



製造業対抗データインサイトチャレンジ

製造業対抗でデータサイエンスのスキルを競う分析コンペティション！

株式会社SIGNATE



締切	投稿	参加
残り29日	97件	130人
2025.11.14 23:59		

製造業向け業務プロセス改善DX商材の購...

説明

課題

データ

リーダーボード

ナレッジ

投稿ファイル

投稿

課題内容

お知らせ

課題概要

デジタルトランスフォーメーション（DX）は、今や多くの企業にとって避けては通れないテーマです。企業が競争力を維持し、変化の激しい社会に対応していくためには、日々の業務プロセスを効率化・高度化し、生産性を高めることが欠かせません。そのための手段のひとつが「業務プロセス改善DX商材」です。これらの取り組みは、単にツールを導入することにとどまらず、企業が自らの業務をどう変えていくかという重要な意思決定に関わっています。

今回のコンペでは、参加者の皆さんに「どの企業が業務プロセス改善DX商材を導入するのか」を予測するモデルを作っていただきます。ご用意しているのは、DXの中でも特に多くの人が関わりやすい「業務プロセス改善」をテーマにしたデータで、コンペのために設計された企業データとなります。従業員数や業界区分、主要な業務内容、組織構造、財務指標、アンケートの情報など、さまざまな切り口で企業をとらえられる内容になっており、さまざまな形でDXに関わる皆さんが、より身近に感じられるテーマとして設計しています。

これらのデータを分析し、機械学習等の手法を駆使して予測モデルを構築し、「どのような企業が業務プロセス改善に積極的に取り組む傾向にあるのか」をデータから導き出すことが、このコンペの課題となります。単に予測精度を競うだけでなく、結果を通じて「なぜその企業が導入するのか」といった背景を考察することで、データ分析ならではの発見につながるかもしれません。製造業においても、「積極的なプロセス改善」の取り組みを行うには、どのような傾向が必要か、データに基づいて「なぜその企業が導入するのか」などの因果関係を考察することが重要です。

身近な企業活動をモチーフにしたデータを扱うことで、現実社会に近い課題に挑戦できるのも、このコンペの大きな魅力です。データサイエンスのスキルを存分に発揮しながら、楽しみながら取り組んでいただければ幸いです。

目次

課題内容

評価方法

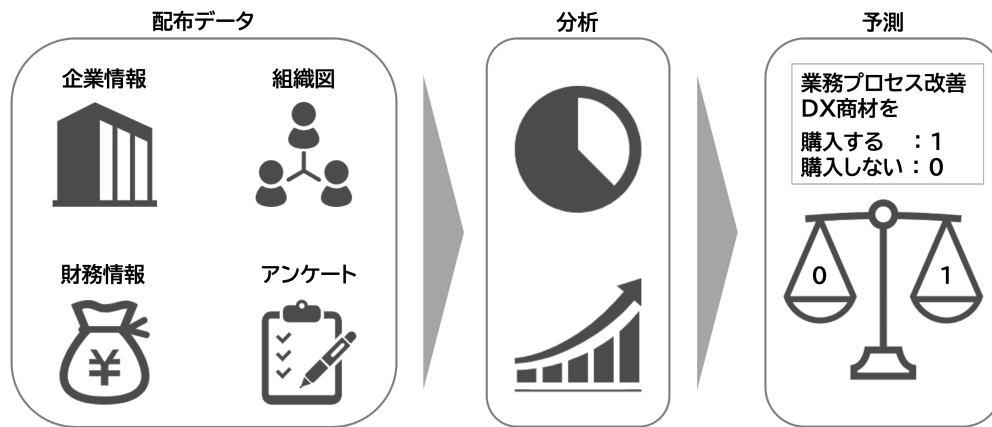
ルール

タスク内容及びデータ概要

本コンペティションでは、架空の企業データセットを用いて、**各企業が「業務プロセス改善DX商材」を導入するかどうか（導入=1, 非導入=0）を予測する二値分類モデル**を構築していただきます。

ご提供するデータセットには、企業の多面的な特徴を示す、以下のような合計42個の変数が含まれており、現実的な企業をイメージした内容になっています。(各カラムの詳細説明は、データタブより「description.csv」ファイルをダウンロードして確認してください。)

- 企業の**基本情報**：従業員数、業界、組織といった、企業の規模や性質を示す基本的なデータ。
- 財務**三表データ**：B/S（貸借対照表）、P/L（損益計算書）、C/F（キャッシュフロー計算書）からなる財務情報。
- 企業への**アンケート結果**：企業文化やDXの現状、改善ニーズ、導入意欲に関する複数のアンケート結果。特に、自由記述形式のテキストデータが含まれており、現場の課題意識や改善要望といった定性的な意見も分析の対象となります。



このタスクでは、**数値**、**カテゴリ**、そして**自然言語テキスト**といった形式の異なるデータを統合的に扱う必要があります。これらの異種混合データから、いかにして企業の業務プロセス改善に関する投資判断に繋がる本質的な特徴を抽出し、有効な特徴量を設計するかが、モデルの精度を向上させる鍵となります。

皆様には、データ分析と機械学習のスキルを駆使して「**どのような企業が業務プロセス改善DX商材を導入する傾向にあるのか**」という**インサイトを導き出す**とともに、高精度な予測モデルを構築していただくことを期待しています。

単なる分類問題としてだけでなく、データの背後にある“企業の意思決定や改善意識”を読み解く視点も重視して取り組んでいただければ幸いです。

※「業務プロセス改善DX商材」とは、RPAツールや業務自動化プラットフォーム、ワークフロー管理システム、データ連携基盤など、企業の業務効率化・省力化を支援する各種ソリューションを指します。

評価方法

- 精度評価は、「F1Score (binary)」を使用します。
 - F1ScoreはPrecisionとRecallのスコアのバランスを考慮した指標です。
 - スコアは0~1で計算されます。
 - 競合する分類の場合は1、そうでない場合を0とします。
 - 各画像に対してTP, FN, FPを判定して下記式に従ってF1Scoreを算出します。
- 評価値は0以上の値をとり、精度が高いほど1に近づきます。
- Precisionは**モデルの正の予測(競合する分類の予測)がどれほど正しいか**を評価する指標です
- Recallは**実際の正のデータ(競合する分類のデータ)をどれほど正しく当てられるか**を評価する指標です

$$F1Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

TP : 正解が1かつ、予測も1

FN : 正解が1、予測は0

FP : 正解が0、予測は1

- 暫定評価は評価用データセット全体の30%、最終評価は評価用データセット全体の70%となります。
 - データ分割の都合上、暫定評価のスコアに対して良い投稿ファイルを選択するだけでは、最終評価のスコアが大きく落ち、順位が大きく変動することもありますので、ご注意ください。

ルール

📖 ルール詳細



