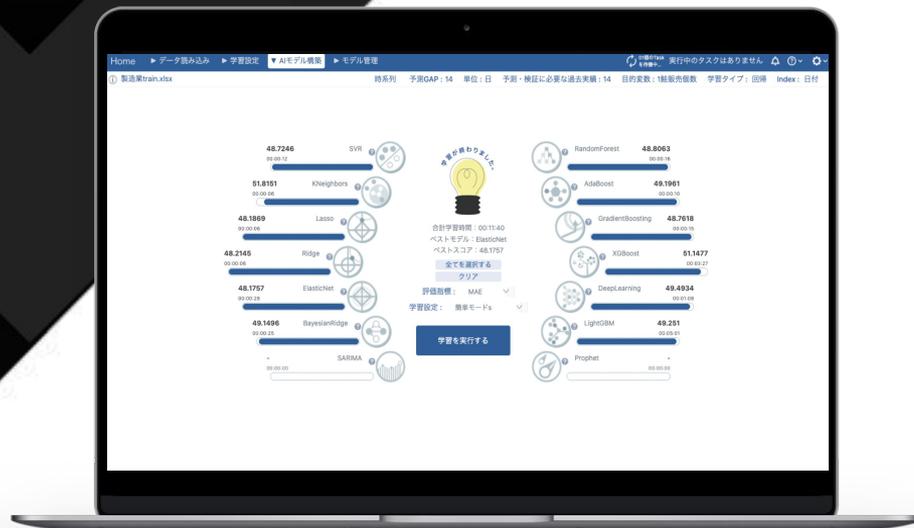


Learning Center Forecast

マニュアル



*バージョンによって実際の画面と若干違う場合がございます。

Learning Center Forecast は、今まではデータサイエンティストと呼ばれる専門家たちが、プログラミングや高度な統計学的知識を使って行っていた「AIによるモデル構築と予測」を、GUI上で、誰でも簡単にできるように作られたツールです。

この小冊子はLearning Center Forecast でAIモデルの構築と予測を開始するための手順を記したものです。

• 従来



• Learning Center Forecast



AGENDA

- 1 初回登録～ログイン
- 2 プロジェクト作成
- 3 タスクについて
- 4 特徴量タイプの設定
- 5 データ型の設定
- 6 ヒストグラムの表示
- 7 特徴量自動選択
- 8 相関ヒートマップの作成
- 9 フィルター・ソートの使い方
- 10 次元圧縮

AGENDA

- 11 時系列① 時系列データの基本設定
- 12 時系列② 時系列データの詳細設定
- 13 時系列③ イベントの設定
- 14 時系列④ 周期性の確認と設定
- 15 時系列⑤ 時系列設定の仕様
- 16 トランザクションデータ
- 17 学習設定
- 18 作成特徴量一覧
- 19 AIモデル構築
- 20 モデル検証

AGENDA

- 21 行ごとの影響度の計算
- 22 デプロイとデータ予測
- 23 APIを使用したデータ予測
- 24 プリセットデータ登録
- 25 規約について
- 26 お問い合わせ

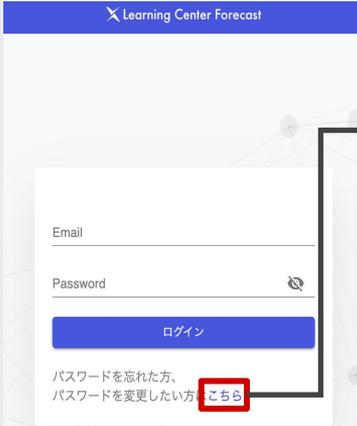
AGENDA

1 初回登録～ログイン

Learning Center Forecast の使用を開始するときに、**セキュリティを担保するために必ず行なっていただく「パスワードの再設定」と、ログイン方法について記載しています。**

①パスワード設定

②ログイン



「こちら」をクリックし、パスワードの再設定画面へ遷移



登録したID(メールアドレス)を入力し、「送信」ボタンをクリック



パスワードの再設定のメールが配信されて、メール内の「パスワードの設定を行う」ボタンをクリックし、新しいパスワードを設定

※メールが届かない場合、「support_lcf@inside.ai」までお問い合わせください。



ログインページより、ご登録のメールアドレス・パスワードを入力し、「ログイン」ボタンをクリック

Learning Center Forecast

Email
*****@inside.ai

Password
.....

ログイン

パスワードを忘れた方、
パスワードを変更したい方はこちら

「ログイン」ボタンをクリック

Learning Center Forecast 使用規約

Learning Center Forecast 使用規約

1. 当社は、両社の間に就いて、本規約を締結する旨、貴社側の承諾の内容およびその効力発生時期を、効力発生時期の到来までにインターネットでの掲載その他の方法により周知することにより、本規約を変更することができるものとします。

第19条 (権利義務の譲渡の禁止)

1. 使用者は、当社の書面による事前の承諾なく、使用契約上の地位又は本規約に基づく権利若しくは義務を第三者に譲渡し、又は担保に供することはできません。

第20条 (準拠法・裁判管轄)

1. 本規約および使用者と当社との諸契約に関しては、日本法を準拠法とします。

2. 本サービスに関して紛争が生じた場合には、東京地方裁判所を第一審の専属的合意管轄裁判所とします。

第21条 (名称の使用)

1. 当社は、使用者の名称等の情報を、営業又はマーケティングにおいて、本サービスの導入実績として使用することができるものとします。

第22条 (存続条項)

1. 第6条、第10条、第13条第2項から第4項まで、第14条から第15条

本規約に同意する **使用を開始する**

「Learning Center Forecast 使用規約」についてのポップアップ画面が表示され、契約内容をご確認の上で、「**使用を開始する**」もしくは「同意しない」を選択してください。タスクの自動更新に必要な機能なので、できる限り「使用を開始する」を選択してください。

※同意しない場合は、手動でタスクの更新を行っていただく形になります。

プロジェクトダッシュボード

Home Learning Center Forecast 実行中のタスクはありません

新しくプロジェクトを作成する

プロジェクト一覧

飛行機乗客数予測 [アプレット](#)
2022/06/19 18:27

タイタニック生還予測 [アプレット](#)
2022/06/19 18:27

Information

AMATERAS RAY の名称変更 およびこれに付随する変更について
2022/06/22

ver2.11 プリセットデータを刷新しました
2022/06/02

ver2.10 外れ値の丸め込み機能をリリース
2022/04/15

【重要】SARIMAのハイパーパラメータチューニング方法を修正
2022/02/16

プロジェクトダッシュボード画面が表示される

AGENDA

2

プロダクト作成

ログインが完了したら、いよいよプロジェクトを作成していきます。
学習用のデータをお手元に用意して、先にお進みください。



画面左上にある「新しくプロジェクトを作成する」ボタンをクリック



プロジェクト作成時のルールが表示され、「プロジェクトを作成する」ボタンをクリック



<記入項目>
 ・File
 AIに学習させたいデータをアップロードする
 ・プロジェクト名
 ※現在の対応形式は「csv、xls、xlsx」です。
 ※今後ファイルの対応形式は変更される可能性がありますので、ご注意ください。

すべての記入が完了したら、「作成」ボタンをクリック



プロジェクトの作成が完了すると、プロジェクト一覧にプロジェクト名が表示され、クリックするとAIプロジェクト画面に遷移することができる

Step.1 コピーしたいプロジェクトを選択

Step.2 プロジェクトをコピー

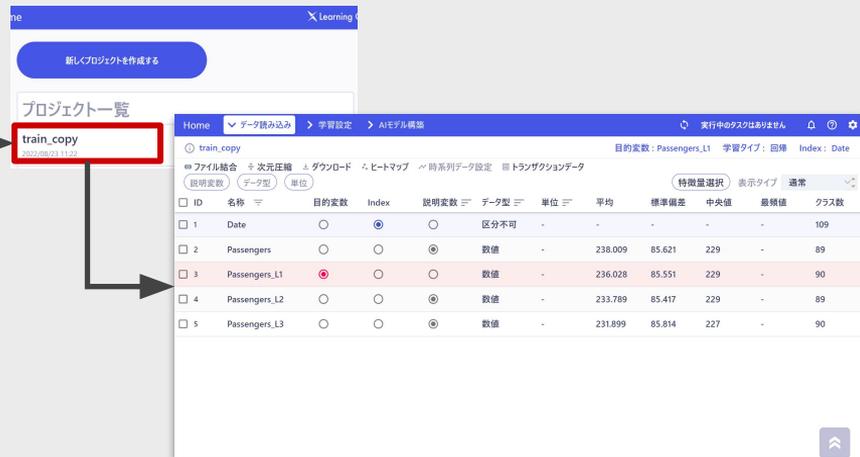
Step.3 コピーしたプロジェクトを確認



コピーしたいプロジェクトを選択



「コピー」ボタンをクリック



プロジェクトのコピーが完了すると、プロジェクト一覧にプロジェクト名が表示され、クリックするとAIプロジェクト画面に遷移することができる。

コピーされた項目の一覧:

- ・学習データ
- ・変数設定
(データ型、目的変数、Index、説明変数の不使用)
- ・時系列設定
- ・トランザクションデータ
- ・結合したデータ

※学習内容などはコピーされませんので、ご注意ください。

AGENDA

3

タスクについて

Learning Center Forecast は、データに関わる処理や、AIの学習など、コンピュータパワーを使う処理をタスクで処理していきます。
ここからタスクの種類や削除について説明します。

プロジェクト
作成、コピー

プロジェクト
結合、解除

次元圧縮データ
作成

次元圧縮プロ
ジェクト作成

ヒートマップデー
タ作成

ヒストグラムの作
成

特徴量自動選択

AIモデル学習

モデル検証

データ予測

データ型の変更

行ごとの影響度
の計算

行ごとの影響度
のダウンロード

タスクをクリックすると、タスクの詳細を確認することができます。
 ※「タスク開始時間」「現在の状況」「タスク進捗」は、進行中のタスクの場合のみ表示される。



タスクの詳細

id	66082b96da4bb2eed7a5c6ecc2d5a9a
taskState	doing
taskProgress	51.8472222222219
createdUser	dec178fef156883fedcb943627d8f30c
createdAt	08/23/2022
type	ライトGBMで学習しています
projectId	a043cd63c5c2bed7584b8d2e5e989aae
startedAt	2022/08/23 12:19
algorithmId	25



ゴミ箱のアイコンをクリック

画面右上のタスクバーをマウスでフォーカスすると、
 現在処理中のタスクの一覧を確認することができる

削除実行の注意ポイント:

- ・実行中のタスクを削除すると、タスクの処理は中断されます。
- ・学習タスクを削除すると、現在の学習状況を保存して、処理が中断されます。
- ・トランザクションデータの検証処理、ホールドアウト検証時の追加データの検証処理など削除できないタスクもありますのでご留意ください。

AGENDA

4

特徴量タイプの設定

プロジェクトが作成できたら、次はデータの中の「何を予測するか」をLearning Center Forecast に教えてあげる必要があります。

AIで何を予測するかという目的となるデータを「**目的変数**」と呼びます。

① 目的変数

目的変数とは、AIに「予測させたい対象」のことを意味します。
プロジェクト作成後に、まず行うのは「目的変数」の設定です。

Home ▼ データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: (なし)

ファイル結合 次元圧縮 ダウンロード ヒートマップ 時系列データ設定 トランザクションデータ

説明変数 データ型 単位 特徴量選択 表示タイプ 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	区分不可	-	-	-	-	-	109
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	239,009	85.621	229	-	89
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236,028	85.551	229	-	90
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233,789	85.417	229	-	89
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231,899	85.814	227	-	90

予測させたい対象の「目的変数」の列の丸いボタンをクリック

目的変数のデータ型が「数値」の場合、学習タイプは自動的に「回帰」になります。為替や株価等数値を予測する場合、この手法を用います。

データ型が「カテゴリ」の場合は、学習タイプは「分類」になります。明日の天気、与信評価など、いくつかあるクラスに分類したい場合、この手法を用います。

② Index

Indexに指定する特徴量は、必ず「一意の値」でなければいけません。
 (一意の値とは、ある一つの値を指定すると、ある一つの行を特定できるような特徴量の事です。
 いわゆる、どれ一つとして同じ値がない、すべてがユニークな値、ということになります)

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: Date

ファイル結合 次元圧縮 ダウンロード ヒートマップ 時系列データ設定 トランザクションデータ

説明変数 データ型 単位 特徴量選択 表示タイプ 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	区分不可						109
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90

設定したい「Index」の列の丸いボタンをクリック

Indexを指定すると、予測結果ダウンロード時に、予測結果データとIndexを結合してダウンロードすることができるようになるほか、Indexデータのデータ型を「日付型」に指定すると、時系列データとして学習することができます。

※時系列データの詳細や、設定方法「時系列設定」で解説しています。

③ 説明変数

説明変数とは、学習に使用する変数のことを指します。デフォルトでは、すべての変数が説明変数に指定されています。チェックボックスを使って、不要な変数を「使用しない」状態に変更することができます。

「使用」を選択すると、チェックを入れた変数をすべて「説明変数」に変更します。

「不使用」を選択すると、チェックを入れた変数をすべて「使用しない」に変更します。

まず、当該変数の左にあるチェックボックスをクリックし、チェックを入れます。

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: Date

ファイル結合 次元圧縮 ダウンロード ヒートマップ 時系列データ設定 トランザクションデータ

説明変数 データ型 単位 選択中: 1

特微量選択 表示タイプ 通常

使用	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
<input type="checkbox"/>	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	109
<input type="checkbox"/>	2 Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
<input checked="" type="checkbox"/>	3 Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
<input type="checkbox"/>	4 Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
<input type="checkbox"/>	5 Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90

3 Passengers_L1 数値 - 236.028 85.551 229 - 90

「不使用」を選んだ状態では、使用しない変数は灰色に表示されます。

AGENDA

5

データ型の設定

データ型を変更することで、より適切に変数を学び、学習効率や精度が向上することがあります。

Learning Center Forecast では4種類のデータ型を選ぶことができます。

Learning Center Forecast には「**数値型**」「**カテゴリ型**」「**日付型**」「**文字列型**」の4つのデータ型があります。

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: Date

ファイル結合 次元圧縮 ダウンロード ヒートマップ 時系列データ設定 トランザクションデータ

説明変数 データ型 単位 選択中: 1

特微量選択 表示タイプ 通常

<input type="checkbox"/>	ID	名称	数値	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
<input checked="" type="checkbox"/>	1	Date	カテゴリ	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	109
<input type="checkbox"/>	2	Passengers	文字列	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
<input type="checkbox"/>	3	Passengers_L1	日付型	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
<input type="checkbox"/>	4	Passengers_L2		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
<input type="checkbox"/>	5	Passengers_L3		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90

データ型が、「**区分不可**」と表記されているデータも存在します。「区分不可」は、「Learning Center Forecast には、データ型が推測できなかった」データということになります。

お客様の手で、適切なデータ型に変更して頂くことになります。
(変更しない場合、学習には使われません)

数値型

名前の通り、数値で量を表すデータを扱います。年齢、各国の人口、株価などは数値データです。変換を施すことなく、そのまま機械学習のデータとして扱うことができます。

日付型

日時を示すデータを扱います。日付データからは様々な特徴量を作ることができます。

Learning Center Forecast では、日付型をIndexに指定することで、時系列処理を施すことができます。詳しくは**時系列設定**の解説をご覧ください。

学習設定の**特徴量作成**の設定を変更することで、特徴量に変換することができます。基本的には、日付データそのものは学習には使われません。

カテゴリ型

いくつかの「クラス」に分けることができるデータを扱います。性別、天気、都道府県などです。

クラスとは、そのデータの分類可能なラベルのことです。たとえば性別であれば、生物学上人間は2つのクラス(男・女)に分類することができます。

カテゴリ型が文字の場合、アルゴリズムは文字を学習できないため、必ず何かの数値に変換して扱います。たとえば性別であれば、男> 0 女-> 1 という形です。

文字列型

名前、SNSでの投稿や製品番号などの文字が連なったデータを扱います。

学習設定の**特徴量作成**の設定を変更することで、文字列を形態素解析(単語に分けること)で分割し、特徴量に変換する機能があります。

詳細は**学習設定**の章をご覧ください。基本的には、文字列データそのものは学習には使われません。

「データ型」ボタンをクリックし、プルダウンメニューから変更したい型を選択する

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: Date

説明変数 データ型 単位 選択中: 1

ID	名称	数値	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	カテゴリ	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	109
2	Passengers	文字列	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
3	Passengers_L1	日付型	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
4	Passengers_L2		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
5	Passengers_L3		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 時系列 予測GAP: 1 単位: 月(月初) 予測・検証に必要な過去実績: 3 目的変数: Passengers 学習タイプ: 回帰 Index: Date

説明変数 データ型 単位 選択中: 1

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	日付型	-	-	-	-	-	109
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90

データ型が変わると、取得できる統計量(平均、クラス数など、データを代表する様々な数値)も変わってきます。

左上のヘッダーの
チェックボックスに
チェックを入れる
と、データすべてを
選択した状態にな
ります。

Home > データ読み込み > 学習設定 > AIモデル構築

train 時系列 予測GAP: 1 単位: 月(月初) 予測・検証に必要な過去実績: 3 目的変数: Passengers 学習タイプ: 回帰 Index: Date

ファイル結合 次元圧縮 ダウンロード ヒートマップ 時系列データ設定 トランザクションデータ

説明変数 データ型 単位 選択中: 5

特微量選択 表示タイプ 通常

<input checked="" type="checkbox"/>	ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
<input checked="" type="checkbox"/>	1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	日付型	-	-	-	-	-	109
<input checked="" type="checkbox"/>	2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
<input checked="" type="checkbox"/>	3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
<input checked="" type="checkbox"/>	4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	233.789	85.417	229	-	89
<input checked="" type="checkbox"/>	5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	231.899	85.814	227	-	90



型変更ができないデータがあります

Date

Dateの型変更に失敗しました。

指定した変数はすべてユニークな値なので、カテゴリカルな変数として扱えません。
「クラス数」が行数よりも少ない場合のみカテゴリカルな変数として指定できます

OK

この状態ですべての型を一度に変更することもできます。

ただし、左の画像のように、型変更ができない変数もあるので注意してください。

たとえば「各国の人口」のようなあまりにもバラバラな数値（一致するものが一つもない）データは、カテゴリ型に変更することはできません。

次のページに各データ型への変更基準をまとめました。

数値型

対象のデータが以下の条件を満たす場合、数値型として扱うことができます。

- ・ 1つでも数値に変更できるデータを含んでいる

例①: ['a', 'b', 'c', 1, 'd'] -> 変換可能

例②: ['a', 'b', 'c', 'c', 'd'] -> 変換不可能

日付型

対象のデータ型以下のどちらかを満たす場合、日付型として扱うことができます。

- ・ 数値データ(数値データはUNIX時間として計算されます)
- ・ いずれかの日付のフォーマットに準拠している

例①: [1, 2, 3, 4, 5] -> 変換可能

例②: ['2019/10/09', '2019/10/10', '2019/10/11', '2019/10/12']
-> 変換可能

カテゴリ型

対象のデータが「完全にユニーク」でない場合は、カテゴリ型として扱うことができます。

「完全にユニーク」とは、データのすべての値が重複しない状態のことを意味します。

*また重複しない値は、分析時に「Rare」というカテゴリに置き換えられます。

文字列型

すべてのデータを変換することができます。

AGENDA

6

ヒストグラムの表示

ヒストグラムは、データの偏りを直感的に確認するための優れた可視化手法です。

Learning Center Forecast では、これを簡単に、すべての変数に対して行うことができます。



ヒストグラムを確認するには、画面右上の「表示タイプ」を「ヒストグラム」に変更する



ヒストグラムの作成が終わると、作成できたものは、左の画像のように表示されます。数値データの場合、binは10で固定されています。

カテゴリデータの場合、そのクラスの数だけbinが設定されます。またデータによってはヒストグラムが作成できないものもあります。先ほど変更した「日付型」はヒストグラムを作成することができません。

