

03

故障予兆判定モデルを
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！



演習④：演習の進め方

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

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で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。
- 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を参考に作成
 - 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」が含まれないようにしてください。

2列	
data_000.csv	予測結果0
data_001.csv	予測結果1
data_002.csv	予測結果2
⋮	
data_00x.csv	予測結果xxx

xxx行

注) 予測結果は、正常を0、故障予兆を1で示す。

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、大型プレス機の主軸破損におけるAIによる予知保全が、売上ロスの低減・コスト削減・納期遅延リスクの低減・属人化の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、故障予兆検知を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

PoC対象である大型プレス機に設置した振動・音響センサーのデータから、正常か故障予兆かを予測するモデルを構築し、その精度を評価せよ。

【留意点】

成績優秀者には、コンペ後集合日程での発表、コードの提出をお願いする可能性があります。

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・運用することになりますので、単に予測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

要求
定義

要件
定義

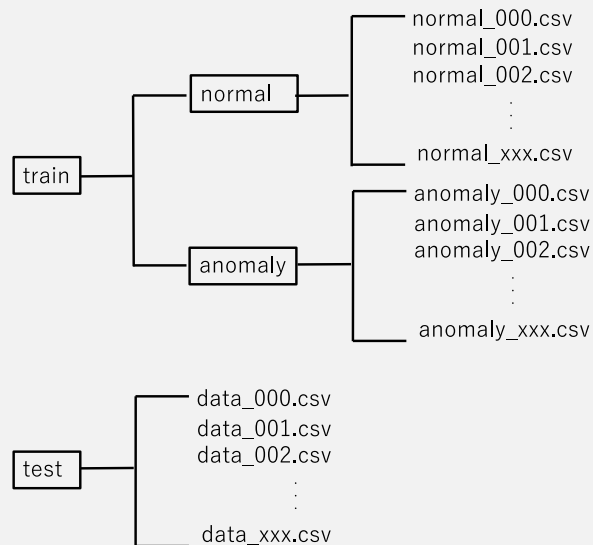
モデル
開発

実装・
運用
計画

最終
プレゼン

PoC対象である大型プレス機に設置した振動・音響センサーのデータを対象とします。検証に使用するデータは、製造部で蓄積されている正常データと故障予兆データからある時間単位でランダムに抽出したものです。以下が使用するデータの概要です。

フォルダ構成



ファイルの中身(normal_001.csvのイメージ)
サンプリング周波数：50kHz、データ長さ：15,000行(0.3秒相当)
1～6列目：振動センサー1～6 (g)、7列目：音響センサー (dBV)

	A	B	C	D	E	F	G
1	-8.61E-01	-2.62E-01	-5.19E-02	2.72E-01	1.34E-04	-7.08E-02	2.12E-01
2	2.15E-01	3.16E-01	4.62E-02	3.18E-01	8.01E-03	-2.23E-02	-1.28E-01
3	7.54E-03	-1.04E-01	-2.71E-02	3.01E-01	-1.88E-02	-4.56E-02	1.55E-01
4	-6.17E-01	5.07E-02	-2.59E-03	2.52E-01	-2.68E-02	-8.37E-02	-2.78E-02
5	8.79E-01	1.11E-01	1.73E-02	3.19E-01	-1.88E-02	-2.83E-03	2.51E-02
6	-1.29E+00	-2.23E-01	-4.27E-02	2.01E-01	-2.11E-02	-9.55E-02	8.93E-02
7	1.30E+00	2.91E-01	5.75E-02	3.14E-01	-1.78E-02	1.60E-02	-8.96E-02
8	-1.20E+00	-3.53E-01	-6.43E-02	1.87E-01	-2.59E-02	-9.45E-02	1.93E-01
9	9.69E-01	3.50E-01	5.38E-02	2.85E-01	-2.65E-02	7.01E-04	-1.48E-01
10	-6.99E-01	-2.83E-01	-5.30E-02	2.13E-01	-2.68E-02	-5.74E-02	2.05E-01
11	8.89E-02	2.04E-01	3.14E-02	2.37E-01	-2.53E-02	-1.96E-02	-1.14E-01

データの種類と用途

データは①メインデータ、②応募用サンプルファイル、③正解ラベルデータの3種類あります。それぞれの用途は以下の通りです。

- ①メインデータ：大型プレス機に設置した振動・音響センサーから得られた*.csv形式のデータが、学習データフォルダ(train)と評価データフォルダ(test)に分かれております。学習データフォルダには、正常データ(normal)と故障予兆データ(anomaly)の2つのサブフォルダが入っております。
- ②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です。
- ③正解ラベルデータ：評価データフォルダ(test)内の各csvファイルの正解ラベルが含まれるデータです。

※各csvファイルの正解ラベルはコンペ終了後に配布されます。なお、今回のコンペでは過剰に評価データ(test)に適合したモデルで高い精度が出ることを防ぐため、評価データ(test)を暫定評価用と最終評価用に分けております。提出の締切までは、リーダーボード上には、暫定評価用データをもとにした暫定評価結果をPublic scoreとして掲載します。提出締切後は、リーダーボード上には、最終評価用データをもとにした最終評価結果をPrivate scoreとして掲載します。Private scoreに基づく順位が最終順位となりますので、ご注意ください。