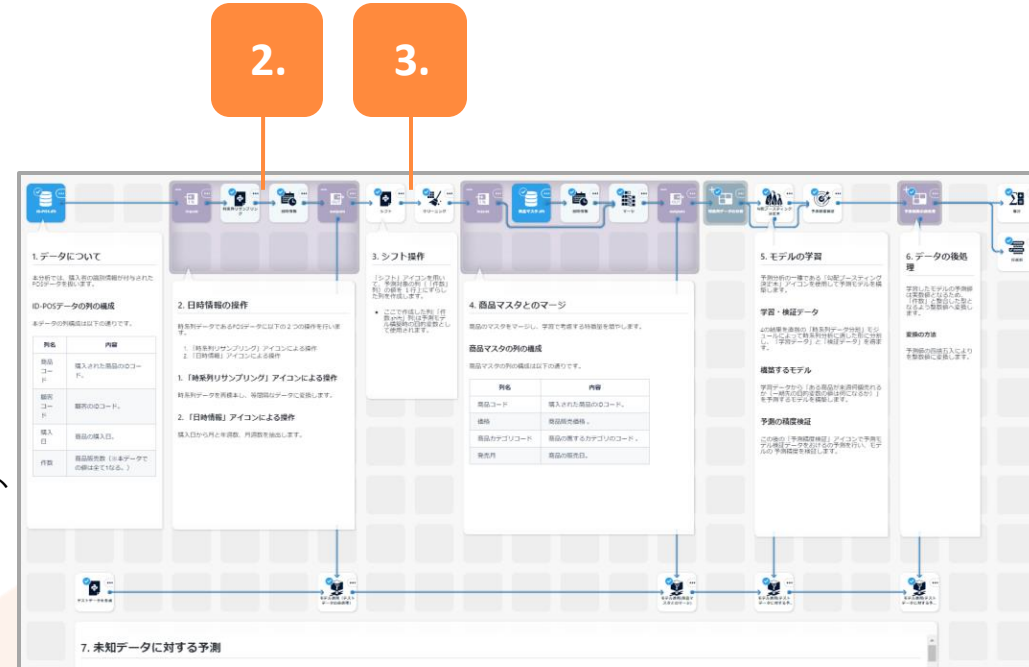
(時系列分析では、多くの場合、各データが等間隔であることを前提としています。)

また、日時情報アイコンを用いて、購入日から月と年週数、月週数を抽出します。

3. シフト操作

シフトアイコンを用いて、**対象となる計算対象列を1つ上にずらします**。

これにより各行に**1期先のデータ**（翌週の販売件数）が格納され、学習の目的値として利用できるようになります。



プロジェクト 解説

4. 商品マスタとのマージ

学習に用いる特徴量を増やします。

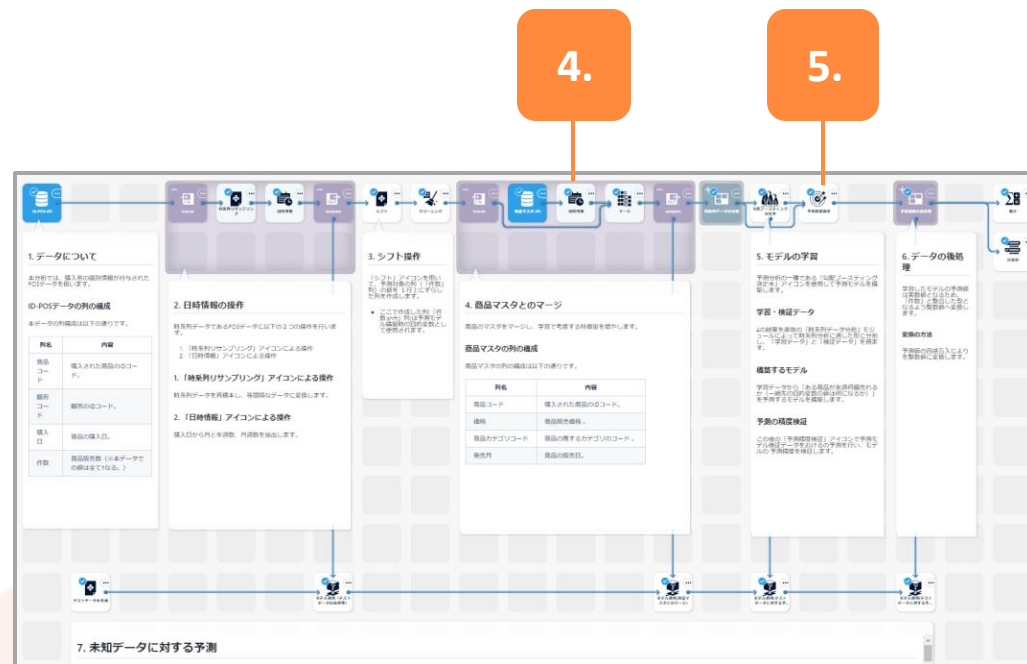
具体的には、商品のマスターをマージ、続いて発売日から年や月の情報も抽出して特徴量とします。