AGENDA

7

特徴量自動選択

モデルの精度を高め、学習を効率よく進めるためには特徴量選択が不可欠です。

専門的な知見を通常必要とするこの工程を、Learning Center Forecast は自動で、かつ高い精度で実現します。

特徴量自動選択を行うには、データ読み込み画面の「特徴量選択」ボタンをクリックします。
※このボタンは「目的変数」が指定されていない場合、クリックすることができないので注意してください。

Home データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 分類 Index: Date

特徴量選択 表示タイプ: 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	文字列	-	-	-	-	-	109
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	カテゴリ	-	-	-	-	229	89
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	229	90
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	229	89
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	229	90



Home データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

train 目的変数: Passengers 学習タイプ: 回帰 Index: Date

特徴量選択 表示タイプ: 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	文字列	-	-	-	-	-	109
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-	89
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	233.789	85.417	229	-	89
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	231.899	85.814	227	-	90

特徴量自動選択が終わると、このように、使用しないほうが良い特徴量が自動的に「使用しない」に変更されます。

この状態で再度学習してみると、より良い結果が得られるかもしれません。

AGENDA

8

相関ヒートマップの作成

学習データを選択する際に一つの指標となるのが変数同士の「**相関係数**」です。

Learning Center Forecast はこの相関係数を一度に、**明確に可視化する機能**を備えています。

相関ヒートマップとは、各特徴量間の相関係数をヒートマップの形で表示する機能です。



画面の左側に表示されている、赤い枠で囲ってある部分が「目的変数との相関係数ランキング」です。

上の赤いラインのグラフが「目的変数との正の相関トップ30」です。
下の青いラインのグラフが「目的変数との負の相関ワースト30」です。

なお、このデータは目的変数が設定されていない場合、表示されません。

相関ヒートマップの作成は以上で完了です。



作成が完了すると、自動的にヒートマップ画面に遷移します。

画面の中央から右にある赤枠で囲った大きな表が「相関ヒートマップ」です。目的変数が指定されている場合、目的変数が最上位に表示されます。

列と行が交わったところに表示されている数値が、相関係数です。全ての説明変数と目的変数について、総当たりで表示されます。

AGENDA

9

フィルター・ソートの使い方

フィルター・ソートの使い方

データの特徴量が膨大にある場合、目的の特徴量を見つけるのに苦労することがあります。そういう場合は、表示する特徴量を絞り込むことができる「フィルター」機能を使用します。

フィルター機能を使用するにはデータ読み込み画面の「名称」列にあるアイコンをクリックしてください。

The screenshot shows the 'AIモデル構築' (AI Model Building) interface. The '名称' (Name) column is selected, and the filter menu is open. The '全解除' (Reset) button is highlighted in red. The filter menu shows a list of variables with checkboxes, including PassengerId, Name, SibSp, Fare, Survived, Sex, Parch, and Cabin.

「全解除」をクリックすると、すべての選択がはずれ、データが表示されなくなります。

「全選択」をクリックすると、すべてのチェックボックスが選択され、データがすべて表示されるようになります。

フィルター・ソートの使い方

ソート機能は、データを並び替える機能です。「説明変数」と「データ型」について使用することができます。各列名の右側にある、矢印ボタンをクリックすると、変更することができます。デフォルトでID順となります。

説明変数

データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

① タイタニック生還予測 目的変数: Survived 学習タイプ: 分類 Index: PassengerId

② ファイル結合 ③ 次元圧縮 ④ ダウンロード ⑤ ヒートマップ ⑥ 時系列データ設定 ⑦ トランザクションデータ

説明変数 データ型 単位

特徴量選択 表示タイプ 通常

ID 名称 目的変数 Index 説明変数 データ型 単位 平均 標準偏差 中央値 最頻値 クラス数 欠損値

説明変数のソート順:

「デフォルト -> 説明変数 -> 不使用」の順番で切り替わります。

Home データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

① タイタニック生還予測 目的変数: Survived 学習タイプ: 分類 Index: PassengerId

② ファイル結合 ③ 次元圧縮 ④ ダウンロード ⑤ ヒートマップ ⑥ 時系列データ設定 ⑦ トランザクションデータ

説明変数 データ型 単位

特徴量選択 表示タイプ 通常

<input type="checkbox"/>	ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数	欠損値
<input type="checkbox"/>	3	Pclass	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	2.305	0.837	3	-	3	0
<input type="checkbox"/>	5	Sex	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	male	2	0
<input type="checkbox"/>	6	Age	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	29.871	14.544	28	-	86	163
<input type="checkbox"/>	7	SibSp	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	0.519	1.064	0	-	7	0
<input type="checkbox"/>	8	Parch	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	0.374	0.801	0	-	7	0
<input type="checkbox"/>	10	Fare	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	33.039	51.525	14.500	-	234	0
<input type="checkbox"/>	12	Embarked	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	S	3	1
<input type="checkbox"/>	1	PassengerId	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	数値	-	400.500	231.084	400.500	-	800	0
<input type="checkbox"/>	2	Survived	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	カテゴリ	-	-	-	-	No	2	0
<input type="checkbox"/>	4	Name	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	800	0
<input type="checkbox"/>	9	Ticket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	625	0
<input type="checkbox"/>	11	Cabin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	136	615

Home データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

① タイタニック生還予測 目的変数: Survived 学習タイプ: 分類 Index: PassengerId

② ファイル結合 ③ 次元圧縮 ④ ダウンロード ⑤ ヒートマップ ⑥ 時系列データ設定 ⑦ トランザクションデータ

説明変数 データ型 単位

特徴量選択 表示タイプ 通常

<input type="checkbox"/>	ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数	欠損値
<input type="checkbox"/>	1	PassengerId	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	数値	-	400.500	231.084	400.500	-	800	0
<input type="checkbox"/>	2	Survived	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	カテゴリ	-	-	-	-	No	2	0
<input type="checkbox"/>	4	Name	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	800	0
<input type="checkbox"/>	9	Ticket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	625	0
<input type="checkbox"/>	11	Cabin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	区分不可	-	-	-	-	-	136	615
<input type="checkbox"/>	3	Pclass	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	2.305	0.837	3	-	3	0
<input type="checkbox"/>	5	Sex	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	male	2	0
<input type="checkbox"/>	6	Age	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	29.871	14.544	28	-	86	163
<input type="checkbox"/>	7	SibSp	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	0.519	1.064	0	-	7	0
<input type="checkbox"/>	8	Parch	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	0.374	0.801	0	-	7	0
<input type="checkbox"/>	10	Fare	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	33.039	51.525	14.500	-	234	0
<input type="checkbox"/>	12	Embarked	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	カテゴリ	-	-	-	-	S	3	1

「説明変数」のソートボタンを1度押すと、「説明変数」がリストの上部に来るように並び替えられます。

「説明変数」のソートボタンをもう1度押すと、「説明変数」以外の特徴量がリストの上部に来るように並び替えられます。

「データ型」をソートする場合も同じ操作で切り替えていきます。

データ型

み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

目的変数: Survived 学習タイプ: 分類 Index: PassengerId

特微量選択 表示タイプ 通常

説明変数	データ型	単位	特微量選択	表示タイプ
<input type="checkbox"/> ID	名称	目的変数	Index	説明変数
<input type="checkbox"/> 1	PassengerId	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/> 数値
<input type="checkbox"/> 2	Survived	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/> カテゴリ
<input type="checkbox"/> 3	Pclass	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 4	Name	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可
<input type="checkbox"/> 5	Sex	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> カテゴリ
<input type="checkbox"/> 6	Age	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 7	SibSp	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 8	Parch	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 9	Ticket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可
<input type="checkbox"/> 10	Fare	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 11	Cabin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可
<input type="checkbox"/> 12	Embarked	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> カテゴリ

データ型のソート順:

「デフォルト -> 学習に使用するもの(カテゴリ・数値)
-> その他のもの(区分不可、文字列、日付)」の順で、
切り替わります。

Home > データ読み込み > 学習設定 > AIモデル構築

実行中のタスクはありません

目的変数: Survived 学習タイプ: 分類 Index: PassengerId

特微量選択 表示タイプ 通常

説明変数	データ型	単位	特微量選択	表示タイプ
<input type="checkbox"/> 2	Survived	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/> カテゴリ
<input type="checkbox"/> 5	Sex	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> カテゴリ
<input type="checkbox"/> 12	Embarked	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> カテゴリ
<input type="checkbox"/> 1	PassengerId	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/> 数値
<input type="checkbox"/> 3	Pclass	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 6	Age	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 7	SibSp	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 8	Parch	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 10	Fare	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> 数値
<input type="checkbox"/> 4	Name	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可
<input type="checkbox"/> 9	Ticket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可
<input type="checkbox"/> 11	Cabin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> 区分不可

AGENDA

10

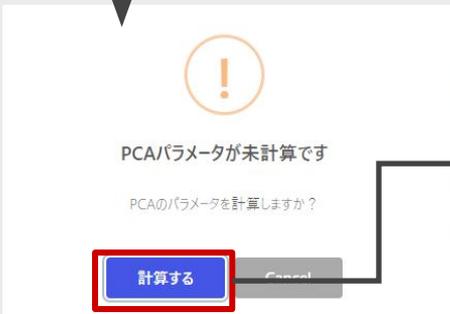
次元圧縮

次元圧縮は、データ量を削減する手法の一つです。データの特徴をできる限り保持しながら、データ全体を変形させることで、次元(列)を減らします。

「特徴量選択」と違い、元のデータとは全く違うデータに変換されるのが特徴ですが、ノイズに隠れて見えなかったデータの特徴が可視化されることもあります。



「次元圧縮」ボタンをクリック



PCAとは、次元圧縮の手法の一つです。Learning Center Forecast では現在PCAのみを使用することができます。

PCAのパラメータとは、次元を圧縮した際に、特徴量の情報量をどれだけ保持できるかの情報のことです。これをまず作成する必要があるため、「計算する」を選択します。なお、一度作成すると、データに何らかの変更（データ型変更、特徴量タイプ変更）を行わない限り、PCAデータは保持されます。

PCAパラメータの計算が終わると、左のようなモーダルが表示されます。

「次元数」とは削減目標の次元の数です。ここで指定した数まで「説明変数」が圧縮されます。このとき、Indexと目的変数は含まれていないことにご注意ください。Indexと目的変数は、圧縮後のデータに結合されます。

「説明力」とは、指定した次元数まで列を圧縮する場合、元のデータの情報量に対して、何%の情報量になるのかという目安です。説明力は次元数を変更すると、変わります。

処理が終わると、次元圧縮後のプロジェクトが追加されます。
{元プロジェクト名} 次元圧縮{圧縮したデータ名}の形式で名称が決まります。

プロジェクト一覧

次元圧縮Airpassengers_train.csv

2022/08/24 12:11

train

2022/08/23 11:46

プロジェクト名をクリック

Home ▼ データ読み込み 実行中のタスクはありません

① 次元圧縮Airpassengers_train.csv 目的変数: (なし) 学習タイプ: (なし) Index: (なし)

② ファイル結合 次元圧縮

説明変数 (学習変数) (単位) 特徴量選択 表示タイプ 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値	クラス数	欠損値
1	Date	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	区分不可	-	-	-	-	-	109	0
2	Passengers	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	238.009	85.621	229	-	89	0
3	c_0	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-	90	0

圧縮後のプロジェクトです。

Indexと目的変数は保持され、その他の特徴量が、指定した次元まで削減されていることがわかります。

次元圧縮

デフォルトで設定されている次元数は、データ全体に対して80%以上を説明できるようになっています。

次元数

説明力 100%

圧縮してプロジェクトを作成する

圧縮してダウンロードする

圧縮後のデータをダウンロードしたい場合は「圧縮してダウンロードする」をクリックします。

	A	B	C
1	Date	Passenge c_0	
2	1949/4/1	129	208.9471
3	1949/5/1	121	218.8793
4	1949/6/1	135	220.5582
5	1949/7/1	148	222.1618
6	1949/8/1	148	233.2353
7	1949/9/1	136	248.8806
8	#####	119	249.4566
9	#####	104	232.6842

タスクが完了すると、自動的にデータがダウンロードされます。ダウンロードしたデータは左のような形式になっています。

AGENDA

11

時系列設定

時系列データとは、たとえば天気予報や、為替の価格、交通状況など
時間ごとの状態を記録したデータのことです。

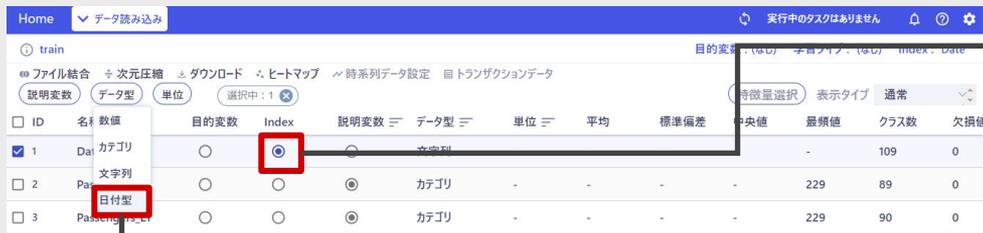
「過去の事象が未来に影響を与えている」という仮定に立っているため、通常のデータと同じように扱おうと、「未来のデータ」を学習データに含んでしまうことがあり、適切な学習ができない可能性があります。(「target leakage」といいます。)

本職のデータサイエンティストでさえ、こうしたデータの「target leakage」を完全に避けるのは難しく、結果的にモデルの精度が不当に高くなり、モデルの運用に失敗するケースもあります。

Learning Center Forecast には、こうした**取り扱いの難しい時系列データ**を適切に学習するための「時系列設定」を、簡単に行うことができる機能を実装しています。

時系列設定① 時系列データの基本設定

時系列データとして扱うための基本設定を行います。



Step.1 時系列設定を行う時は、まずデータ読み込み画面で **Index**を指定する必要があります。

Indexのデータ型は「**日付型**」である必要があります。
日付型に変換できないIndexは、時系列データとして扱うことができませんので、注意してください。
(ex: 商品番号、人名や地名等)

Step.2 データ型ボタンをクリックし、プルダウンメニューから「日付型」を選択



Step.3 最後に **目的変数** を設定し、準備完了です。



Step.4 次に画面上部のツールバーから「時系列データ設定」をクリック



①予測GAP設定

予測のGAPを設定します。

未来のデータの予測には必ずGAPが生じます。

たとえば、明日の株価を予測するための今日までのデータが、手元に完全に揃っていることは稀なことです。

こうした場合「一昨日のデータで予測する」というように、手元に完全に揃っているデータで予測できるように、学習データを自動で調整します。

直感的には「何日(単位による)後を予測したいか」を設定すれば良いと考えて下さい。

A. 予測単位

予測するデータの単位です。

単位は学習用データから自動的に読み取られます。

B. 決定ボタン

上記の変更を保存し、確定します。

②詳細設定ボタン

詳細設定に遷移します。

高度な設定が含まれるため、必ずしも必要ではありません。

詳細は「時系列データの詳細設定」の項で解説します。

③コレログラム

自己相関のグラフを表示します。

詳細は「周期性の確認と設定」の項で解説します。

Learning Center Forecast の時系列設定では、より詳細な学習設定を行うことができます。上手に設定すれば、モデルの精度向上に繋がるだけでなく、データからより良い知見を得ることができるはずですが、ただし、中には高度な内容も含まれているため、必須の設定ではありません。

時系列データ設定

Index名: Date 1 月(月初) 後を予測 設定 ① 詳細設定

時系列詳細設定

イベント設定 ①

設定済みイベント 0日 ① イベントを設定する

② 休日設定 ①

週末判定 祝日設定 日本

③ 飽和状態 ①

最大値 数値 最小値 数値

周期変数設定 ①

自動作成: する 周期変数設定

① イベント設定画面

周期イベントを作成することができます。詳細は「時系列設定③ イベントの設定」にて解説します。

② 休日設定

週末の設定の有無と、祝日の設定を変更できます。たとえば、レストランの売り上げ予測などを行う時は、週末や祝日などの休日に売り上げが上がるという仮説が成り立ちそうなので、有効にした方が精度が向上するかもしれません。祝日については、現在は日本の「国民の祝日」のみに対応しています。

③ 飽和状態

飽和状態とは、成長モデルにおける、成長の理論的な限界値のことです。たとえば、劇場や球場など、箱の大きさが決まっている場合、1日の入場者数には限界があります。こうした限界を超えてモデルが成長しないように、飽和状態を設定することで制限をかけることができます。

時系列設定③ イベントの設定

時系列モデルにおいて周期的に発生するイベントは非常に重要です。

たとえば、出前の寿司の売り上げを予測する場合、家族が集まる元旦の影響を無視することは絶対にできません。他に家族が集まる大型連休といえば、GWやお盆なども考えられます。GWとお盆とでは、お盆の方がより遠方から家族が集まる、という推測が成り立ちそうなので、影響力はGWより大きくなるかもしれません。

Learning Center Forecast は、こうした複雑なイベント同士の関係性をシンプルに設定し、自動で学習することができます。

時系列詳細設定

イベント設定 ①

設定済みイベント 0日

イベントを設定する

イベント設定

2022年8月

月	火	水	木	金	土	日
1	2	3	4	5	6	7
8	9	10	11 山の日	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28

まずは週次イベントから設定しましょう。

イベントの「起点」として使いたい日付をクリックします。

起点とは、設定上のイベント開始日のことで、目的となる周期に設定できれば、いつでも構いません。

たとえば、毎週土曜日に定期的に行うイベントであれば、今月の土曜日を指定しても構いません。

イベントの詳細設定で、細かい周期の設定についてもカバーできます。

起点を決定したら、「新規作成」をクリックします。

イベント設定

① イベント名

② タグ

繰り返し

③ 数値 1 年 に1度 数値 2022 年 まで ⓘ

この日を指定

週位置で指定

④ 8月13日

8月 2週目 土曜日

⑤ 該当する日付がない時 実施しない

⑥ この日より前にもイベントを適用

⑦ イベントのインパクト

範囲前

数値

後

数値

影響力 ⓘ

数値

① イベント名

必須要素です。

② タグ

タグが同じイベントは「同一のイベント」として学習されます。たとえば、オリンピックのような、日付が定まっていないイベントを同じイベントとして扱いたい場合、共通のタグをつけることで、周期とは関わりなく同一のイベントとして学習することができます。

