

Student Cup 2024

3rd place solution

user name : kyo99999
date : 2024.9.21

現状

どんな人が旅行プランを成約するかなんとなくわかる
≒ 情報を共有しづらい



課題

マーケティングコストの高さ



目的

旅行パッケージを成約するか機械学習で予測

1. 年齢の欠損を回帰予測
2. AND演算子による特徴の強調
3. 決定木とNNを組み合わせる

動機：年齢は約100か所欠損している
年齢の欠損補完を代表値で行うことは不適切

△ Age

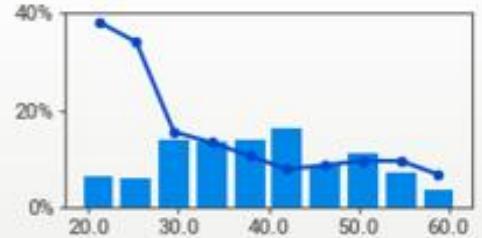
VALUES: 3,389 (97%)
MISSING: 100 (3%)

DISTINCT: 43 (1%)

ZEROES: ---

MAX 61.0
95% 55.0
Q3 47.0
AVG 38.9
MEDIAN 38.0
Q1 31.0
5% 22.0
MIN 19.0

RANGE 42.0
IQR 16.0
STD 9.97
VAR 99.4
KURT. -0.809
SKEW 0.106
SUM 132k



アプローチ：lightGBMで年齢を回帰予測

要点

目的変数を年齢の欠損箇所に設定

過学習を避けるためハイパーパラメータは調整しない

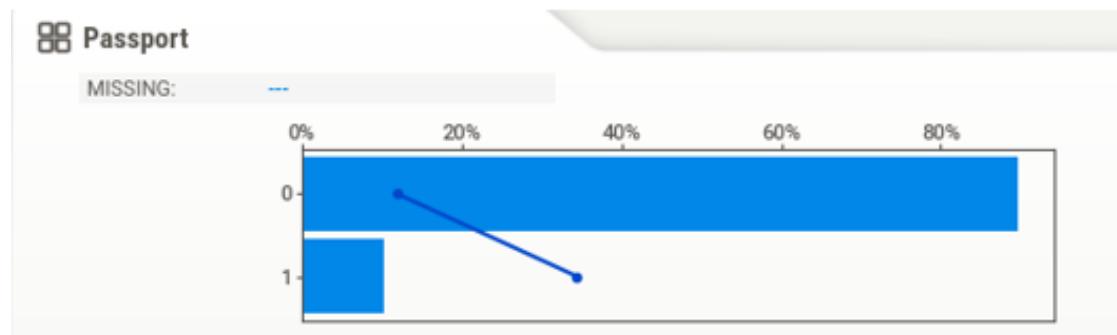
1. 年齢の欠損を回帰予測
2. AND演算子による特徴の強調
3. 決定木とNNを組み合わせる

動機：成約した人の特徴を強調したい

成約：非成約 = 14 % : 86 %
⇒成約した人の情報が少ない

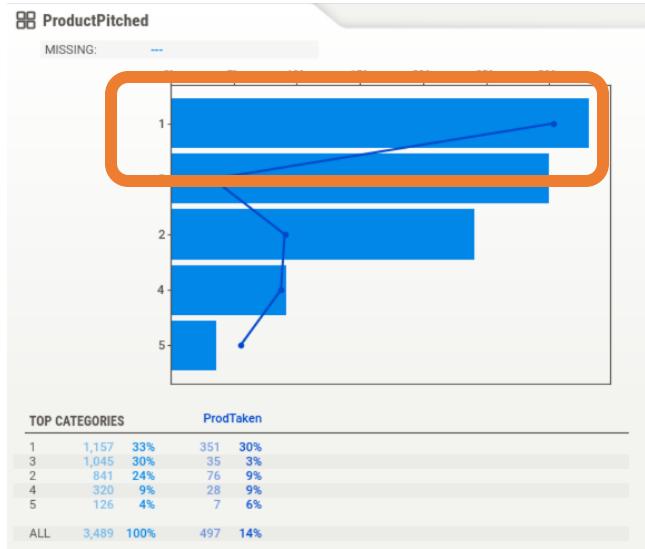
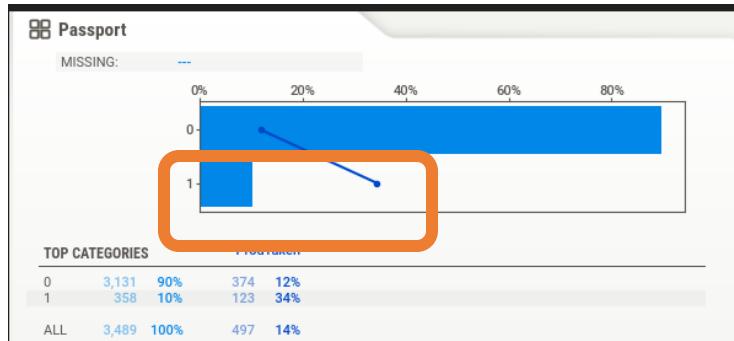
解法：5つの情報のうち2つ組み合わせ, AND演算

↳ パスポート有, 独身, 30歳以下, 子供なし, basicプラン

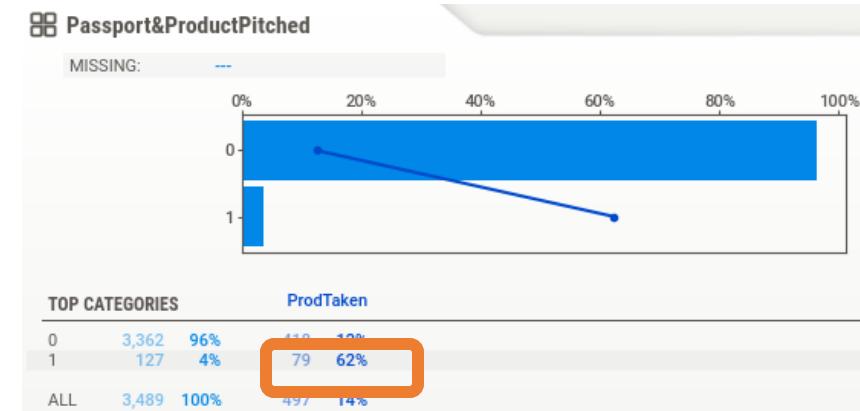


例：パスポートを持っている人は, 34%が成約している
⇒成約した人の特徴が取りやすそう

解法：5つの情報のうち2つ組み合わせ, AND演算



Λ (かつ)



成約した人が62%へ

1. 年齢の欠損を回帰予測
2. AND演算子による特徴の強調
3. 決定木とNNを組み合わせる

動機

NN系と決定木系のアンサンブル学習が良い
⇒異なる特徴をとらえ、汎用性を上げる

解法：4つのモデルを加重平均

決定木

- ① catboost
- ② lightGBM

NN

- ③ AutoML
- ④ Mamba[1]

$$① : ② : ③ : ④ = 1 : 2 : 1 : 4$$

→LBの精度が良い順に重みづけ

コンペを終えて

うまくいかなかったこと（うまくいきそうで）

仮定：若者はお金がない、中年は時間がない

具体案：20歳-39歳と40歳-60歳で分割

⇒ 年代ごとに明確に特徴が異なっていた

例：中年は同伴する子供が3人以上なら全員非成約

感想

機械学習の活用について考える良い機会だった

研究や他コンペと同時並行で充実した時間を過ごせた

ご清聴ありがとうございました