5. モデルの学習

予測分析の一種である「勾配ブースティング決定木」アイコンにより、予測モデルを構築します。

今回は「**ある商品は（1期先の）来週、何個売れるか**」を予測するモデルとなります。

なお、モデルの学習に用いる訓練データおよびモデルの精度検証に用いるデータは、直前の「時系列データ分割」モジュールによって得られます。



プロジェクト 解説

6. データの後処理

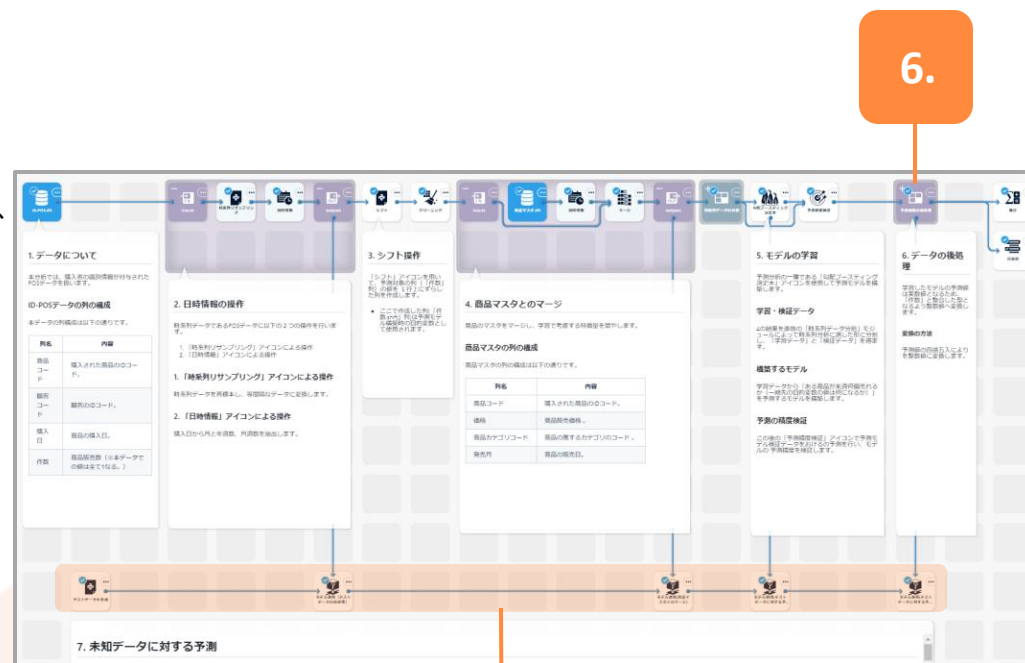
学習したモデルは、予測値が実数値で出力されているため、四捨五入による整数値への変換を実行します。