③ 周期

イベントの周期を設定できます。発生スパンは数値で設定し、周期は年、月、週、なしの4種類から設定できます。

④ イベント位置の設定

日付を直接指定するか、週の位置で設定するかを決められます。

⑤ 該当する日付がない時

年や月によっては、存在しない日付や週位置があります。(たとえば2月には30日は存在しません、第5週が存在しない場合もあります) こうした該当する日付がないイベントを、**実施しない**、**月末**、**前週**の3種類から設定できます。

⑥ 過去へのイベント適用

ONIにすることで、起点となった日よりも過去へ、イベントを適用できます。学習期間全てにわたって適用したい場合、ONIにすると便利です。

⑦ イベントへのインパクト

影響力や、影響範囲を設定できます。次のページで詳述します。

イベント設定

イベント名
週末セール

タグ セール

繰り返し

数値 2 週 に1度 数値 2022 年 8 月まで

この日を指定 2022/08/13

週位置で指定 土曜日

該当する日付がない時 実施しない

この日より前にもイベントを適用

イベントのインパクト

範囲前 数値 後 数値 影響力 3

決定

たとえば、2週間に1度行う週末セールであれば、左のように設定できるかもしれません。
タグに「セール」属性をつけることで、週末以外で行われるセールも、同じセールとして学習できます。

影響力は3を設定しています。
影響力は10段階で設定できますが、あくまでも相対的な値です。
まずは感覚的にこの値に決めています。

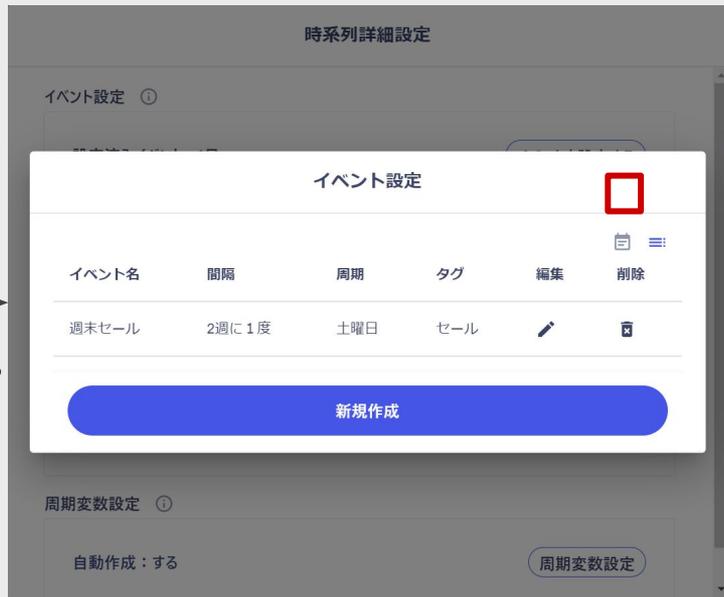
影響範囲については、週末セールは、日付の前後には大きな影響を与えないと考え、設定していません。

設定が終わったら「決定」ボタンをクリックします。



自分が設定したイベントの起点日がわからなくなった時は、右上の**一覧ボタン**をクリックすると、イベントを一覧で表示することができます。

起点日を2020/02/15としたので、2週間後の02/29に次のイベントが発生しているのを確認できます。
これで週末セールの設定は完了です。



このように一覧でイベントを表示できます。
イベントの編集、削除も行うことができます。
カレンダーボタンをクリックすると、カレンダーに戻ります。
次に、月末セールを設定してみましょう。

・月末セール

タグの項目に「セール」で共通設定をします。

・発生周期

月に1度。日付は31日を指定しました。

※月によっては31日が存在しない場合があるため、「該当する日付がない時」の設定として、**月末**を設定します。これで**「すべての月の末日で発生する」**という設定にすることができます。

・影響力

5に設定します。

これは、月末セールは週末セールよりも大規模なため、集客能力があるという判断からです。

イベント設定

イベント名 月末セール

タグ セール

繰返し

数値 1 月 に1度 2022 年 9 月まで

この日を指定 31日

週位置で指定 5週目 水曜日

該当する日付がない時 月末

この日より前にもイベントを適用

イベントのインパクト

範囲前 数値 後 数値 影響力 5

削除 修正

イベント設定

2022年9月

月	火	水	木	金	土	日
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19 敬老の日	20	21	22	23 秋分の日	24	25
26	27	28	29	30 月末セール	1	2

新規作成

このように31日が存在しない月でも月末に設定できています。

時系列分析では周期性の確認と、周期に合わせたデータを作ることが非常に重要になります。
Learning Center Forecast では、そうした高度な操作を完全に自動化できる他、周期を自分で確認してラグ変数を作成することもできます。

時系列データ設定

Index名 : Date 1 月(月初) 後を予測 設定 ⓘ 詳細設定

コレログラム 階差 ⓘ

数値 6 ⓘ

偏自己相関データがありません 数値 ⓘ

周期性の確認には「コレログラム」を利用することができます。
Learning Center Forecast では、2種類のコレログラムを確認することができます。



まず「自己相関」を確認しましょう。
「自己相関データはありません」の右にある入力欄に数値を入力してください。

その後、その右側にある「決定ボタン」をクリックしてください。計算には、少し時間がかかる場合があります。

しばらくすると、このようにマッチ棒のようなグラフが表示されます。

グラフの棒は、左から「元のデータ」と「元データからn個ずらしたデータ」の相関の大きさを表しています。
相関係数なので、一番左の棒の値は1です。



しばらくすると、このようにマッチ棒のようなグラフが表示されます。

グラフの棒は、左から「元のデータ」と「元データからn個ずらしたデータ」の相関の大きさを表しています。
相関係数なので、一番左の棒の値は1です。

偏自己相関も自己相関と同様の操作で、確認することができます。



偏自己相関と自己相関の違いは、「ずらしたデータより前のデータの影響を考慮するか、しないか」という点にあります。

自己相関は、左の画像のようにきれいな階段状になっており、これは前日の影響が次の日に残っているためです。

一方、偏自己相関は、前日の影響などを除いて、元データとずらした後のデータの相関を直接計算します。

左の画像の様に、通常の自己相関とは大きく結果が違うのがわかります。



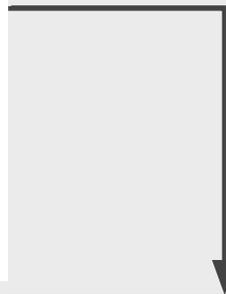
「階差」のチェックボックスをONにすると、「データの階差の自己相関」を表示します。階差とは、今の値から一つ前の値を差し引いたものを配列にしたデータのことです。

たとえば[1,2,4,8]で左から古いデータだとすると、[Nan,1,2,4]という階差を取得できます。

階差をとることにより、いわゆる「見せかけの回帰(相関)」を回避することができます。

左の画像のように、先ほどまでのコレログラムとは大きく形が違うのがわかります。

コレログラムを確認した上でラグを自分で設定して学習することもできます。



手動で設定したい場合は、時系列データ設定画面より、「詳細設定」>「周期性変数設定」をクリックして設定することができます

ラグを使用せずに、Learning Center Forecast にラグの作成を任せることもできます。その場合は、「自動作成チェックボックス」をONにしてください。



手動でラグを作りたい場合は、自動作成をOFFにし、「周期の作成」をクリックして、ラグの値を入力し、「決定」ボタンをクリックします。



時系列設定の仕様一覧です。

特徴量作成*

- 目的変数に対する**ラグ、差分、微分、移動平均、誤差系列、分散**
- 日付の分解
- イベント、週末、月末、祝日の影響の反映

学習時の仕様

- 欠損値の補完は**線形補完**、あるいは**前回の値**で補完
- CrossValidationには「**TimeSeriesCV**」を採用
- カテゴリデータの変換には「**OneHotEncoding**」を使用する*
- スケーリングのパラメータは、**学習データのもの**を使用する

*「学習モードが旧学習モード以外のときのみ適用」されます。

AGENDA

16

トランザクションデータ

トランザクションデータとは、業務に伴って発生した出来事(たとえば取引)を記録したデータのことです。トランザクションデータには、非常に大きな情報が含まれていますが、それ単体では学習に使用できないことが多いです。

Learning Center Forecast には、学習データのIndexをもとに、トランザクションデータから特徴量を作成する機能が実装されています。

date	id	value
2020/8/5	1100	100
2020/8/6	3311	200
2020/8/6	1219	10
2020/8/6	9865	1
2020/8/7	1100	100
2020/8/8	3928	150
2020/8/8	4738	20
2020/8/9	5940	50
2020/8/10	1100	100
2020/8/11	3904	10
2020/8/11	8569	20
2020/8/11	4859	100
2020/8/11	3311	120

トランザクションデータのイメージは左のようなデータです。

時系列順に、何らかの出来事が記録され、その記録に紐づいた(左のデータのid)と値(左のデータのvalue)が並んでいます。

たとえば、id は会員カードと紐づいた「顧客ID」で、value は「購入した金額」といったイメージです。

このid(学習データと、1対多の関係にある)は、学習データのIndex と紐づいている必要があります。

Learning Center Forecast では、こうしたデータから特徴量を抽出し、自動的に学習データ、検証データ、予測データに結合することができます。

ID	名称	目的変数	Index	データ型	単位	平均	標準偏差	中央値	最頻値	欠損値
1	SK_ID_CURR			数値	-	276,596	103,700	278,551	-	15,375
2	TARGET			数値	-	0.080	0.272	0	-	2
3	NAME_CONTRACT_TYPE			カテゴリ	-	-	-	-	Cash loans	2
4	CODE_GENDER			カテゴリ	-	-	-	-	F	2
5	FLAG_OWN_CAR			カテゴリ	-	-	-	-	N	2
6	FLAG_OWN_REALTY			カテゴリ	-	-	-	-	Y	2
7	CNT_CHILDREN			数値	-	0.414	0.715	0	-	7
8	AMT_INCOME_TOTAL			数値	-	168,867	101,784	153,000	-	349
9	AMT_CREDIT			数値	-	602,171	402,311	518,562	-	2,189
10	AMT_ANNUITY			数値	-	27,172	14,200	25,146	-	5,262
11	AMT_GOODS_PRICE			数値	-	541,645	369,348	450,000	-	399
12	NAME_TYPE_SUITE			カテゴリ	-	-	-	-	Unaccompanied	7
13	NAME_INCOME_TYPE			カテゴリ	-	-	-	-	Working	5
14	NAME_EDUCATION_TYPE			カテゴリ	-	-	-	-	Secondary / secondary special	5
15	NAME_FAMILY_STATUS			カテゴリ	-	-	-	-	Married	5
16	NAME_HOUSING_TYPE			カテゴリ	-	-	-	-	House / apartment	6

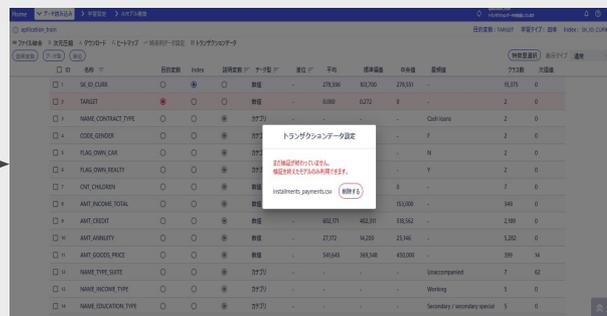
まず、トランザクションデータを読み込ませるには**Indexと目的変数**が指定されている必要があります。

条件が整ったら、データ読み込み画面上の「トランザクションデータ」という項目をクリックしてください。

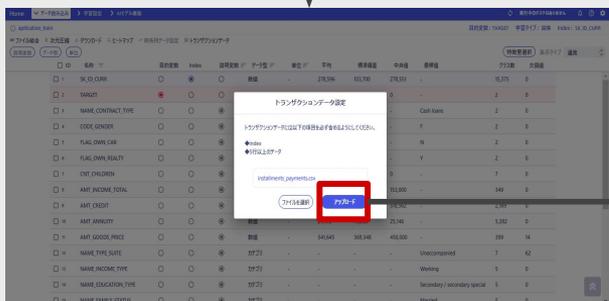
読み込ませるデータを選択します。



不適切なデータ(Indexが存在しない等)の場合



検証に失敗します。再度適切なデータに整えてから、アップロードし直してください。



アップロードが終わると、データの検証が行われます。この処理には少し時間がかかります。



読み込ませるデータを設定したら、アップロードボタンをクリックします。

データの検証に成功すると、このような画面に変わります。これでトランザクションデータを利用する準備が整いました。

トランザクションデータ ⓘ

前回の値 ⓘ
 前回からの距離 ⓘ
 統計量 ⓘ
 平均距離 ⓘ

平均時間 ⓘ
 直近n単位の登場回数 ⓘ
 直近n単位の統計量 ⓘ

トランザクションデータによる特徴量作成を有効にするには、学習設定を変更する必要があります。詳細は「学習設定」の章で解説します。

学習設定 > 特徴量作成設定 > トランザクションデータで変更できます。

現在対応しているのは「前回の値」「統計量」です。

今後作成できる特徴量が増えていく予定です。

The screenshot shows the 'Model Management' page for 'application_train'. It lists the 'Ridge' model with a learning time of 0.23 and accuracy of 0.143. Below, a 'Feature Importance' window is open, showing a list of features with their importance scores. A grey arrow points to the 'REGION_POPULATION_RELATIVE' feature, which has an importance score of 17. The text below the list states: 'ホールドアウト検証が済んでいません。検証結果は、ホールドアウト検証終了後のみ、表示可能です。'

★印がついており「トランザクション」と名称が入っているものが、トランザクションデータから作られた特徴量です。

特徴量重要度を確認すると、作成された特徴量が、モデルでの予測に役立っているのがわかります。

トランザクションデータの設定は以上です。

AGENDA

17 学習設定

機械学習には、より良くデータにフィットさせるための様々なパラメータがあります。Learning Center Forecast は、単に自動で学習を行うだけでなく、こうしたパラメータの設定も行うことができます。

やや上級者向けの機能ですので、初めてLearning Center Forecast に触れる方は、読み飛ばして頂いても構いません。

Home > データ読み込み > 学習設定 > AIモデル構築

飛行機乗客数予測 時系列 予測GAI 単位: 月(月初) 予測・検証に必要な過去実績 目的変数 学習タイプ: 回帰 Index: Date

説明変数 データ型 単位 特微量選択 表示タイプ 通常

ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	日付型	-	-	-	-	-
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	233.789	85.417	229	-
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	231.899	85.814	227	-

学習設定は、画面上部のナビゲーションバーの**学習設定**をクリックすることで遷移することができます。

なお、学習設定は「目的変数」を設定しなければ表示されません。



学習設定の各コンポーネントについて説明します。

① サイドバー

作成済みの学習設定が一覧で表示されます。default設定のものには「default」のマークが表示されます。なお、新規作成は下部の「新規作成」ボタンをクリックしてください。

② ツールバー

名称の変更、コピーや、削除、設定データのエクスポートなどの操作が行えます。

③ 検証データ可視化view

交差検証、ホールドアウト検証設定(後述します)の分割状況を可視化します。

④ 設定一覧

変更可能な設定情報が表示されます。



サイドバーの各コンポーネントについて説明します。

①設定一覧

作成済みの学習設定が一覧で表示されます。default設定のものには「default」のマークが表示されます。

プリセットの学習設定には「プリセット」マークが表示されます。プリセット学習設定は削除したり、変更することができません。

②設定済みプロジェクト

当該学習設定が反映されたプロジェクト数を確認できます。クリックすると、反映済みのプロジェクトが一覧で表示され、そのプロジェクトに遷移することもできます。

③サイドバーメニューボタン

サイドバーのメニューを開くことができます。詳細は次のページで解説します。

④新規作成ボタン

学習設定を新規作成します。

名称とインポートするファイル(任意)を選択するだけで簡単に作成できます。

他の人が作った設定ファイルをインポートするだけで、簡単に同じ設定で学習することができます。

⑤メニュートグルボタン

サイドバーを格納したり、展開することができます。



<サイドバーのメニュー>

① デフォルト設定ボタン

新しいプロジェクトを作成した際に、初期設定されている学習設定を登録できます。

② コピーボタン

学習設定をコピーします。コピー後の名前を入力する必要があります。

③ 削除ボタン

学習設定を削除します。一度削除すると元には戻りません。

④ エクスポートボタン

学習設定を「amt」という拡張子でファイルとして出力します。学習設定の新規作成時に、このファイルをインポートするだ



<ツールバーの各コンポーネント>

① 名称変更ボタン

現在開いている学習設定の名称を変更します。

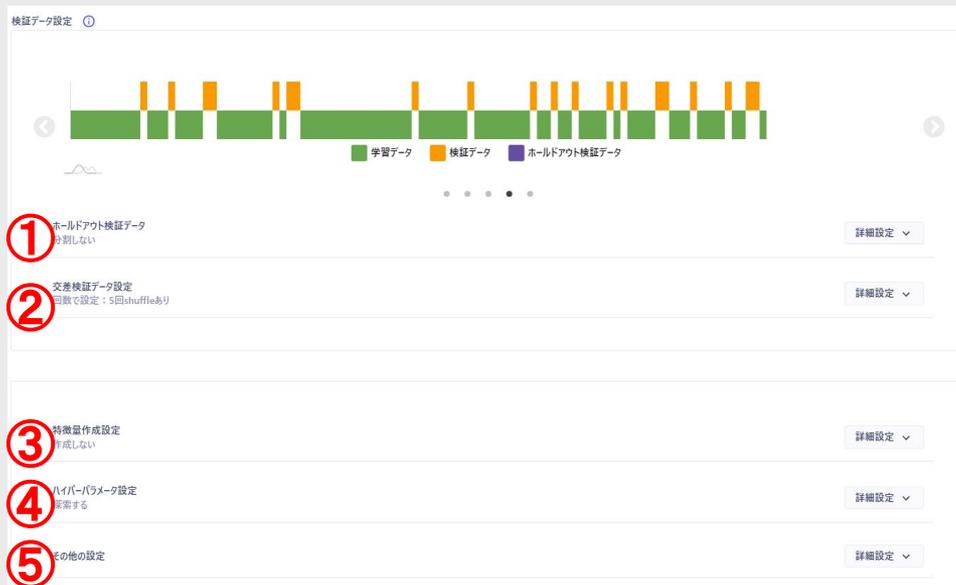
② 保存ボタン

現在開いている学習設定を保存します。保存しなければ、反映されないので注意してください。

また、すでに学習に利用中の学習設定を変更すると、既存の学習モデルは破棄されますので、ご注意ください。

③ 概要変更ボタン

現在開いている学習設定の概要を変更します。



検証データ設定

学習データ 検証データ ホールドアウト検証データ

- ① ホールドアウト検証データ
分割しない
- ② 交差検証データ設定
回数で設定: 5回shuffleあり
- ③ 特徴量作成設定
作成しない
- ④ ハイパーパラメータ設定
探索する
- ⑤ その他の設定

