

第5回 住宅価格を予測せよ！

課題

Austinの住宅価格を、
マルチモーダルな住宅情報から予測する

数値・カテゴリ・テキスト・画像・位置情報といった多様なデータを組み合わせて、Austinの住宅価格を推定する“マルチモーダル回帰”の課題です。住宅価格予測は、入門用データセットとして有名な Boston Housing に代表されるように、教師あり学習の典型的なテーマ。株価のように時系列変動が大きいわけではなく、「どこにある物件か」「どんな物件か」といった属性が価格に強く影響する、いわば“どっしりした”データであることが特徴であるといえます。

実施期間 2021/10/5～2021/11/22 参加チーム数 56 Teams

第7回 中古車価格を予測せよ！

課題

中古車情報からその価格を予測する

中古車に関するさまざまな情報から、その販売価格を推定する回帰タスクです。典型的なテーブルデータを用いた教師あり学習の基礎を身につけるのに最適な構成で、特徴量の理解や前処理など、実務にも通じる基本スキルを磨ける課題でした。また、本コンペティションでは（正解データそのものを除き）外部データの利用が認められており、スクレイピングなどの実践的なデータ収集も学べる形式となっていました。

実施期間 2022/3/22～2022/5/24 参加チーム数 85 Teams

第6回 ゲームレビューの評価結果を予測せよ！

課題

ゲームのレビューデータから
評価結果を予測する

ゲームに関するレビュー情報をもとに、その評価結果を分類するテーブルデータ中心の内容です。しっかり前処理に取り組む必要があるデータ構造で、実践的な特徴量設計スキルが問われました。（おまけで画像データも含まれていますが）基本はテーブルデータによる分類タスクで、評価指標にも標準的なものが採用された、基礎をしっかりと固められる課題でした。

実施期間 2021/12/13～2022/2/7 参加チーム数 73 Teams

第8回 中古車価格を予測せよ！

課題

中古車情報からその価格を予測する

中古車に関する多様な情報をもとに、その販売価格を推定する回帰タスクです。データセットには中古車の価格情報が含まれており、前回コンペティション「Data Science Osaka Spring 2022」で参加者によって収集された関連画像なども追加で利用することができました。典型的なテーブルデータを用いた教師あり学習の基礎を身につけることができる、実務にもつながる基本スキル習得に適した課題でした。

実施期間 2022/6/20～2022/8/23 参加チーム数 55 Teams

上位者のVoice



Voice | 第6回上位者 | K.Oさん

欠測値から特徴量を作る発想や、CatBoostのカテゴリ変数処理、BOWなどの古典的手法が意外と効果的でした。CVとLBの相関を追いながら進めることで、シェイクダウンを防げます。コミュニティでの情報共有も大きな学びになります。



Voice | 第7回上位者 | T.Iさん

Kaggleは勝つか学ぶか。失敗も糧にして挑戦を楽しむ姿勢が大切だと思います。ぜひ『Win or Learn』でチャレンジしてください。