7. 未知データに対する予測

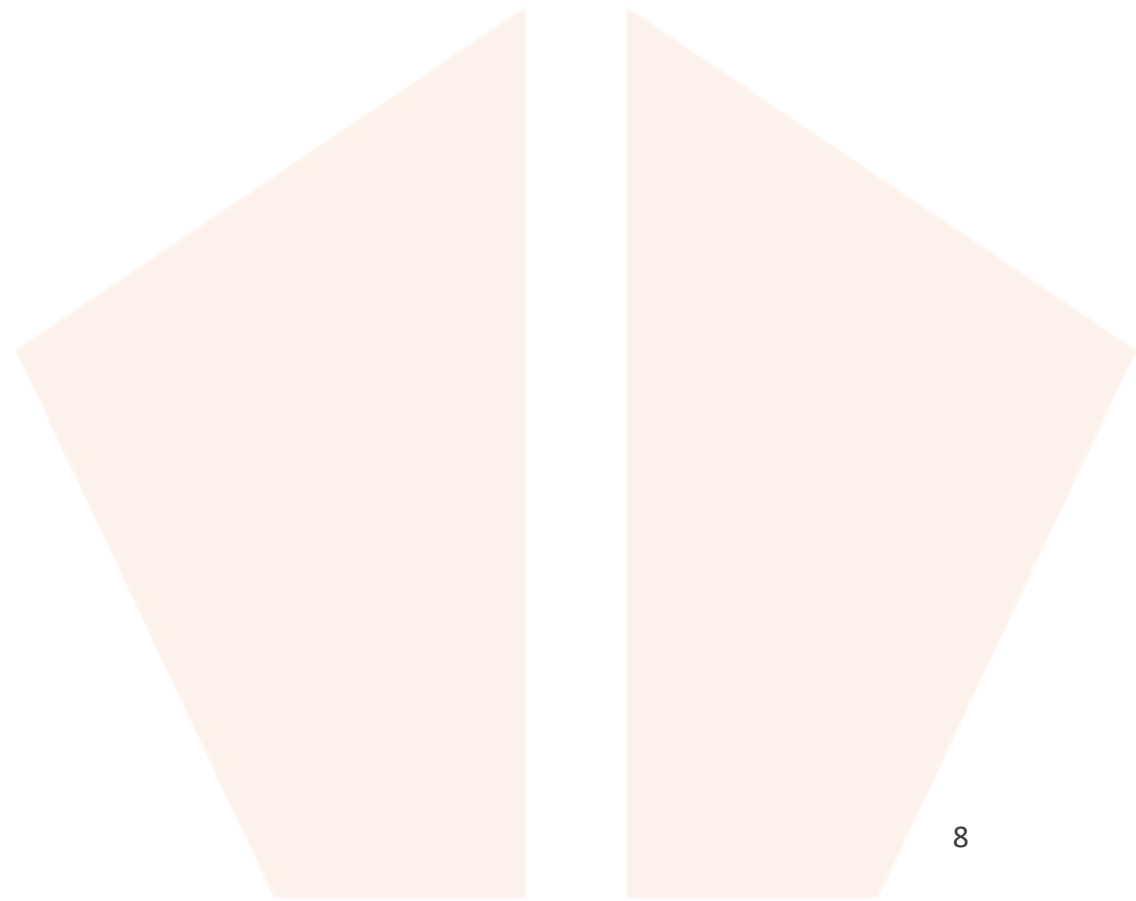
未知データに対しても学習データと同じ前処理を行ったうえで、学習済みモデルによる予測を行います。

本プロジェクトでは、未知データを Python スクリプトで仮想的に生成しております。

- ▶ 実際の運用時には、左端にある未知データを予測したい対象のデータに差し替えて利用します。



アウトプットの説明



アウトプット

予測値

本プロジェクトでは、仮想的に生成した未知データに対する予測を行った結果が出力されます。

可視化画面:[テストデータを生成]ノード

No.	商品コード CATEGORY	購入日 DATE	件数 FLOAT
1	400000001	2021/09/26	1.000000
2	400000001	2021/09/26	1.000000
3	400000003	2021/09/26	2.000000