メイン画面の各コンポーネントについて説明します。

①ホールドアウト検証設定

学習データの一部を分割して、検証データとして使用する設定です。割合、範囲、条件等様々な設定で分割できます。

②交差検証データ設定

回数、シャッフルの有無、層化抽出等、交差検証時の設定を行います。

③特徴量作成設定

作成する特徴量をチェックボックスで簡単に設定できます。

④ハイパーパラメータ設定

ハイパーパラメータの探索範囲、方法に関する設定です。

⑤その他の設定

上記に当てはまらない設定を行います。

Home > データ読み込み > 学習設定 > AIモデル情報

飛行機乗客数予測

目的変数: Passengers 学習タイプ: 回帰 Index: Date

簡単モード
通常モード
しつかりモード

簡単モード

ホールドアウト検証データ
分割する

分割しない (手動で分割) 分割する

割合で分割

割合は 30 %

シャッフルする

同じ割合でクラスを含むように分割

範囲指定で分割

分割する範囲

条件で分割

条件を作成

条件はありません

交差検証データ設定
回数で設定: 5回 (有効)

特徴量作成設定
作成しない

保存する

新規作成

ホールドアウト検証設定について説明します。

ホールドアウト検証については「検証用データとは？」にて詳述しています。

①分割の有効化

「分割する」をONにすると、学習データの一部を分割して、ホールドアウト検証を行うようになります。

②割合で設定

学習データ全体を100として、そのうちのどのくらいを検証に使用するかを設定できます。通常設定だと、データの末尾から分割します。シャッフルを有効にすると、全体からランダムにデータを抽出します。

「同じ割合でクラスを含むように分割」は分類問題の時のみ有効です。ONにすると、目的変数のカテゴリのクラスの割合を、学習、検証データで同じ程度になるよう分割します。

③範囲指定で分割

データの行番号を指定して自由に分割できます。左のボックスのみ入力し、右を空欄にすると、その値以上の行を分割します。右のボックスのみの場合、その値以下の行を分割します。

④条件で分割

詳細な条件を作成して分割します。後述します。

条件を指定して分割 - 条件作成

特徴量の名称 **①**

データ型 **②**

値 欠損値 **③**

欠損値でない

条件 **④**

作成

条件分割について説明します。

①特徴量の名称

条件指定の対象にしたい特徴量の名称です。

学習データに当てはまる特徴量がない場合、無視されます。

②データ型

特徴量のデータ型を指定します。データ型によって、入力できる値や条件が変わります。

実際に条件を当てはめる際に、ここで設定したデータ型に変換した上で、条件抽出を行います。

詳細は次のページで説明します。

③値

条件になる値です。

「欠損値」をONにすると、当該特徴量の欠損と一致(もしくは不一致)するかどうかで判定します。

④条件

条件をプルダウンで指定します。

次のページで詳述します。

条件分割で入力、設定可能な値、条件とデータ型の関係性をまとめました。
条件は複数設定可能で、現在はand(AかつB)にのみ対応しています。

データ型	入力できる値・方式	設定できる条件
数値型	数値のみ	一致、一致しない、以上、以下、割り切れる、 割り切れない
カテゴリ型	文字列	一致、一致しない
日付型	カレンダーから、日付と時間 yyyy/MM/dd hh/mm/ssの形式	一致、一致しない、以上、以下
文字型	文字列	一致、一致しない、含む、含まない
欠損値	-	一致、一致しない

The screenshot shows the 'Cross-validation data settings' (交差検証データ設定) interface. At the top, it says 'Number of iterations set to 5 (5回shuffleあり)'. Below this, there are four settings with red circled numbers 1 through 4 pointing to them:

- ① Iterations (fold count) (回数 (fold数)): A text input field containing the number '5'.
- ② Shuffle (シャッフルする): A checked checkbox.
- ③ Split by class (同じ数だけクラスを含むように分割): A checked checkbox.
- ④ Exclude future data (未来のデータを含まないように分割): An unchecked checkbox.

交差検証設定について説明します。

交差検証については次のページで詳述します。

①回数 (fold数)

交差検証の回数を指定します。デフォルトは5回です。

②シャッフルする

交差検証時に、データをシャッフルする設定です。

一般にシャッフルする方が、より汎化性能が高まります。シャッフルしない場合は、左側からデータを分割します。

③同じ数だけクラスを含むように分割

分類問題の時、強制的に有効になります。目的変数のカテゴリのクラスの割合を、学習、検証データで同じ程度になるように分割します。

これを層化抽出といいます。

④未来のデータを含まないように分割

時系列データの時、強制的に有効になります。

通常の交差検証と違い、常に未来のデータ(画面右側のデータ)を検証データにするように分割します。

学習済みモデルが、どの程度データにfitしているかを確かめるために「ホールドアウト検証」を行うことを説明しました。

一方で、そういった未知のデータを使った検証だけでなく、学習データそのものへのモデルの当てはまりを確かめる手法として、交差検証という方法があります。

交差検証は非常にシンプルな手法です。

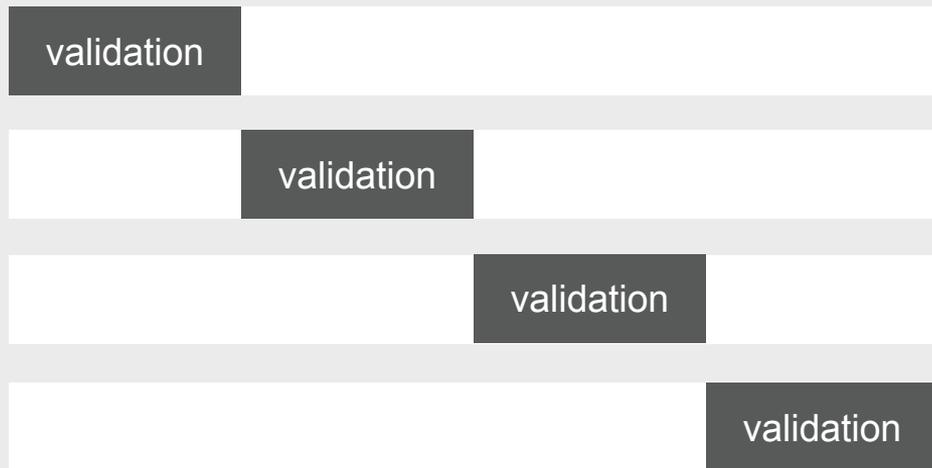
たとえば、fold4の交差検証であれば、右の図のように、データを4分割します。オレンジ色のvalidationを検証用データとし、残りのデータで学習し、構築したモデルで検証を行います。それをn回分言い、各foldのスコアを平均(ここにも様々な手法があります)し、モデルのスコアとします。

交差検証は、データ全体を偏りなく学習と検証に使うことで、より信頼性の高い精度の検証を行うことができます。

しかし、fold数を増やせば良いというわけではありません。

データが少ない場合には、あまりに多いfold数では、検証データが少なく、不当に精度が向上(あるいは低下)してしまう恐れがあります。また、大きなデータに対してfold数が少ないと、データの偏りを十分に補正できない可能性があります。

データの特徴をよく表現した最適なfold数と分割方法を決めることは、信頼性の高いモデルを構築する上で、必ず必要な工程です。データをよくみて、さまざまな手法を試してみてください。



ハイパーパラメータ設定について説明します。
ハイパーパラメータについては「ハイパーパラメータとは？」のコラムをご覧ください。

①パラメータサーチの有効化

「探索する」で、ハイパーパラメータサーチを有効にします。探索しない場合は各アルゴリズムのデフォルトパラメータで探索を行います。

②探索範囲

デフォルトの状態で学習した中で最も良い精度のアルゴリズムにのみで探索を行う「ベストモデルのみ」と、すべてのアルゴリズムで探索を行う「すべてのモデル」から選びます。

The screenshot shows the 'Hyperparameter Setting' (ハイパーパラメータ設定) interface. It includes a search button (探索する) and a 'Detailed Setting' (詳細設定) link. The 'Search' (探索) radio button is selected and circled in red with a '1'. The 'Search Range' (探索範囲) section has 'All Algorithms' (すべてのアルゴリズム) selected and circled in red with a '2'. The 'Search Method' (探索方法) section has 'Exhaustive' (総当たり) selected and circled in red with a '3'. 'Bayesian Optimization' (ベイズ最適化) is also circled in red with a '4'. The 'Number of Trials' (試行回数を決める) field is set to 10 and circled in red with a '5'. The 'Search Time' (探索時間を決める) field is set to 10 seconds and circled in red with a '6'.

③探索手法 - 総当たり

パラメータを総当たりで探索する方法です。
時間はかかりますが、精度が安定して出やすいのが特徴です。

④探索手法 - ベイズ最適化

現在の学習状況とパラメータの状況を確認しながら、より良いパラメータを探索していく手法です。

探索時間と精度のバランスが最も良いおすすめの手法です。

⑤探索手法 - 試行回数を決める

探索の回数を決めて、ベイズ最適化で探索することができます。指定できる回数はアルゴリズムごとの探索回数です。

(探索回数を 10回にして5種類のアルゴリズムで学習すると、合計で 50 回分の探索を行います。)

⑥探索手法 - 探索時間を決める

探索の時間を決めて、ベイズ最適化で探索することができます。指定できる時間はアルゴリズムごとの探索時間です。(探索時間を10秒にして5種類のアルゴリズムで学習すると、各アルゴリズムで探索時間が10秒を超えた時点で探索を終了するように学習します。最小で合計 50 秒分の探索を行います。)ただし、学習する時間が10秒を超えることもあるため、必ず指定した秒数以内に学習が終わるわけではありません。

・ハイパーパラメータとは？

アルゴリズムごとに存在する「人の手で設定しなければならない数値設定」です。
たとえばRidgeというアルゴリズムには「alpha」というパラメータがあります。
これは「データの外れ値」をどこまで重要視するかという度合いの設定です。
この設定によってモデルの振る舞いは大きく変わり、精度にも影響します。

このように「データから学ぶことができない」パラメータは、人が手で設定する必要があります。
しかしハイパーパラメータは、アルゴリズムによっては数十～数百種類にのぼり、手動で設定していくのは非常に骨が折れます。Learning Center Forecast は自動でこの値を調整し、優れたモデルを生成します。

学習設定

- その他の設定 -

その他の設定

① データ増幅 ①

増幅しない

増幅する

② 外れ値を丸め込む ①

①データ増幅

時系列分析の際に「データを増幅(データの行数を増やす)する」機能です。

データを増幅することで、少ないデータでも学習ができたり、精度が向上する可能性があります。増幅する際は、必ず「増幅する年数」を設定する必要があります。

また増幅に対応しているのは「月次」「週次」「日次」のデータのみであり、増幅には最低でも「1年分」のデータが必要です。

月次: 12行以上

週次: 52行以上

日次: 365行以上

②外れ値を丸め込む

外れ値を丸め込み、モデルの予測が外れ値によって歪められるのを防ぎます。

トレンドはあるが、値の変動が大きい場合などに、モデルの性能が向上する可能性があります。外れ値の丸め込みはアップロードしたデータのうち「数値型」のみが対象であり、それ以外のデータ型や、Learning Center Forecast 内で作成された特微量には実施されません。

AGENDA

18

作成特徴量一覧

選択方法	追加されたバージョン	計算に使用するアルゴリズム	概要
通常	ver2.6.0	LightGBM	デフォルトで設定されている特微量の選択方法です。比較的高速に動きますが、特微量を削りすぎる場合があります。
じっくり	ver2.6.0	RandomForest	「通常」に比べて動作は遅いですが、特微量を比較的好く残します。
しっかり	Ver2.8.0	Ridge, RandomForest	Learning Center Forecast 独自の選択アルゴリズムであるRPMFS (recursive permutation-mediated feature selection)を使用します。特微量作成に最も時間がかかりますが、非常に大きなデータに対しても対応しているのが特徴です。しっかりの詳細設定については次のページで説明します。

RPMFSの特徴

RPMFSは従来のLearning Center Forecast の特微量選択とは明確に違う3つの特徴があります。

1つは、巨大なデータにも対応可能ということです。

RPMFSはデータを分割して特微量選択を繰り返すため、通常学習が難しいと思われる巨大なデータでも特微量選択を実施できます。(理論上はどんな大きなデータでも可能です)

学習が困難な巨大なデータも特微量選択でデータ量を減らすことで、スムーズな学習が可能になります。

2つ目の違いは、Learning Center Forecast で作成された特微量だけでなく、学習データも削減対象として指定することができることです。

これにより学習データと作成した特微量の関係を捉えながら、選択を実施することができます。

もちろん、従来通り学習データの特微量選択を行わないことも可能です。

① 選択に利用するアルゴリズム

Ridgeのみ RandomForestのみ 両方を利用

② 学習データを選択対象に含めない

含めない

① 選択に利用するアルゴリズム

Ridgeのみ	比較的高速に動きますが、特微量を削りすぎてしまうことがあります。木形以外のアルゴリズム (SVM (SVR, SVC), Kneighbors, 線形アルゴリズム (Ridge, Lasso, ElasticNet, BayesianRidge)、ナイーブベイズ (BernoulliNB, MultinomialNB)、DeepLearning) を利用する際に、より有効なことがあります。
RandomForestのみ	特微量選択のアルゴリズムに RandomForestのみを使用します。選択に時間がかかりますが、Ridgeよりも良い特微量を選択することができます。木形のアルゴリズム (RandomForest, AdaBoost, GradientBoosting, XGBoost, LightGBM) を利用する際に、より有効なことがあります。
両方を利用	特微量選択のアルゴリズムに RidgeとRandomForestの両方を利用します。木形のアルゴリズムには RandomForestでの選択結果を利用します。木形以外のアルゴリズムには Ridgeでの選択結果を利用します。アルゴリズムに合わせた選択を行うので、より柔軟に特微量を選択できる可能性があります。

② 学習データを選択対象に含めない

含めない	この設定を ON にすると、RPMFS を実施する際に学習データは含めずに選択を行います。(Learning Center Forecast で作成された特微量のみを選択対象にします) この設定を OFF にすると、RPMFS を実施する際に学習データも含めて選択を実施します。(学習データも削られてしまい、学習に使われない可能性があります) デフォルトは OFF です。
------	---

時系列:周期

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
ラグ	ラグ {n}	通常としっかりモード(時系列のみ)	ver2.1.8	数値	n単位前の目的変数	遅れて影響を与える値を表現します
差分	差分 {n}	通常としっかりモード(時系列のみ)	ver2.1.8	数値	目的変数 - n単位前の目的変数	トレンドの効果を除いた値を表現します
微分	微分 {n}	通常としっかりモード(時系列のみ)	ver2.1.8	数値	n単位前の目的変数 - (n + 1)単位前の目的変数	変化量を表現し、変曲点を見つけやすくします
移動平均	移動平均 {n}	通常としっかりモード(時系列のみ)	ver2.1.8	数値	n単位前までの目的変数の平均	値の変化を平滑化し、トレンドを捉えやすくします
誤差	誤差 {n}_{m}	通常としっかりモード(時系列のみ)	ver2.1.8	数値	ラグ {n} - 移動平均 {m}	移動平均からの差の推移をとらえることで、周期性のあるデータにおいて予測をしやすくします
分散	分散 {n}	通常としっかりモード(時系列のみ)	ver2.3	数値	n単位前までの目的変数の分散	直近の値の変化の大きさを表現します

時系列:周期(説明変数)

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
ラグ	ラグ{n}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	n単位前の目的変数	遅れて影響を与える値を表現します
差分	差分{n}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	目的変数 - n単位前の目的変数	トレンドの効果を除いた値を表現します
微分	微分{n}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	n単位前の目的変数 - (n + 1)単位前の目的変数	変化量を表現し、変曲点を見つけやすくします
移動平均	移動平均{n}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	n単位前までの目的変数の平均	値の変化を平滑化し、トレンドを捉えやすくします
誤差	誤差{n}_{m}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	ラグ{n} - 移動平均{m}	移動平均からの差の推移をとらえることで、周期性のあるデータにおいて予測をしやすくします
分散	分散{n}_{特微量名}	しっかりモード (時系列のみ)	ver2.6.0	数値	n単位前までの目的変数の分散	直近の値の変化の大きさを表現します

時系列: 日付

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
イベント	イベント: {イベント名}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	時系列設定で作成したイベント特微量の値です。設定した影響力、影響範囲によって値が変化します。影響範囲は、イベントより前に影響がある場合、影響開始地点から、イベント当日に向かって影響力が増大します。イベントより後に影響がある場合、影響終了地点に向かって線形に、影響力が減衰します。
年	日付分解:年_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の年の値を抽出したものです。nは当該の年を表現しています。(ex: 年_2021 は 2021年を表現します)
月	日付分解:月_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の月の値を抽出したものです。nは当該の月を表現しています。(ex: 月_1 は 1月を表現します)
週	日付分解:週_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の週の値(1年のうちの何週目か)を抽出したものです。nは1年の中の週の位置を表現しています。(ex: 週_10 は その年の10週目を表現します)
日	日付分解:日_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の日の値を抽出したものです。nは当該の月の日にちを表現しています。(ex: 日_24 は その月の24日を表現します)
時	日付分解:時_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の時間(hour)の値を抽出したものです。nは当該の日の時間(24時間表記)を表現しています。(ex: 時_20 は その日の20時を表現します)
分	日付分解:分_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の分(minutes)の値を抽出したものです。nは当該の時間の分を表現しています。(ex: 分_30 は 30分を表現します)
秒	日付分解:秒_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の秒(seconds)を抽出したものです。nは当該の分の秒を表現しています。(ex: 秒_45 は 45秒を表現します)
曜日	日付分解:曜日_{n}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の曜日を抽出したものです。nは曜日を表す数値で、日曜日 =1 として、曜日を追うごとに1ずつ加算され、土曜日 =7 の数値で表現されています。
週末	週末	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の週末(土、日)のフラグです
月末	月末	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日付特微量の月末(月の最終日)のフラグです

時系列:日付

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
祝日	祝日:{祝日名}	通常としっかりモード(時系列のみ)	ver2.1.8	日付	-	日本の祝日を、祝日ごとにフラグとして作成します。
調整済み祝日	調整済み祝日: {祝日名}	通常としっかりモード(時系列のみ)	ver2.6.0	日付	-	日本の祝日を、祝日ごとにフラグとして作成します。 祝日の発生日を「予測 GAP」分調整します。 「n 単位後に発生する祝日」として学習することができます。
調整済み祝日フラグ	調整済み祝日フラグ	通常としっかりモード(時系列のみ)	ver2.6.0	日付	-	日本の祝日をフラグとした特微量を作成します。
調整済みイベント	調整済みイベント: {イベント名}	通常としっかりモード(時系列のみ)	ver2.6.0	日付	-	詳細は「学習設定 -特微量作成設定 -時系列-日付」ページの「イベント」を確認してください。 イベントの発生日を「予測 GAP」分調整します。 「n 単位後に発生するイベント」として学習することができます。
調整済み月末	調整済み月末	通常としっかりモード(時系列のみ)	ver2.6.0	日付	-	詳細は「学習設定 -特微量作成設定 -時系列-日付」ページの「月末」を確認してください。 月末を「予測 GAP」分調整します。 「n 単位後の月末」として学習することができます。

