

03

需要予測モデルを
構築し、評価せよ



はじめに：本パートで学べる内容

本パート「**AIモデル構築**」では、実際の企業の導入に求められる精度レベルを超えたAIモデルを開発するために必要なスキルを学ぶことができます。

(具体的には、以下の内容を学びます)

- データ前処理
- データ集計/可視化
- モデル開発

企業へのAI導入を進めるために必要なことを、
このパートを通じて理解しましょう！



進め方と マイルストン

- 3.5週間をかけAIモデル開発及びその評価を行っていただきます。
- 本モデル開発はコンペ形式で行います。参加者はコンペ期間中何回もモデルの評価結果を投稿することができます。ただし、1日あたりの投稿回数上限は5回で、毎日深夜0時にリセットされます。また投稿結果はランキングボードに反映されます。
- 自主学習だけではモデル開発が進まない受講者向けにガイドコンテンツを用意しています。
 - ガイドコンテンツには、学習補助教材と、モデリングのサンプルコードがあります。
 - 必要に応じ、PBLサイト内のページよりダウンロードし、活用してください。
 - サンプルコードは、内容を理解し、ご自身で改修して頂くことで、学びを深めて頂くことを意図したものです。特に、モデリングになかなか手が付けられないという場合には有効に使っていただけるものとなっておりますので、必要に応じ適宜ご活用下さい。
- 成績上位者には、記載したコードの提出をお願いする可能性があります。

提出物

- 提出用ファイルを作成、提出
 - ファイルの形式は、sample_submission.csvを参考作成
 - 例：`submission.to_csv('my_submission.csv', index=False, header=False)`
 - 右の例のように、提出するCSVファイルには「ヘッダー」や「インデックス」が含まれないようにしてください

2 列	
0	予測0
1	予測1
2	予測2
⋮	
2697	予測2697
2698	予測2698
2699	予測2699

2700 行

演習④：モデリングコンペの実施

要求
定義

要件
定義

モデル
開発

実装・
運用
計画

最終
プレゼン

これまでの検討を踏まえ、発注量検討のための需要予測のAI化が、業務工数削減や在庫最適化、業務属人化解消につながるとの社内コンセンサスが得られた。これを受け、現行調達部門において経験ベースで行われている需要予測を、AIにより自動化する方向で検討が進められることになった。

課題

需要予測対象である170商品の、2019年12月の各店舗における売上数を予測するモデルを構築し、その精度を評価せよ

【留意点】

本課題は実プロジェクトでのPoCにあたるものとして、コンペティション形式で、モデルを構築し、2019年12月ひと月分の予測を行っていただきます。実際のAI導入プロジェクトにおいては、PoCで精度が確認でき、本格導入が決定した場合は、実業務環境に組み込み、本番実装・運用することとなり、その場合には、当然12月以外の予測も行うことが想定されます。したがって、本コンペにおいても、ターゲットの2019年12月の精度を追求するだけでなく、実際に今後運用していくことが可能となるような予測モデルの作成を心がけてください。

演習④：対象データ

わんだーゲーム本舗の2018/1～2019/10までの過去売上データを対象にします。売上データは日次で記録されており、商品IDや店舗ID、価格なども確認することができます。今回の予測ターゲットは2019年12月における18店舗×170商品における予想売上数です。以下が使用するデータの概要です。

メインデータの詳細

sales_history.csv (121216 rows × 6 columns)

カラム	ヘッダー名称	データ型	説明
0	日付	datetime	yyyy-mm-dd形式
1	店舗ID	int	店舗の識別番号
2	商品ID	int	商品の識別番号
3	商品カテゴリID	int	商品カテゴリの識別番号
4	販売価格	int	その日その店舗で販売した価格
5	売上個数	int	売上数

データの種類と用途

データは①学習データ、②評価データ、③暫定評価用正解データ、④最終評価用正解データ、⑤応募用サンプルファイルの5種あります。それぞれの用途は以下の通りです。

- ① 学習データ (sales_history.csv) : 2018/1～2019/10までの過去売上データ
- ② 評価データ (test.csv) : 予測対象である2019/12の店舗×商品を示したデータ
- ③ 応募用サンプルファイル (sample_submission.csv) : コンペへ投稿する際のフォーマットデータ
- ④ 暫定評価用正解データ: 暫定評価結果をスコアボードに掲載するためのデータ (受講者には配布されません)
- ⑤ 最終評価用正解データ: コンペ終了後の最終評価を行うためのデータ (コンペ終了後に配布されます)