

### 第1回 アプリ評価を予測せよ!

#### 課題

iTunes Storeのアプリ情報とレビュー内容から、アプリの評価点を予測する

「操作が快適」「広告が多い」といったレビューの文章や、アプリのカテゴリ・価格・リリース情報など、さまざまな特徴量をもとに、アプリの評価(星の数)を推定するという内容。評価指標には正規化ジニ係数が用いられ、予測値と実際の評価点の並びがどれだけ一致しているかが問われました。

実施期間 2020/12/1~2021/1/19 参加チーム数 53 Teams

### 第3回 ボクササイズの動きを分類せよ!

#### 課題

ウェアラブルセンサーデータを用いた、ボクササイズ動作の分類

Xenoma社製のウェアラブルセンサー「e-skin DK」を用いて、フィットボクシングのような動作を行い取得したデータから、12種類のコンビネーション動作のいずれかを予測するという内容です。センサーデータ特有のノイズや時間方向の特徴をどう捉えるか。データの量とモデリングの難度が程よく設計された、“実践的な機械学習”を体感できるコンペティションとなりました。

実施期間 2021/4/23~2021/6/29 参加チーム数 76 Teams

### 第2回 住宅価格を予測せよ!

#### 課題

大阪府の取引価格を、その他地域の住宅情報を用いて予測する

大阪府のデータを使わずに、近畿地方のその他地域の住宅情報——広さや築年数、最寄駅までの距離、地価などの特徴量をもとにして、大阪府の住宅価格を推定するという内容。地域差をどう捉えるか、そして“未知の大阪”をどこまで正確に予測できるかが勝負どころとなりました。

実施期間 2021/2/1~2021/3/23 参加チーム数 43 Teams

### 第4回 アニメの評価スコアを予測せよ!

#### 課題

アニメのデータからレビューサイトにおける評価スコアを予測する

数値・カテゴリ・テキストといったベーシックなテーブルデータを用いて、アニメ作品に関するデータからレビューサイトの評価スコアを予測するという内容です。シンプルな構造ながら、特徴量の組み合わせやマルチテーブルの扱い方など、ビジネスデータ分析にも通じる基本的なモデリングスキルを磨ける課題でした。

実施期間 2021/8/5~2021/9/16 参加チーム数 50 Teams

#### 上位者のVoice



#### Voice | 第1回上位者 | N.Oさん

第1回ということでもわからず、この回最多の58回もサブミットしてしまいました。内容は、当時の勾配ブースティングから今はLLMへと変わってきており、時代の変革を肌で感じています。今の時代を生きる必須の知識がコンペを通じて身に付くことは確かです。



#### Voice | 第3回上位者 | K.Iさん

達成感と焦りを味わいながら、クラスタリングを使った教師なし学習に挑戦しました。モデルやハイパーパラメータへの理解は、その後のE資格取得や実務にも生きています。ChatGPTでコーディングの敷居が下がった今、Kaggleは良い練習の場だと思います。