日付(説明変数)

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
日付	{特微量名}_日付分解:{日付単位}_{n}	-	ver2.6.0	日付(説明変数)	-	日付の説明変数を分解し、7種類の日付単位(年、月、週、日、時、分、秒)の特微量を作ります。(各日付単位の詳細は「学習設定 -特微量作成設定 -時系列-日付」ページにあります)

数値間

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
交互作用(2次)	{A} × {B}	しっかりモード	ver2.1.8	数値	ある行における特微量 Aの値と特微量Bの値の積	数値特微量間の影響を表現します。
和	{A} + {B}	しっかりモード	ver2.1.8	数値	ある行における特微量 Aの値と特微量Bの値の和	数値特微量間の影響を表現します。
比	{A} / {B}	しっかりモード	ver2.1.8	数値	ある行における特微量 Aの値と特微量Bの値の商	数値特微量間の影響を表現します。
差	{A} - {B}	しっかりモード	ver2.1.8	数値	ある行における特微量 Aの値と特微量Bの値の差	数値特微量間の影響を表現します。

カテゴリ間

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
LDAによるトピック分析	トピック分析_{n}	しっかりモード	ver2.1.8	カテゴリ	-	カテゴリ間の共起行列から共通のトピックを想定し、カテゴリ特微量同士の関係性を表現します。 各カテゴリ間でトピックが3つあるという仮定で特微量を作成するため、"カテゴリ特微量の数 * (カテゴリ特微量の数 - 1) * 3"個の特微量が作成されます。

文字列

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
形態素解析 (bug of words)	{A}_{文字列}	通常としっかりモード	ver2.1.8	文字列	-	文字列から「名詞」「動詞」「形容詞」を抽出し、ベクトル化します。特に予測に与える影響が大きかった値を最大50列まで学習に用います。

トランザクションデータ

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
前回の値	{A} トランザクション 前回の値	-	ver2.3	カテゴリ,数値	Indexの、ある値の、特微量 A における、直近の値	直近の値を取得することで、データ全体の傾向をつかみます
統計	{A} トランザクション 統計 {x}	-	Ver2.3	カテゴリ,数値	Indexの、ある値の、特微量Aにおける、統計量 x (平均、最大、最小、合計、標準偏差、中央値)	Indexのある値の特徴を抽出します

その他

特微量名	特微量重要度での表示	作成されるモード	追加されたバージョン	作成元の特微量の型	計算方法	概要
行の統計量 (欠損)	行の欠損の数	通常としっかりモード	ver2.1.8	-	ある行における欠損の数の合計値	行ごとの統計量を取ることで、特定の特微量に寄らない外れ値やデータ間の影響を表現します。
行の統計量 (負の値)	行の負の値の数	通常としっかりモード	ver2.1.8	-	ある行における0より小さい値の特微量の合計値	行ごとの統計量を取ることで、特定の特微量に寄らない外れ値やデータ間の影響を表現します。

AGENDA

19

AIモデル構築

予測する対象と、学習に使用する特徴量が決まったら、いよいよAIによる学習を始めていきます。

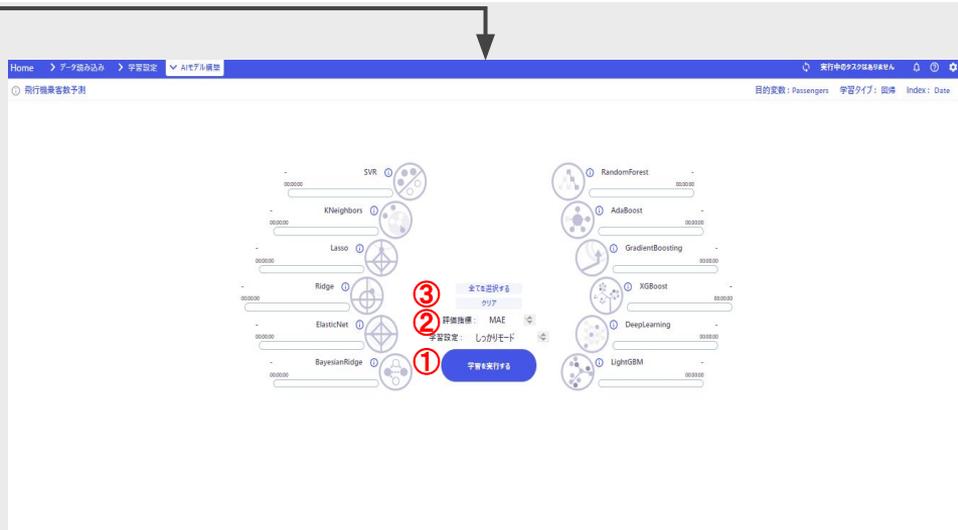
Learning Center Forecast では驚くほど簡単に、AIによるモデル構築が可能です。



ID	名称	目的変数	Index	説明変数	データ型	単位	平均	標準偏差	中央値	最頻値
1	Date	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	日付型	-	-	-	-	-
2	Passengers	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	238.009	85.621	229	-
3	Passengers_L1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	数値	-	236.028	85.551	229	-
4	Passengers_L2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	233.789	85.417	229	-
5	Passengers_L3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	数値	-	231.899	85.814	227	-

目的変数を設定すると、「AIモデル構築」画面に遷移できるようになります。

画面上部のパンくずリストの「AIモデル構築」をクリックしてください。



AIモデル構築では12～14種類のアルゴリズムを使って、最適なモデルを構築することができます。

- ①画面中央の「全てを選択する」ボタンを押すと、すべてのアルゴリズムが選択されます。
- ②評価指標の下にあるプルダウンは「学習設定」です。学習設定については「学習設定」で詳述しています。
- ③中央の評価指標ボタンで、評価指標を変更することができます。評価指標とは、モデルの性能を評価するための基準のことです。

数値予測はデフォルトではMAE(平均絶対誤差)としており、これは予測値と実績値の誤差の絶対値の平均だと考えていただいて問題ありません。



特徴量の作成が完了した時、左の様なInformationが表示されることがあります。

これは時系列データの特徴量であるラグなどの周期を持ったデータ独自のものです。

ラグデータとは、ある変数を任意の数、行方向にずらしたデータのことです。データが連続的に変化する、周期性を持っている時、適切なラグデータは予測に大きな効果を発揮します。

たとえば、①のようなデータがあるとして、 b_j 列が目的変数だとします。 a 列について $\text{lag}=1$ と $\text{lag}=2$ のデータを作成してみます。

ラグを作成すると、②のようになります。

データをずらしたことによって、 $aL1$ と $aL2$ にNan(欠損)ができてしまっています。時系列データでは、未来のデータを先取りすることはできず、これ以上過去のデータもありません。そのため、ラグで生じたNanについては、下手に補完するのではなく、削除する方が無難です。

削除して、最終的には③のようになります。

このように、ラグデータは場合によっては非常に有効ですが、データの元の数が少ない場合は貴重な学習データを減らしてしまう危険性があります。また検証データや予測用データも同数のラグを作るだけの余裕を持ったデータを用意する必要があります。

①

a	b	c
1	2	3
2	4	6
3	6	9



②

a	aL1	aL2	b	c
1	Nan	Nan	2	3
2	1	Nan	4	6
3	2	1	6	9

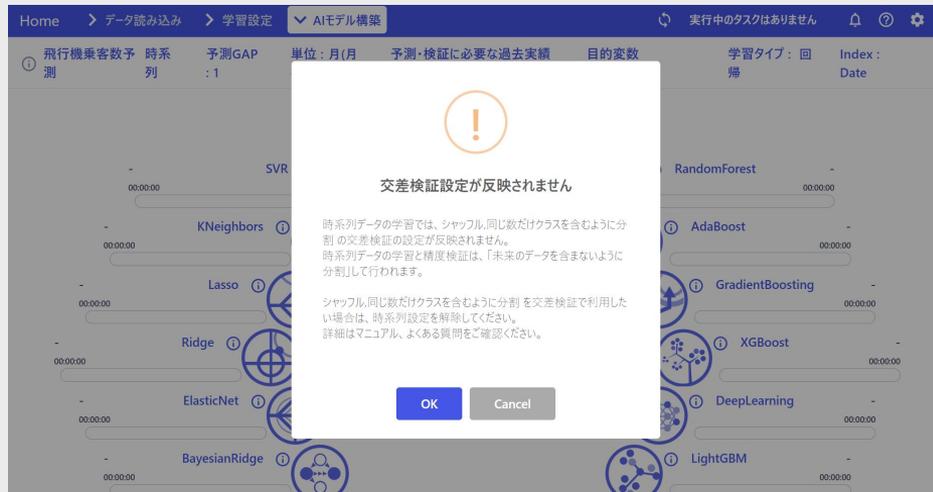


③

a	aL1	aL2	b	c
3	2	1	6	9



アルゴリズムを一つ以上選択し、中央にある**学習を実行する**ボタンを押すと、学習が始まります。



時系列データを学習する場合、学習開始時に、このような警告が表示されることがあります。

長い警告文ですが「時系列データの場合、交差検証は(交差検証については「交差検証とは?」で詳述します)のやり方を変更することができなくなります」という意味のメッセージが表示されています。

基本的には気にしなくても問題ありません。

ただ、交差検証や、学習設定を詳細に知りたい方は次のページの解説をご覧ください。

時系列データのための交差検証とは？

時系列データには、通常のデータ(時間の影響が少ないと考えられるデータ)とは違う特徴があります。

特に重要な特徴は「予測(検証)データよりも未来の情報は手に入らない」という前提です。

そのため、交差検証時のシャッフルや層化抽出を行うことができません。

(正確には行うこと自体はできますが、そのモデルの信頼性は低いものと考えられます)

Learning Center Forecast では、交差検証も時系列データ専用の方法を取るようになっています。

それが「Time Series Cross Validation (TSCV)」です。

右図はTSCVの模式図です。

青い部分が学習データ、グレーの部分が検証データです。

データの向かって左側がより古い時間、右側がより新しい時間(違和感のある言い方ですがお許しください)です。

学習データは、常に検証データよりも古い時間(検証データより過去)で学習しているのが確認できるかと思います。

このようにすることで「未来」のデータからのleakを防ぎ、直感に近い精度を表現することができます。

過去

未来





学習の暫定のベストスコアは、中央の「ベストスコア」で確認できます。

この場合だと「ElasticNet」というアルゴリズムがもっとも性能が良かったことがわかります。

アルゴリズムを選択し直せば、何度も学習をすることができます。

AIモデル構築は以上で完了です。

学習中は、ナビゲーションバーのタスクウィンドウで進捗を確認することができます。

学習が進むほど、モデルの精度は上がっていきます。

アルゴリズムアイコンの横に表示される数値が「精度」です。

この場合はMAEですので、「予測値と実績値の誤差の絶対値の平均」ということになります。

つまり、誤差が小さいほど精度が高いため、この値が小さいものほど優れたモデルだと判断することができます。



AGENDA

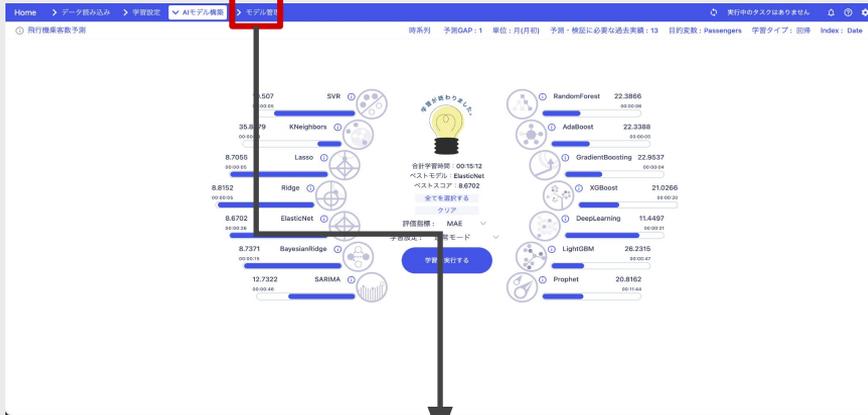
20

モデル検証

学習しただけでは、モデルを運用することは困難です。
なぜなら、そのモデルの精度を客観的に測れていないためです。
モデルの性能をできる限り客観的に計測するために行われるのが「**ホールドアウト検証**」です。ホールドアウト検証では、**学習をする前にあらかじめデータを分割し、一方を学習用に、一方を検証用に**使用します。

検証を行っていないモデルは、試合を一度もしたことがないスポーツ選手のようなものです。練習の結果を、練習試合で試すことで、初めて実戦での実力が向上したかどうか確認できます。

AIモデルも同じで、「**未知の検証データ**」を予測させることで、実際に運用に耐えうる精度が出せるのか確認する必要があります。



AIモデル構築が完了した後は、モデルの未知のデータへの精度を確認するために「モデル検証」を行う必要があります。

画面上部の「モデル管理」をクリックしてください。

飛行機乗客数予測

時系列 予測GAP: 1 単位: 月(月初) 予測-検証に必要な過去実績: 13 目的変数: Passengers 学習タイプ: 回帰 Index: Date

詳細を確認したいアルゴリズムをクリックしてください

アクション

アルゴリズム名	学習時間	学習精度
ElasticNet	0.03	8.704
Lasso	0.03	8.766
Ridge	0.03	9.186
BayesianRidge	0.04	9.324
SVR	0.04	10.041
DeepLearning	0.68	10.695
SARIMA	0.41	12.732
XGBoost	1.45	19.767
Prophet	69.02	20.816
GradientBoosting	0.39	22.794

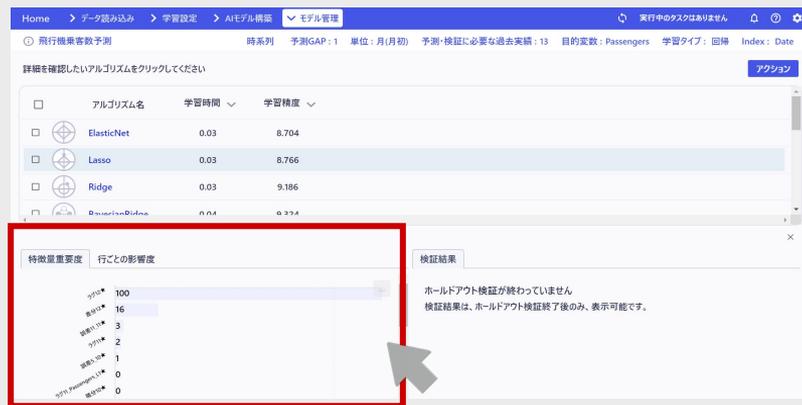
モデル管理画面では、各モデルの学習時のスコアと、学習時間を一覧で見ることができます。

検証前の状態では、学習スコアが高い順番に上から並んでいます。もしスコアが同一の場合は、学習時間が短いモデルが上になります。

モデルをクリックすると、モデルについての詳細を見ることができます。



右上にあるボタンからグラフを画像やデータとしてダウンロードすることができます。



左側のウィンドウは、特微量重要度です。

特微量重要度とは、モデルがデータを予測する際、最も影響力が大きかった特微量を100とした時に他の特微量の影響力がどの程度あるか、相対的に表したものです。

たとえば、左の画面からは、最も影響力のある特微量に対して、影響力のない特微量は低い割合でしか予測に寄与していないことがわかります。

特微量重要度は各モデルごとに作成できます。

左の画面からは「最も影響力がある特微量が、二番目に影響力がある特微量の約6倍結果に影響する」ということがわかります。

多くのモデルにおいて、上位に位置した特微量は、学習に十分に有用な特微量であるともいえるでしょう。



それでは、検証を始めていきましょう。
右上の「アクション」ボタンから「ホールドアウト検証」を選択してください。



検証に使用する「検証データ」を選択します。

検証データはあらかじめ用意しておく必要があり、かつ「実績値」が入ったデータである必要があります。

多くの場合、学習用に用意したデータの一部を切り分けておき、検証用データとして残しておきます。

次のページに検証用データについて簡単にまとめました。

検証用データとは？

検証用データとは、モデルの性能をより客観的に図るために行われる**ホールドアウト検証**で使用するデータのことで、
検証用データは、**必ず実績値**が入っていなければいけません。

学習だけでは、モデルの性能はわかりません。

学習に使用したデータを「過学習」してしまい、却って未知のデータへの対応力が失われてしまうことがあるからです。

そこでモデルの「未知のデータ」への**対応力=汎化性能**を測るため、あらかじめ学習用に用意した実績値入りのデータを分割しておき、一部を学習用に、一部を検証用に使用します。

学習するデータ量が減ってしまうため、あまり多くを割くことはできませんが、性能を正しく評価できなければ、結局モデルは役に立ちません。

こうしたことに対応するためにも、用意するデータは多ければ多いほど良いと言えます。

用意した実績値入りのデータ



学習用データ

検証用データ

データを選択したら、アップロードします。



時系列データの場合



アップロードが完了すると、「ホールドアウト検証を開始する」というボタンが表示されます。このボタンを押すと、検証が始まります。

時系列データは、過去のデータから未来を予測します。また、特徴量の作成にも過去のデータを利用します。

そのため、「検証したいデータ」に加えて「検証期間を予測するためのデータ」と「特徴量作成を行うためのデータ」を用意する必要があります。

画面のように、Learning Center Forecast が必要なデータの期間を表示してくれます。「ファイル選択」でデータを選択してください。

ホーム > データ読み込み > 学習設定 > AIモデル構築 > モデル管理

飛行機乗客数予測 時系列 予測GAP: 1 単位: 月(月初) 予測・検証に必要な過去実績: 13 目的変数: Passengers 学習タイプ: 回帰 Index: Date

詳細を確認したいアルゴリズムをクリックしてください

アルゴリズム名	学習時間	学習精度
ElasticNet	0.03	8.704
Lasso	0.03	8.766
Ridge	0.03	9.186
BayesianRidge	0.04	9.324
SVR	0.04	10.041
DeepLearning	0.68	10.695
SARIMA	0.41	12.732
XGBoost	1.45	19.767
Prophet	69.02	20.816

ホールドアウトデータ検証

検証データのアップロード

検証データ.csv 削除する

追加データのアップロード

追加データを検証しています...