可視化画面:[モデル適用(テストデータに対する予測:後処理後)]ノード

No.	商品コード CATEGORY	購入日 DATETIME	件数 FLOAT	件数予測結果 INT
1	400000001	2021/09/26 00:00:00	2.000000	0
2	400000003	2021/09/26 00:00:00	2.000000	0



テストデータ

予測値の出力

0 となっているため、
2021/9/26 ~ の翌週は 1 件も売れないだろう
という予測になっています。

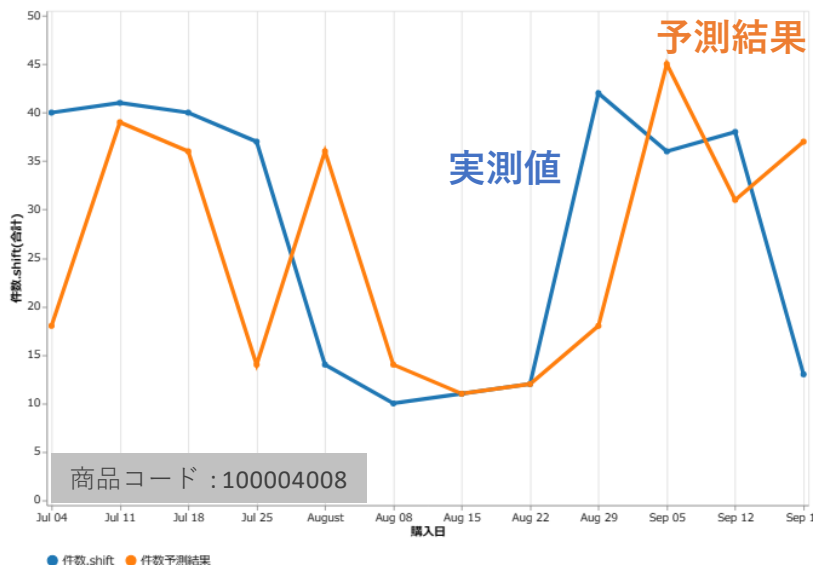
アウトプット

実測値と予測値のプロット

訓練用データによって学習したモデルを検証用データに適用し、予測が適切に行えているかを視覚的に確認します。

全体的な需要の変化傾向の把握ができています

可視化画面:[実測値と予測結果の比較]プロット



アウトプットのための処理

今回は、購入件数の多かった商品コード:100004008の商品に対して、左図のようにプロットを行ってみます。なお、購入件数は「集計」アイコンを利用して商品コード毎に総和を計算します。

アイコンの設定

アイコンの設定

時系列リサンプリング

変数選択

再標本の計算に用いる変数を設定します。

- キー列

1 系列を識別する変数を指定します。

系列数が 1 つの場合は選択しません。

- 計算対象列

集計対象となる変数を指定します。

再標本区間において計算対象列で指定した変数の値を集計します。

- 時刻列

時刻を表す変数を指定します。指定した変数にもとづいて再標本を行います。

時系列リサンプリング
?
×

変数選択

列名	列型	キー列	計算対象列	時刻列
商品コード	カテゴリ	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
顧客コード	カテゴリ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
購入日	日付	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
価格	整数	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
件数	整数	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
商品カテゴリコード	カテゴリ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
発売月	日付	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

