

コンペ XX 説明文

アカウント名

やったことの全体像

まずは demo_basic.ipynb 参考に、ベースラインとして「Sex, Pclass, Age だけ」を使ったシンプルなモデルを作成

そのあと、少しずつ特徴量を追加して、スコアがどう変わるかを確認しながら改善した

前処理で工夫した点

Age の欠損値

単純に「中央値」で埋めた

その上で、Age を「子供 / 大人 / 高齢」の 3 区分に分けた（ビンニング）

Embarked の欠損値

一番多い港（S）で埋めるシンプルな方法にした

特徴量エンジニアリング

FamilySize = SibSp + Parch + 1 を追加

→ 一人旅か家族連れかで、生存率が変わると考えた

IsAlone = (FamilySize == 1) かどうかのフラグを作成

→ 一人旅の人を簡単に判別できるようにした

Title（敬称）の抽出

Name から Mr, Mrs, Miss, Master を取り出してカテゴリ変数にした

レアな敬称は "Other" にまとめて、カテゴリ数を増やしすぎないようにした

モデル

ロジスティック回帰とランダムフォレストを比較した

学び・振り返り

難しいモデルにいきなり行くよりも、「特徴量を 1 つ足して、スコアがどう動くか」を確認するほうが理解が深まった

特に FamilySize, IsAlone, Title は、スコア向上に貢献したと感じた