ホールドアウト検証を開始する

ホールドアウト検証
検証結果のダウンロード
行ごとの影響度の計算
行ごとの影響度のダウンロード
モデルのデプロイ
モデルの削除

追加データをアップロードすると、データの検証が始まります。

しばらくお待ちください。

不適切なデータの場合（指定された期間を含んでいない、Indexが存在しない等）、ホールドアウト検証を行うことができません。

適切な追加データを再度選択して、アップロードしてください。

ホーム > データ読み込み > 学習設定 > AIモデル構築 > モデル管理

飛行機乗客数予測 時系列 予測GAP: 1 単位: 月(月初) 予測・検証に必要な過去実績: 13 目的変数: Passengers 学習タイプ: 回帰 Index: Date

詳細を確認したいアルゴリズムをクリックしてください

アルゴリズム名	学習時間	学習精度
ElasticNet	0.03	8.704
Lasso	0.03	8.766
Ridge	0.03	9.186
BayesianRidge	0.04	9.324
SVR	0.04	10.041
DeepLearning	0.68	10.695
SARIMA	0.41	12.732
XGBoost	1.45	19.767
Prophet	69.02	20.816

ホールドアウトデータ検証

検証データのアップロード

検証データ.csv 削除する

追加データのアップロード

適切な追加データが設定されました。
いつでもホールドアウト検証を開始できます。

検証追加データ.csv 削除する

ホールドアウト検証を開始する

ホールドアウト検証
検証結果のダウンロード
行ごとの影響度の計算
行ごとの影響度のダウンロード
モデルのデプロイ
モデルの削除

追加データの検証が正常に終わると、左のような画面になります。

「ホールドアウト検証を開始する」をクリックしてください。

アルゴリズム名	学習時間	学習精度	MAE	MAPE	RMSE	RMSPE
Decision	0.01	8.766				
Linear	0.01	8.766				
Ada	0.01	8.766				
DecisionTree	0.04	8.766				
XG	0.04	8.766				
DeepLearning	0.04	8.766				
SARIMA	0.04	8.766				
GBout	1.46	16.927				
Prophet	0.02	20.816				
GradientBoosting	0.39	23.774				

検証中はモデルの詳細を確認することができません。しばらくお待ちください。

それぞれの指標の右横にあるボタンをクリックすると、ソートをかけることができます。

アクションボタンから「検証結果のダウンロード」をクリックしてください。

ダウンロードデータの構成に少し時間がかかるので、少々お待ちください。

Home > データ読み込み > 学習設定 > AIモデル構築 > モデル管理

飛行機乗客数予測 時系列 予測GAP: 1 単位: 月(月初) 予測・検証に必要な過去実績: 12ヶ月 変数: Passengers 学習タイプ: 回帰 Index: Date

詳細を確認したいアルゴリズムをクリックしてください

<input checked="" type="checkbox"/>	アルゴリズム名	学習時間	学習精度	MAE	MAPE	RMSE	RMSPE	アクション
<input checked="" type="checkbox"/>	Lasso	0.03	8.766	12.752	3.083	17.347	4.100	ホルダアウト検証 検証結果のダウンロード 行ごとの影響度の計算 行ごとの影響度のダウンロード モデルのデプロイ モデルの削除
<input checked="" type="checkbox"/>	AdaBoost	0.03	8.704	13.002	3.152	17.586	4.169	
<input checked="" type="checkbox"/>	RandomForest	0.03	9.186	14.418	3.532	18.680	4.493	
<input checked="" type="checkbox"/>	XGBoost	0.04	10.041	14.443	3.574	18.891	4.592	91.04%
<input checked="" type="checkbox"/>	AdaBoost	0.04	9.324	14.500	3.556	18.756	4.511	91.17%
<input checked="" type="checkbox"/>	DecisionTree	69.02	20.816	16.337	4.126	20.056	5.179	89.9%
<input checked="" type="checkbox"/>	GradientBoosting	0.68	10.695	23.559	5.453	28.788	6.561	79.19%
<input checked="" type="checkbox"/>	XGBoost	1.45	19.767	31.623	7.234	39.164	8.521	61.49%
<input checked="" type="checkbox"/>	AdaBoost	0.47	22.935	33.807	7.733	41.167	9.056	57.46%
<input checked="" type="checkbox"/>	RandomForest	0.75	23.778	35.057	7.866	43.587	9.128	52.31%

検証結果はダウンロードすることができます。

結果を知りたいモデルの横のチェックボックスにチェックを入れてください。一番上のチェックボックスをクリックすると、すべてのモデルを選択することができます。

検証終了

Date	Passengers	KNights	LGBMReg	RandomForest	AdaBoost	GradientBoosting	Prophet	XGBoost	SARIMA	MLPhog	GB	BayesianRidge	Lasso	ElasticNet
1958/5/1	369	351.5	352.3889	278.21	361.8029	357.4916	291.2249	333.172	355.3711	367.7403	359.3143	364.1821	364.1821	364.0466
1958/6/1	435	388.5	393.4142	413.27	417.5	417.0477	441.7107	442.282	428.4999	455.4895	441.5988	443.7718	444.1975	444.7121
1958/7/1	491	443.5	429.4131	446.17	465	452.4141	501.1243	495.7034	475.8202	472.9341	491.2581	491.4796	491.0892	491.8096
1958/8/1	525	466	444.112	493.78	485	492.9722	491.1433	495.7494	477.8887	483.1346	503.2966	501.7466	500.9867	498.7879
1958/9/1	404	435.5	422.0979	427.96	422.7	436.0518	433.8703	466.2238	469.0253	445.5068	447.8763	445.1492	445.4978	438.9314
1958/10/1	359	286.5	291.1161	268.87	257.4	262.6242	277.1472	262.9719	246.5216	231.4209	244.1716	246.2167	246.5089	244.8442
1958/11/1	310	298	332.0075	349.99	352.3731	334.9034	328.8828	325.1676	306.8431	328.8529	325.5247	311.713	325.5912	325.5369
1959/1/1	337	321	339.1945	314.74	324.0	324.172	373.0171	347.0473	334.5898	332.7921	348.0347	345.3029	345.5001	344.0729
1959/2/1	360	377.5	342.2226	353.72	354.1176	338.0788	380.0888	345.1204	338.669	338.3063	353.5122	362.2663	345.4485	345.2889
1959/3/1	342	305.5	345.1873	342.51	352.5556	346.06	370.3338	340.6486	314.8897	336.5957	332.8781	333.6299	334.4553	340.7125
1959/4/1	406	359	371.6742	380.93	362.1818	373.0316	418.5008	370.9469	363.0475	379.838	370.688	374.7274	374.2294	384.748
1959/5/1	396	348	388.7427	364.39	356.7778	337.2096	400.114	384.1643	347.2146	385.002	383.8978	382.6697	381.3285	381.3512
1959/6/1	420	351.5	393.4484	390.12	403.25	373.7945	430.6141	379.0155	356.2146	409.8217	410.4631	409.4452	409.842	411.963
1959/7/1	472	443.5	427.3828	421.58	422	426.8783	487.2087	437.7243	435.3841	478.9149	487.8493	491.3853	491.7987	495.754
1959/8/1	548	465	443.5588	446.12	465	451.6556	503.0064	464.8187	467.8338	488.8147	520.2193	526.5719	526.0004	527.5176
1959/9/1	559	435.5	445.5988	450.98	465.6667	451.5885	513.7838	466.5199	489.2277	506.3619	559.8018	559.0595	558.0008	555.6228
1959/10/1	460	435.5	397.648	431.3	426.4286	413.4331	474.2214	454.2083	414.5161	454.7921	452.7395	452.5893	451.4964	451.3534
1959/11/1	427	375.5	375.2068	380.05	362.1818	388.4668	411.1188	372.0905	346.2527	378.1218	384.5443	384.2969	384.8606	404.4388
20/1/1	362	320	347.3553	343.81	352.3731	343.6002	356.154	337.7927	295.915	344.5097	344.3914	344.3163	344.9636	350.0462

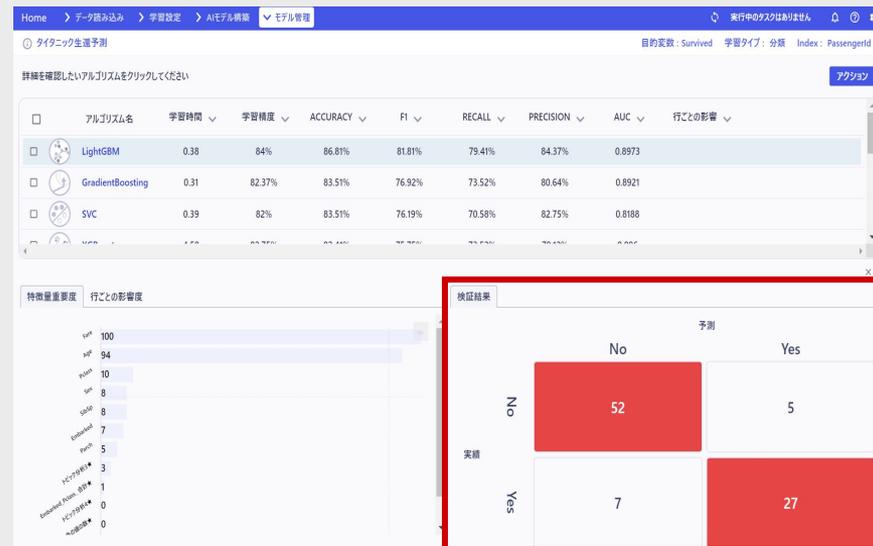
検証結果は左のような形でダウンロードされます。

Indexが指定されていれば、Indexが一番左に配置されます。その横に実績データ、そして各モデル(アルゴリズム名)による予測結果が並んでいます。



検証結果は、数値予測の場合、左のような予測と実績の折れ線グラフが表示されます。

濃い色が実績値、薄い色が予測値です。



クラス分類の場合、右のような、混同行列が表示されます。

モデル検証機能の説明は以上です。

AGENDA

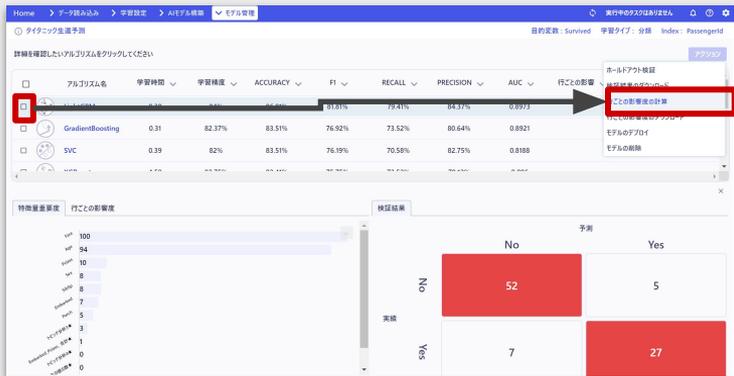
21

行ごとの影響度の計算

機械学習のモデルが予測した結果は、一般には解釈が難しいと考えられています。特徴量重要度などである程度の傾向は確認できるものの「予測値」1つ1つに対して、「なぜこのような予測を行なったのか」を解釈するのは非常に難しいです。

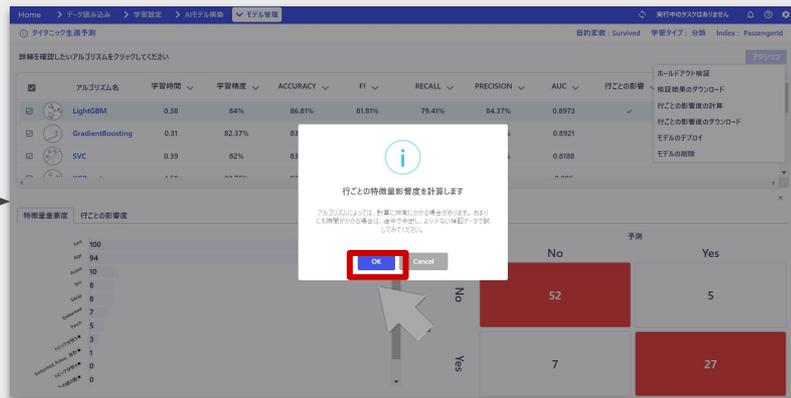
「**行ごとの影響度の計算**」は、予測値1つ1つに対して、それぞれの特徴量がどの程度予測に寄与したのかを表示する機能です。

②画面右上の「アクション」ボタンをクリックし、上から3つ目の「行ごとの影響度の計算」をクリックしてください。



①「行ごとの影響度」を計算したいアルゴリズムをチェックします。アルゴリズム名の左にあるチェックボックスをクリックします。

最上段にあるチェックボックスをクリックすることで、すべてのアルゴリズムを指定できます。



確認のためのポップアップが表示されます。この案内にある通り、一部のアルゴリズムでは行ごとの影響度の計算に非常に時間がかかります。(次のページ参照)

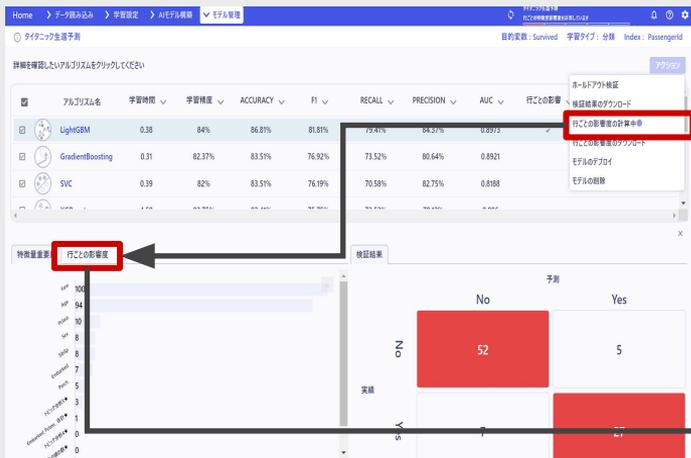
- ・SVM(サポートベクターマシン)
- ・Kneighbors(k最近傍法)
- ・ナイーブベイズ系(BernoulliNB MultinomialNB)
- ・AdaBoost
- ・DeepLearning

これらのアルゴリズムで計算する場合は、十分な時間を確保した上で、もしくは、データ数を絞った上でお試しください。
 また上記のアルゴリズムは、「クラス分類」で、今のバージョンではデータの列数が100列より多い場合は計算できませんのでご注意ください。

- ・SVM(サポートベクターマシン)
- ・Kneighbors(k最近傍法)
- ・ナイーブベイズ系(BernoulliNB MultinomialNB)
- ・AdaBoost
- ・DeepLearning

左で紹介された以下のアルゴリズムで行ごとの影響度を計算すると、稀に「値にゆらぎ」が生じることがあります。
 「値のゆらぎ」とは、計算のたびに、計算結果にほんの少しズレが発生する現象のことです。

値のゆらぎは小数点第三以下にとどまり、意思決定に大きな影響を与えるものではありません。



①計算が開始されると、「行ごとの影響度の計算中」という文言に変わります。

計算が終わるまで気長にお待ちください。

②行ごとの影響度を確認したいモデルをクリックして選択後、画面下部のモデルの詳細タブにある「行ごとの影響度」をクリックしてください。



画面が変わり、影響度が確認できるようになります。

画面では「タイタニック号」(サンプルデータとしてLearning Center Forecast に最初から実装されています)のデータを使用しています。

乗客ごとに「何が影響して助かったのか」のモデルによる解釈が確認できます。この一覧画面では、影響度が高かった上位3つの特徴量と、その影響度を確認できます。影響度の詳細を確認したい場合は、クリックします。



詳細画面では、すべての特徴量についての影響度が確認できるようになります。

この例では、この方はかなり生き残る確率が高いだろうと予測されています。その理由として「Sex(性別)」が非常に大きな影響を与えたことがわかります。タイタニック号の事故では、女性や子供が優先的に救命ボートに乗せるような動きがあったのでしょうか。

また「Pclass(等級)」も影響が大きかったようです。当時の船は身分によって「等級」が分かれており、等級の高い人ほど生き残る傾向があったようです。

画面上部、左にある「基準値」は、モデルの予測値の平均です。この値から、各特徴量がどのような影響を与えて、「予測値」になったのかが確認できます。

今回は分類データなので、影響度は百分率で表示されています。回帰予測の場合は、影響度は変動した値そのもので表示されます。(例えば株価の予測であれば、ある特徴量が株価をいくら変動させたか、という形で確認できます)

また、画面上部の矢印を押すことで、他の行についての影響度を確認できます。左の矢印をクリックすると、一つ前のIndexに戻ります。右の矢印をクリックすると、一つ先のIndexに進みます。



The screenshot shows the 'モデル結果' (Model Results) page. At the top, there are tabs for 'Home', 'データ読み込み', '学習設定', 'AIモデル一覧', and 'モデル結果'. Below this, there's a navigation bar with 'データチェックをダウンロード' and '目的変数: Survived', '学習タイプ: 分類', 'Index: PassengerId'. A table lists models: LightGBM, GradientBoosting, and SVC. The '行ごとの影響度' (Feature Importance) column has a dropdown menu with options: 'ロードアウト検証', '検証結果のダウンロード', '行ごとの影響度のダウンロード', and 'モデルのプロパティ'. A grey arrow points to the '行ごとの影響度のダウンロード' option. Below the table, there's a '検証結果' (Validation Results) section with a confusion matrix for '予測' (Prediction) vs '実観値' (Actual Value).