計算方法

集計方法*

合計 ▼

再標本ごとの集計方法をを設定します。

再標本間隔*

週 ▼

再標本間隔をを設定します。

ウィンドウサイズ

1

整数値を設定します。

実行

保存

アイコンの設定

時系列リサンプリング

計算方法

再標本区間ごとの計算方法を指定します。

- **集計方法**

再標本区間ごとの集計方法を指定します。

- **再標本間隔**

年・月・週・日から再標本間隔を選択します。
例えば、「週」を選択すると週ごとのデータになるようにデータをまとめ上げます。

- **ウィンドウサイズ**

再標本間隔の長さを正の整数で指定します。
再標本間隔が「週」、ウィンドウサイズが「1」の場合、一週間ごとのデータにまとめ上げることを意味します。

時系列リサンプリング
?
×

変数選択

列名	列型	キー列	計算対象列	時刻列
商品コード	カテゴリ	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
顧客コード	カテゴリ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
購入日	日付	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
価格	整数	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
件数	整数	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
商品カテゴリコード	カテゴリ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
発売月	日付	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

計算方法

集計方法*

合計 ▼

再標本ごとの集計方法をを設定します。

再標本間隔*

週 ▼

再標本間隔を設定します。

ウィンドウサイズ

1 _____

整数値を設定します。

実行

保存

アイコンの設定

シフト

変数選択

シフト操作に用いる変数を指定します。

- キー列

1 系列を識別する変数を指定します。
系列数が 1 つの場合は選択しません。

- 計算対象列

シフト対象となる変数を指定します。
指定した列に対してシフト操作を行います。

入力データを出力に含める

チェックを入れると出力テーブルに入力された
データを含めます。

シフト
?
_
×

変数選択

列名	列型	キー列	計算対象列
商品コード	カテゴリ	<input checked="" type="checkbox"/>	<input type="checkbox"/>
購入日	日時	<input type="checkbox"/>	<input type="checkbox"/>
件数	実数	<input type="checkbox"/>	<input checked="" type="checkbox"/>
月	整数	<input type="checkbox"/>	<input type="checkbox"/>
年週数	整数	<input type="checkbox"/>	<input type="checkbox"/>
月週数	整数	<input type="checkbox"/>	<input type="checkbox"/>

入力データを出力に含める

シフト数の設定

シフト数
-1

実行

保存

アイコンの設定

シフト

シフト数の設定

シフト数を整数値で指定します。

正の値を指定した場合

行が下方方向にシフトします。

負の値を指定した場合

行が上方方向にシフトします。

シフト

変数選択

列名	列型	キー列	計算対象列
商品コード	カテゴリ	<input checked="" type="checkbox"/>	<input type="checkbox"/>
購入日	日時	<input type="checkbox"/>	<input type="checkbox"/>
件数	実数	<input type="checkbox"/>	<input checked="" type="checkbox"/>
月	整数	<input type="checkbox"/>	<input type="checkbox"/>
年週数	整数	<input type="checkbox"/>	<input type="checkbox"/>
月週数	整数	<input type="checkbox"/>	<input type="checkbox"/>

入力データを出力に含める

シフト数の設定

シフト数
-1

実行

保存

補足情報

技術的な情報や利用規約について

本文書・プロジェクトファイルのご利用にあたって

本文書ならびにプロジェクトファイルは、（株）NTT データ数理システム（以下「弊社」）が開発・販売する分析プラットフォーム **Alkano** についての情報提供として弊社が作成を行ったものです。

プロジェクトファイルは、ご利用者様の責任のもとで改変して利用することができますが、これに対するリバースエンジニアリングを禁じます。また、本文書ならびにプロジェクトファイルのご利用に際して、ご利用者様および第三者に損害が発生したとしても、弊社は責任を負わないものとします。

プロジェクトファイルは、その中に同梱されているデータを利用し、本文書内で解説している設定可能なパラメータで動作させた場合についてのみ、弊社にて動作の検証を行っております。これを超えるような状況における動作は保証いたしません。

本プロジェクトファイルは、**MSIP1.8.2** および **Alkano1.2.2** にて動作確認を行っております。



データ活用の確かなパートナー

お問い合わせ: 株式会社NTTデータ数理システム 営業部

Tel: 03-3358-6681

E-mail: alkano-info@ml.msi.co.jp

WEB: <https://www.msi.co.jp/alkano/>

株式会社 NTTデータ数理システム

NTT DATA NTT DATA Mathematical Systems Inc.