	実観値: No	実観値: Yes
予測: No	52	5
予測: Yes	7	27

行ごとの影響度はダウンロードすることができます。ダウンロードしたいアルゴリズムを選択後、右上の「アクション」ボタンから、「行ごとの影響度のダウンロード」をクリックして選択してください。ダウンロードには少し時間がかかりますので、気長にお待ちください。

The screenshot shows a data table with columns labeled A through W. The first few rows contain numerical values representing feature importance. A grey arrow points to the 'shop' column in the last row.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W				
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
3	0.1874	0.0709	-0.0022	0.0183	-0.0207	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
4	0.1881	0.1389	0.0178	0.0382	0.1641	0	0	0	0	0.0000	-0.3956	-0.018	-0.3939	-0.3842	-0.1114	0.0334	-0.0773	-0.1009	0	0.0000	-1.0613	-0.1023	0	0.0000			
5	-0.0636	0.5949	0.0021	0.0384	-0.1326	0	0	0	0.0000	0	0.0000	-0.1458	0.07	-0.002	0.0533	0.0011	-0.0006	-0.0007	0.0002	0	0.0000	-0.0001	0.0003	0	0.0000		
6	-0.0698	0.0122	0.0087	0.0484	-0.1202	0	0	0	0.0000	0	0.0000	-0.1551	-0.0623	-0.0629	-0.1848	0.0075	0.0120	-0.1113	0.0004	0	0.0000	0.0002	0.0004	0	0.0000		
7	-0.0852	0.1946	0.0161	0.0369	0.0562	0	0	0	0.0000	0	0.0000	-0.3934	-0.3926	-0.0953	-0.3707	0.0324	0.0080	0.0142	-0.41	-0.0003	0	-0.0002	0.0003	0.0005	0	0.0000	
8	0.1188	0.1483	0.0102	0.0304	-0.2048	0	0	0	0.0000	0	0.0000	-0.3834	-0.1576	-0.0346	-0.3844	-0.1443	-0.1829	0.0324	-0.1305	-0.0005	0	-0.0001	0.0004	0.0004	0	0.0000	
9	-0.2283	-0.0111	0.0027	0.0374	0.0996	0	0	0	0	0.0000	0.0000	0.3943	-0.0101	-0.2451	0.1096	0.0133	0.0000	-0.0007	0	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
10	0	0.0000	0.0000	0.0000	0.1221	0	0	0	0.0000	0	0.0000	-0.2418	-0.0116	-0.0402	0.0122	0.0015	0.0007	-0.0009	-0.0008	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
11	0.0006	0.0442	0.0059	0.0092	0.1301	0	0	0.0000	0	0.0000	0.0000	0.3909	0.0398	0.0382	0.0128	-0.0096	0.0127	0.0242	0.0005	0	0.0000	0.0004	0.0004	0.0004	0	0.0000	
12	-0.0542	-0.0002	0.0006	0.0036	-0.0435	0	0	0.0000	0	0.0000	0.0000	-0.1895	-0.1122	-0.0673	0.0128	0.0099	0.013	-0.01	-0.0004	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
13	-0.0606	-0.0427	0.01	0.0049	-0.0884	0	0	0.0000	0	0.0000	0.0000	-0.1286	-0.0997	-0.1762	0.0107	0.0094	0.0153	-0.0025	-0.0001	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
14	0.0004	-0.0028	0.0019	0.0002	0.0000	0	0	0.0000	0	0.0000	0.0000	-0.2178	-0.0002	-0.0004	0.0107	0.0015	0.0000	-0.0005	-0.0007	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
15	-0.2407	0.145	-0.2201	0.0126	-0.1407	0	0	0.0000	0	0.0000	0.0000	0.0000	0.1958	-0.0128	-0.0773	0.0147	-0.0349	-0.0300	0.0017	0.0144	0	0	0	0	0	0.0000	
16	-0.086	0.2223	0.0161	0.0005	0.023	0	0	0.0000	0	0.0000	0.0000	-0.1549	-0.0993	-0.1703	0.0127	0.0084	0.0141	-0.0099	-0.0003	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
17	0.0014	-0.1642	0.0011	-0.0018	-0.1138	0	0	0.0000	0	0.0000	0	-0.0016	-0.1718	-0.0049	-0.0028	-0.0022	0.0000	-0.0004	0.0004	-0.0133	0	0.0000	-0.0000	-0.0000	0	0.0000	
18	-0.2006	-0.0007	0.0025	0.0006	0.0015	0	0	0.0000	0	0.0000	0.0000	0.3308	-0.0009	-0.0027	0.0044	0.0109	0.0123	-0.0089	-0.0002	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
19	0.0159	0.0037	0.0043	0.0027	-0.0018	0	0	0.0000	0	0.0000	0	0.0000	-0.2671	0.0015	-0.0442	0.1094	0.0042	0.0005	0.0012	0.0002	0	0	-0.0008	0.0004	0	0.0000	
20	-0.0049	-0.0002	0.0003	0.001	-0.1476	0	0	0.0000	0	0.0000	0.0000	-0.0002	-0.0041	-0.0046	-0.0403	0.0142	0.0083	0.0104	-0.0004	-0.0009	0	0.0000	0.0000	0.0000	0.0000	0	0.0000
21	-0.0061	0.0049	-0.1504	-0.0106	-0.0008	0	0	0.0000	0	0.0000	0.0000	-0.0764	-0.0053	-0.0902	0.117	-0.0126	-0.1290	0.0024	-0.0023	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
22	0.0000	-0.0006	0.0003	-0.0001	0.16	0	0	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0004	0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0	0.0000	
23	-0.0400	-0.0079	0.0009	0.0008	-0.1234	0	0	0.0000	0	0.0000	0.0000	-0.2059	-0.0119	-0.0607	-0.0101	0.0073	0.013	-0.0282	0.0005	0	0.0000	0.0000	0.0000	0.0000	0	0.0000	
24	0.1200	-0.0717	0.0009	0.0002	-0.1906	0	0	0.0000	0	0.0000	0.0000	-0.3934	-0.1582	-0.0044	-0.3903	0.0143	-0.0223	0.0122	-0.1073	0.0044	0	0.0000	0.0000	0.0000	0	0.0000	
25	-0.2114	-0.0663	0.0028	0.0012	0.0405	0	0	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	-0.1000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0	0.0000	

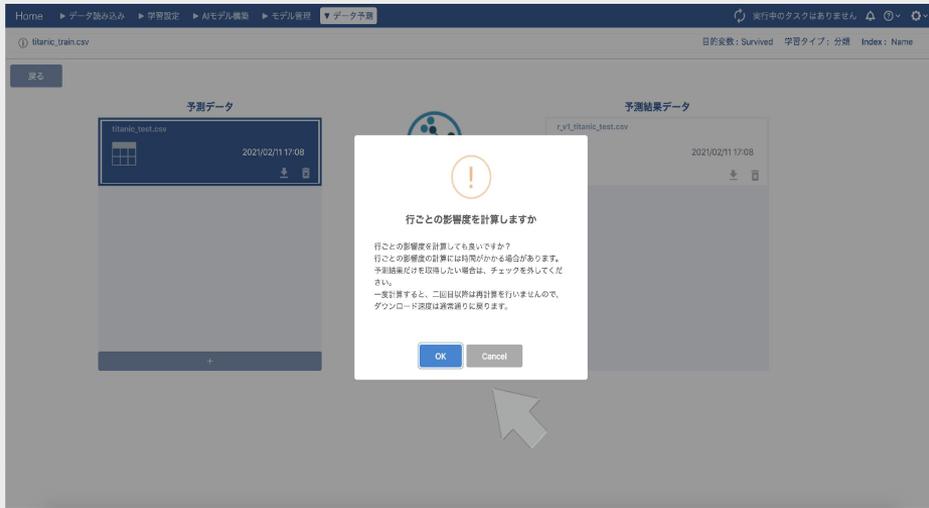
ダウンロードしたファイルは左のようになります。予測結果に行ごとの影響度が結合されているのが確認できます。

アルゴリズムごとに行ごとの影響度があるため、{アルゴリズム名}_{列名}の形式で表示されています。

The screenshot shows a dialog box titled '予測済みデータをダウンロードします' (Download predicted data). It contains a text input field for the file name, currently showing 'c:\AI\検証データ.xlsx'. Below the input field, there are three checkboxes: 'Index' (checked), '検証結果' (unchecked), and '行ごとの影響度' (unchecked). A grey arrow points to the 'ダウンロード' (Download) button.

行ごとの影響度は予測時にも確認できます。予測を行ったファイルをダウンロードする際に、「行ごとの影響度」というチェックボックスが表示されます。

これをチェックした状態でダウンロードを押すことで、予測値に対する行ごとの影響度をダウンロードすることができます。



	A	B	C	D	E	F	G	H	I	J	K
1	Name	result	No	Yes	Pclass	Age	SibSp	Parch	Fare	Sex	Embarked
2	Ponessell, Mr. Martin	No	0.737990377	0.262009623	-0.0002	0.0017	0.0021	0.0067	0.1255	-0.2854	-0.0123
3	Collier, Mrs. Harvey (Charlotte Annie Tate)	Yes	0.040738731	0.959261269	0.189	0.0777	-0.0012	-0.0199	-0.0421	0.4152	0.0018
4	Carter, Master. William Thornton II	Yes	0.393068352	0.606931648	0.1917	0.1582	0.0207	0.0257	0.1789	-0.2884	-0.0186
5	Thomas, Master. Assad Alexander	Yes	0.246223957	0.753776433	-0.0649	0.0212	0.019	0.0297	-0.1267	-0.1367	0.0705
6	Hedman, Mr. Oskar Arvid	No	0.874905684	0.125099316	-0.0611	0.1036	0.0208	0.0041	-0.107	-0.1587	-0.0052
7	Johansson, Mr. Karl Johan	No	0.631497378	0.318502622	-0.0885	0.1445	0.0168	0.0086	0.0562	-0.1029	-0.0005
8	Andrews, Mr. Thomas Jr	No	0.544963398	0.055094632	0.1343	-0.0523	0.0025	0.0006	-0.1983	-0.1058	-0.0044
9	Pettersson, Miss. Ellen Natalia	Yes	0.254481855	0.735518145	-0.2164	-0.0045	0.0248	0.0051	0.1097	0.4961	-0.0118
10	Meyer, Mr. August	No	0.770879193	0.229123007	0.0001	-0.0004	0.0179	0.006	0.1117	-0.2327	-0.0129
11	Chambers, Mrs. Norman Campbell (Bertha Griggs)	Yes	0.009340486	0.990659514	0.0001	0.0445	0.0021	0.0083	0.1337	0.2678	-0.006
12	Alexander, Mr. William	No	0.5948967025	0.051032975	-0.0001	0.0204	0.0076	0.0031	-0.0528	-0.1489	-0.0161
13	Lester, Mr. James	No	0.954247172	0.045752828	-0.0001	0.0476	0.0054	0.0054	-0.0609	-0.1166	-0.0124
14	Slemen, Mr. Richard James	No	0.902752932	0.097247068	-0.0001	0.0248	0.0128	0.0037	0.0055	-0.2275	-0.0099
15	Andersson, Miss. Ebba Iris Alfrida	No	0.85868052	0.14131938	-0.2327	0.2532	-0.2377	-0.0115	-0.1426	0.1875	-0.0135
16	Tottilin, Mr. Ernest Portage	No	0.659099594	0.340960406	-0.0864	0.2146	0.0168	0.0101	0.0143	-0.1556	-0.0116
17	Fry, Mr. Richard	No	0.957011322	0.032988678	0.0914	-0.0461	0.0019	-0.0016	-0.174	-0.1737	-0.0036
18	Haininen, Miss. Wendia Maria	Yes	0.289289444	0.710710556	-0.1891	-0.0806	0.0274	0.0048	0.095	0.539	-0.0245
19	Mallet, Mr. Albart	No	0.848032391	0.151967609	0.0136	0.0454	0.0091	-0.0293	-0.0206	-0.2627	0.0578
20	Holm, Mr. John Fredrik Alexander	No	0.986391362	0.013608638	-0.0365	-0.0505	0.0039	0.0011	-0.1429	-0.0949	-0.0052
21	Skoog, Master. Karl Thorsten	No	0.978938577	0.021061423	-0.048	0.0409	-0.1559	-0.0098	-0.0522	-0.0769	-0.0058
22	Hays, Mrs. Charles Melville (Clara Jennings Gragg)	Yes	0.011398869	0.988601131	0.2087	-0.0428	0.0061	-0.0121	0.1653	0.3286	-0.0039
23	Lulic, Mr. Nikola	No	0.975760507	0.024239493	-0.0421	-0.0376	0.0066	0.0024	-0.1194	-0.1	-0.0144
24	Reuchlin, Jonkhaer. John George	No	0.948161202	0.051838798	0.1366	-0.068	0.0023	0.0004	-0.1869	-0.1674	-0.0038
25	Moor, Mrs. Beilla	Yes	0.414157829	0.585842171	-0.2037	-0.042	0.0416	-0.0193	0.0084	0.4832	-0.0209
26	Parnia, Master. Utho Abraham	No	0.831105988	0.168894012	-0.0651	0.4259	-0.3636	-0.005	-0.335	-0.1213	-0.0081
27	Flynn, Mr. John	No	0.940997024	0.059002976	-0.0305	-0.0345	0.0022	-0.0007	-0.1405	-0.1129	-0.0117
28	Lars, Mr. Len	Yes	0.298810239	0.700189761	-0.082	0.1041	0.025	0.0067	0.4731	-0.1492	-0.016
29	Mallet, Master. Andre	Yes	0.108545919	0.891354081	0.0474	0.5557	0.0244	0.0127	0.0541	-0.1822	0.0406
30	McCormack, Mr. Thomas Joseph	No	0.95843854	0.04155146	-0.0446	-0.0333	0.0049	-0.0009	-0.0445	-0.1595	-0.0191
31	Stane, Mrs. George Nelson (Martha Evelyn)	Yes	0.011752618	0.988247382	0.1821	-0.0177	0.0065	0.0107	0.0847	0.3775	0.0058
32	Yashek, Mrs. Antoni (Salto) Alexander	No	0.922473967	0.077526033	-0.2149	-0.0003	-0.02	0.0017	-0.2183	0.2497	0.0643

ここでも、計算に時間がかかるアルゴリズムがあることにご注意ください。

ただし、

- ・同じデプロイバージョン
- ・同じ予測結果

については、1度計算を完了すれば、2度目の以降のダウンロード時は計算する必要がなくなります。

ダウンロードしたファイルは左のようになります。

予測結果に行ごとの影響度が結合されているのが確認できます。

これで行ごとの影響度の計算の項目は終わりです。

AGENDA

22

デプロイとデータ予測

検証したモデルを運用していくためには、外部から API利用可能なサーバーに移し替える必要があります。

Learning Center Forecast では、この工程を「**モデルのデプロイ**」と呼んでいます。

もちろん予測はAPIだけでなくGUI上からも行うことができます。



モデル検証を終えると、モデルのデプロイを行うことができます。モデルのデプロイとは「モデルを外部から利用可能なサーバーに移す」ことを意味します。デプロイを行うことで、モデルによる予測が可能になります。

①まず検証後のモデルの中から、デプロイしたいモデルを選びます。デプロイしたいモデルのチェックボックスをクリックしてください。

②右上のアクションボタンから「モデルのデプロイ」を選択してください。

③デプロイが完了すると、「データ予測」画面に遷移できるようになります。

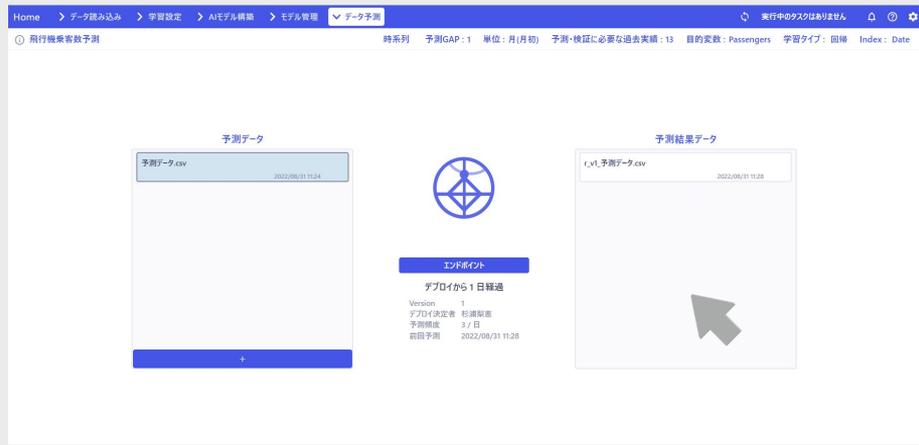
クリックすると、画面が遷移します。



データ予測では、まず予測したいデータをアップロードします。画面左下の「+」ボタンをクリックして、ファイルを追加していきます。



アップロードしたいデータを選択して「アップロード」ボタンをクリックします。



データがアップロードされると、予測データ一覧に表示されます。

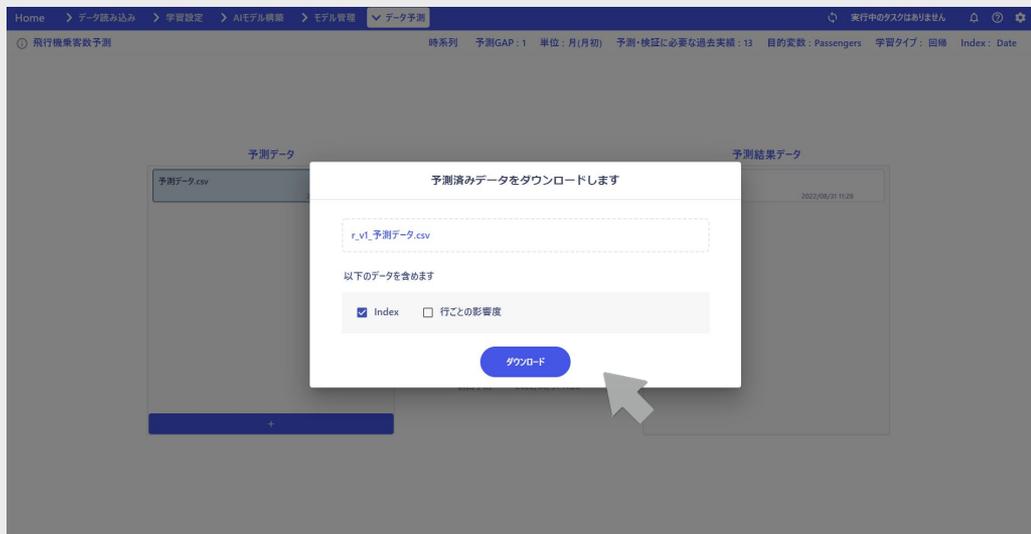
予測を行う場合は、予測したいデータをクリックして選択し、中央にある予測開始ボタンをクリックします。

予測が始まると、タスクが積み上がります。

予測が完了すると、予測結果データが表示されます。

予測結果データの名称は

`r_{モデルのバージョン}_{予測データ名}`という形式になります。同じバージョン、同じ予測データで予測した場合は、予測結果に上書きされます。予測結果をダウンロードする場合は、**ダウンロードボタン**をクリックします。



ダウンロードモーダルが表示されます。

チェックボックスで「予測結果と一緒にダウンロードしたいデータ」を選択します。

たとえば、左の画面のように「Index」にチェックを入れた状態で、「ダウンロード」をクリックすると...

	A	B
1	Date	result
2	1958/5/1	332.565957
3	1958/6/1	364.938694
4	1958/7/1	444.526541
5	1958/8/1	483.44504
6	1958/9/1	487.571095
7	1958/10/1	366.220638
8	1958/11/1	358.993379
9	1958/12/1	301.754677
10	1959/1/1	349.622953
11	1959/2/1	357.35305

このようにIndexと結果が結合したデータがダウンロードできます。



クラス分類の場合「**確信度**」を選択することができます。

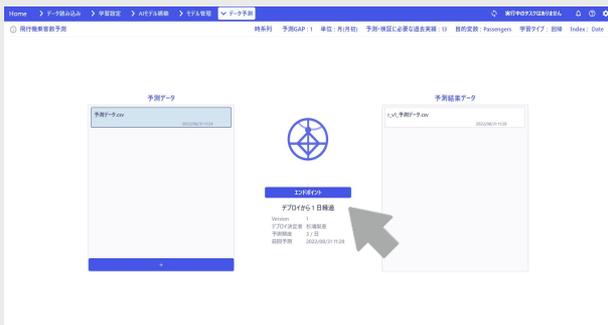
	A	B	C	D	E	F
1	PassengerId	result	No	Yes	Pclass	Age
2	801	No	0.85485971	0.14514029	0.022	0.0013
3	802	Yes	0.03495467	0.96504533	0.1929	0.0714
4	803	Yes	0.38009021	0.61990979	0.0897	0.1832
5	804	Yes	0.1423335	0.8576665	-0.0494	0.6025
6	805	No	0.90480932	0.09519068	-0.0299	0.1056
7	806	No	0.67813938	0.32186062	-0.0568	0.1565
8	807	No	0.97437259	0.02562741	0.0339	-0.0576
9	808	Yes	0.21833785	0.78166215	-0.1977	0.0509
10	809	No	0.86028417	0.13971583	0.0256	-0.0465
11	810	Yes	0.00849019	0.99150981	0.171	0.0569
12	811	No	0.94356221	0.05643779	-0.0281	-0.0061
13	812	No	0.94439626	0.05560374	-0.0475	-0.0585
14	813	No	0.93136028	0.06863972	0.0172	-0.0514
15	814	No	0.87646128	0.12353872	-0.2232	0.2387
16	815	No	0.77043069	0.22956931	-0.0517	0.1421
17	816	No	0.99092419	0.00907581	0.0216	-0.0244

確信度とは、ごく簡単に言うと、「あるモデルによる分類において、あるデータがそのクラスに属する確率」のことです。この確率が最も高いクラスに分類されるとなります。

たとえばABCの3分類の問題の時、確信度が、A:0.6、B:0.3、C:0.1の場合はAに分類されます。A:0.4、B:0.45、C:0.15の場合はBに分類されます。

AGENDA

23 APIを使用したデータ予測



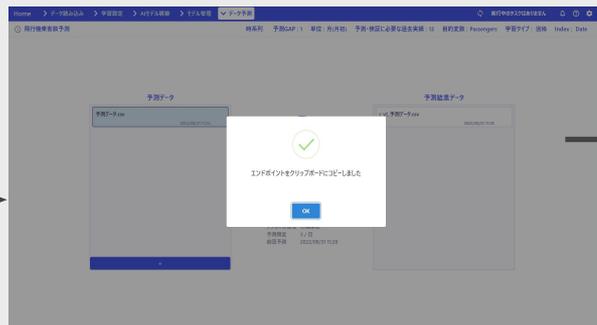
APIによる予測を行うためには、モデルがデプロイされている必要があります。

API予測には

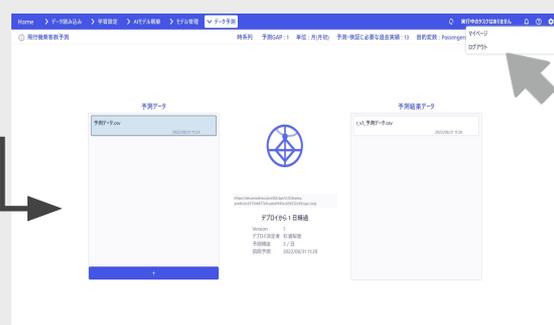
- ・APIエンドポイント
- ・APIキー

が必要です。

APIエンドポイントは、「データ予測」画面の中央からコピーすることができます。

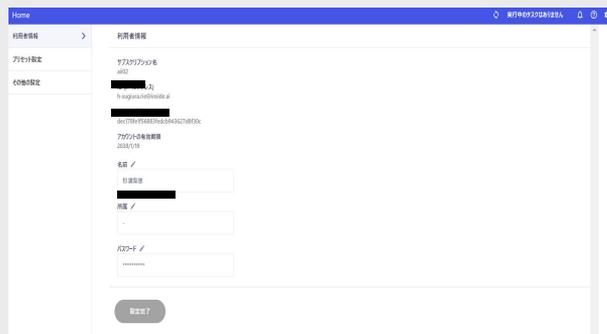


APIエンドポイントをクリックすると、クリップボードにコピーすることができます。



このAPIエンドポイントは保存しておき、次はAPIキーをコピーします。

画面右上の「マイページアイコン」をクリックします。展開されたメニューから「マイページ」をクリックします。



マイページの「APIキー」をコピーします。

これで予測の準備は完了しました。
次は実際のPythonコードでAPI予測を行います。

APIによる予測はPythonを使う場合、非常に少ないコードで行うことができます。

* 推奨Pythonバージョンは**3.6**以上です。この例では3.6を使用してご説明します。

まず、ライブラリをimportします。

サーバーへリクエストするためのrequests、ファイルの読み書きに使用するsys、io、json、最後に、取得したデータを処理するために使用するpandasです。

pandasは必須ではないので、あくまでも予測データを加工したい場合や可視化したい場合のみご使用ください。

この例では、pandasを as文によってpdという名称に縮めて使用できるようにしています。

pandasの詳細情報は以下で閲覧可能です。

<https://pypi.org/project/pandas/>

```
#ライブラリのインポート
```

```
import requests
import sys,io,json
import pandas as pd
```

ヘッダー情報を設定します。
API経由で予測できるデータの文字コードは**UTF-8**のみとなっています。

```
#ヘッダーの設定
```

```
headers = {'Content-Type': 'text/plain; charset=UTF-8'}
```

予測したいファイルを読み込みます。文字コードは**UTF-8**、一度に予測できるデータサイズは**10MB**までです。

Excelなどで UTF-8で書き出した場合は、下記のように**encoding** の引数を`utf-8`で指定してください。

```
#予測したいファイルのpath
file_path = '.././../Datas/AMATERAS RAY2.2.0 sampledata/テストデータ_旅客機乗客者数予測/Airpassengers_test.csv'

#ファイルの読み込み
fp = open(file_path, 'r').read()

#Excel等でUTF-8で書き出した場合は以下のように読み込んでください
fp = open(file_path, 'r', encoding='utf-8').read()
```

アップロードするデータをリクエスト可能な形式(辞書型)に整えます。

```
#送信するためのデータ  
data = {'data':fp}|
```

この時、下記のようなパラメタを、オプションで渡すことが可能です。

パラメタ名	データ型	概要
file_name	str	アップロードするファイルの名称を指定します。 GUI上でAPI経由のファイルを判別したい際に使用します。
on_index	int	1を指定することで、Index列を結合した予測結果を取得できます。
on_proba	int	1を指定することで、予測の確信度を結合した予測結果を取得できます。(分類のみ)
on_predict	int	1を指定することで、予測データを結合した予測結果を取得できます。

* 現在の最新バージョン(v2.5)では、APIから「行ごとの影響度」の結果を取得することはできません。今後のバージョンアップで対応予定です

先ほどコピーした **エンドポイント**、**APIキー**を一つに連結し、データと一緒に **POST**でリクエストします。
V2.0のAPIを使用する場合はエンドポイントのバージョンを変更してください。

```
#AMATERAS上でコピーしたエンドポイント v1.0  
end_point = 'https://amateras.tk/6gou/api/v1.0/deploy-predict/77997b8b1056e6da2a993401dfb27459'  
  
#AMATERAS上でコピーしたエンドポイント v2.0  
#end_point = 'https://amateras.tk/6gou/api/v2.0/deploy-predict/77997b8b1056e6da2a993401dfb27459'  
  
#api_key  
api_key = '881390a1ecdac1ebb5129e4b0f1ee0a1'  
  
#リクエスト  
predictions_response = requests.post(f'{end_point}/{api_key}',data)
```

予測結果を取得します。

```
#response
```

```
message = predictions_response
```

予測結果はJSON形式で受け取ることが可能です。
V1.0のフォーマットは以下のようになっています。

```
{
  result:{
    column1:{'1':1,'2':2...},
    column2:{'1':'test','2':'test_2'...},
    ...
  }
},
error:null
}
```

V2.0のフォーマットは以下のようになっています。

```
{
  result:[
    {column1:1,column2:'test'},
    {column1:2,column2:'test_2'},
    ...
  ]},
error:null
}
```

処理中にエラーが発生するとerrorメッセージが非nullになります。

予測結果をJSON形式でロードし、Pythonのライブラリである **pandas** で利用可能な形に戻しています。

```
#データをjson化 v1.0
load_json = json.loads(message.json()['result']['data'])

#データをjson化 v2.0
load_json = json.loads(message.json()['result'])

#pandas dataframeに変換
result_df =pd.DataFrame(data=load_json)
```

今回のコード例はPythonを使用しましたが、**RESTfulAPI**なので、あらゆる言語やリクエストに対応可能です。

たとえば、curlを利用したリクエストは、以下のように表記できます。

```
curl -X POST '{apiエンドポイント}/{apiキー}' -F data=@{予測ファイルの絶対path}
```

APIを使用したデータ予測の説明は以上です。

AGENDA

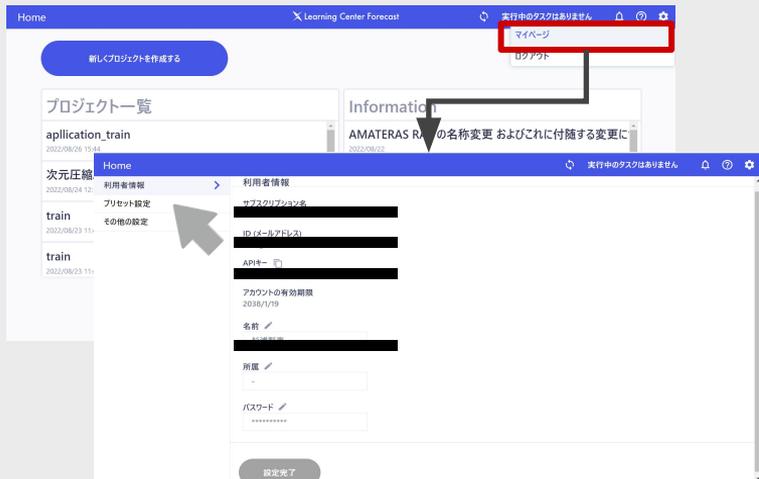
24

プリセットデータ登録

Learning Center Forecast を使用するメンバーで共通のデフォルトプロジェクトやデフォルト学習データを設定できる機能です。

たとえば、分析担当に新しいメンバーが入ってくるときに、練習問題や既存のプロジェクトを効率よく共有することができます。

権限がないと使用できない機能なので、実行したい場合は admin 権限の申請をお願い致します

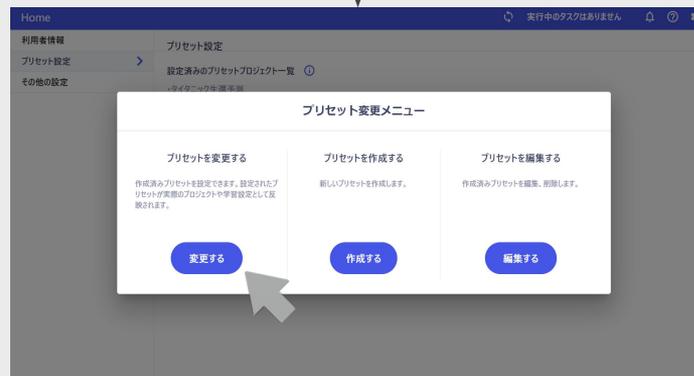


右上の「歯車マーク」をクリックして、メニューを表示し、「マイページ」を選択してください。その後左側のメニューから、「プリセット設定」を選択してください。

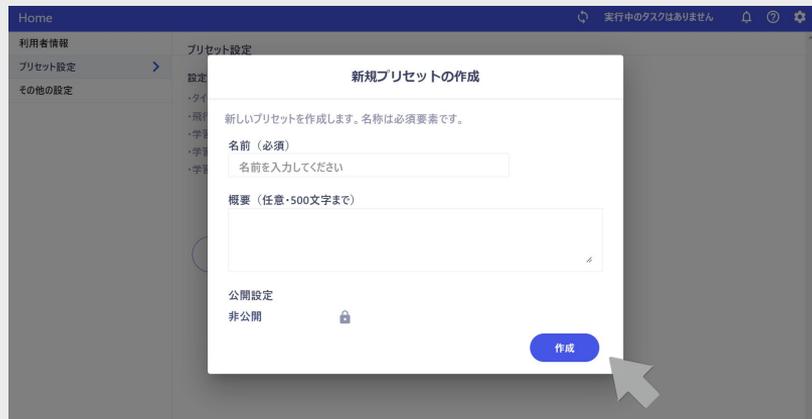


「プリセットをリセットする」をクリックすると、現在のプリセットプロジェクトを削除し、再作成します。
(この際、学習済みのモデルについても削除されてしまうので、ご注意ください)
プリセット学習設定については削除されず、「プリセット」の設定がはずれ、新しく作成されます。

設定を変更したい場合は「プリセットを変更する」をクリックしてください。



プリセット変更メニューから、必要な操作を実施します。
新しいプリセットを登録したい場合は、「プリセットを作成する」をクリックします。



「名前」「概要」「公開設定」を入力します。
「名前」のみが必須です。

公開設定は、プリセットを他のサブスクリプションに公開するかどうかを設定できます。

公開したプリセットは「他のサブスクリプション(他の企業)」からも閲覧可能になってしまうため、設定には注意してください。(社内メンバーのみが利用したい場合は、非公開の設定にしてください)

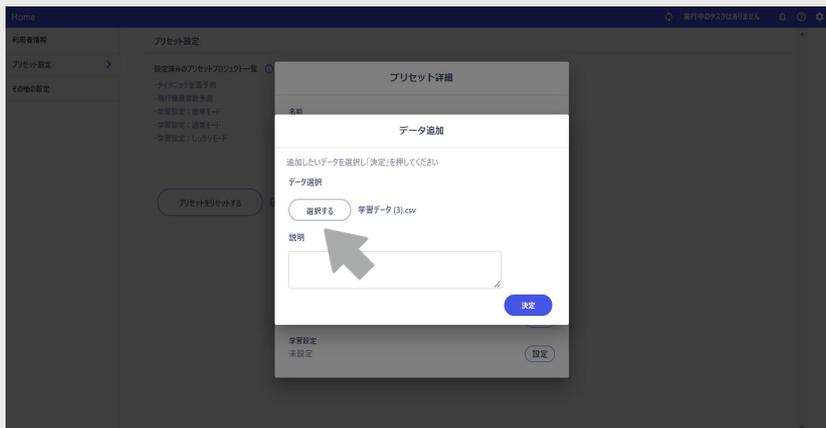
入力が完了したら「作成」をクリックしてください。



学習データ、検証データ、検証追加データ、学習設定、それぞれをプリセットで登録できます。

プロジェクトとして登録する場合は、「学習データ」が必須になります。
プリセット学習設定として登録したい場合は、学習設定のみ登録します。

ここでは学習データと学習設定を登録します。
学習の横の「設定」ボタンをクリックします。

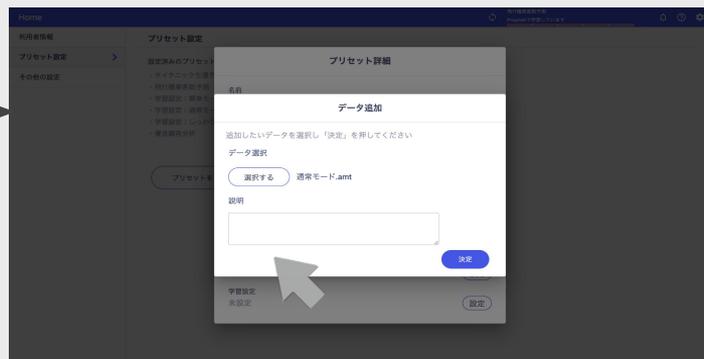


「選択する」でデータを選んでアップロードします。
データの選択は必須項目です。
「説明」は学習データに関する概要などを記載できます。

設定が完了したら「決定」をクリックしてください。



設定が完了すると、「学習」の横に先ほどアップロードしたデータが表示されます。
次に「学習設定」を登録します。

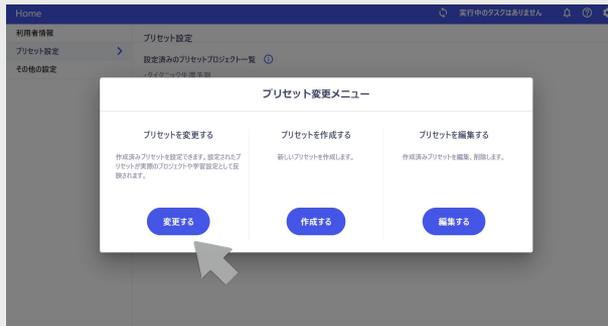


学習設定は、Learning Center Forecast の「学習設定エクスポート機能」を利用して出力した amt ファイルをご利用ください。
(詳細は「学習設定」の章をご確認ください)
設定が完了したら「決定」をクリックします。



これで「学習データ」と「学習設定」の登録が完了しました。これだけでは「プリセットを作成」しただけですので、利用できません。今度は作成したプリセットをサブスクリプションに登録する作業を行います。

「キャンセル」をクリックしてメインメニューに戻ります。



サブスクリプションに登録するプリセットを変更する場合は、「プリセットを変更する」をクリックしてください。

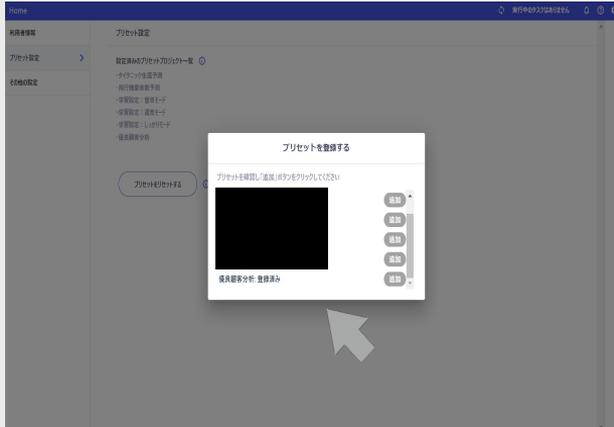


この画面では現在登録されているプリセットが一覧で表示されます。現在は作成しただけで、登録されていません。

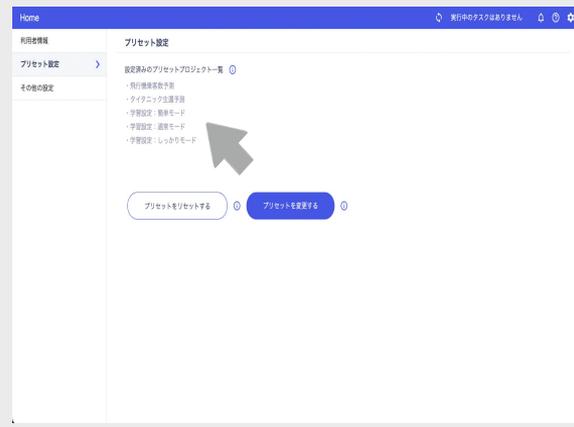
先ほど作成したプリセットを登録していきます。ポップアップ右下の「追加」をクリックしてください。



作成済みのプリセットが表示されますので、登録したいプリセットの横にある「追加」をクリックしてください。



追加が完了すると「登録済み」という表示に変わります。これで登録完了です。



プロジェクト一覧に戻れば、先ほど設定したプリセットをご確認いただけます。



登録したプリセットが表示されていたら成功です。プリセットのプロジェクトには右下に「プリセット」というマークが表示されています。プリセットプロジェクトの登録方法の解説は以上です。

AGENDA

25 規約について

**Learning Center Forecast の規約は以下からご覧いただけます。
ご利用いただく際に必ずお読みください。
規約を改定する際にはこちらのページに告知を掲載します。**

現行規約はこちら

AGENDA

26 お問い合わせ

お問い合わせ

操作方法や、Learning Center Forecast の仕様でご不明な点があれば、下記のメールアドレスまでご連絡ください。5営業日以内にご返信いたします。

もし5営業日以内に返信がない場合、お手数ですが、その旨ご連絡いただければ幸いです。

今後ともLearning Center Forecast をどうぞよろしくお願い致します。

support_lcf@inside.ai