



**SIGNATE Cup 2024 夏**

**学生部門 2位解法**

saarrrrruu

# 目次

- ▶ コンペの概要
- ▶ 解法について
- ▶ コンペを終えて

01

コンペの概要

## ▼ コンペの概要

旅行会社の保有する顧客データ(属性や志向、営業担当との接触履歴等)を元に、旅行パッケージの成約率を予測

### 特徴量

id  
Age  
TypeofContact  
CityTier  
DurationOfPitch  
Occupation  
Gender  
NumberOfPersonVisiting  
NumberOfFollowups

ProductPitched  
PreferredProperty  
Star  
NumberOfTrips  
Passport  
PitchSatisfactionScore  
Designation  
MonthlyIncome  
customer\_info

### 目的変数

ProdTaken

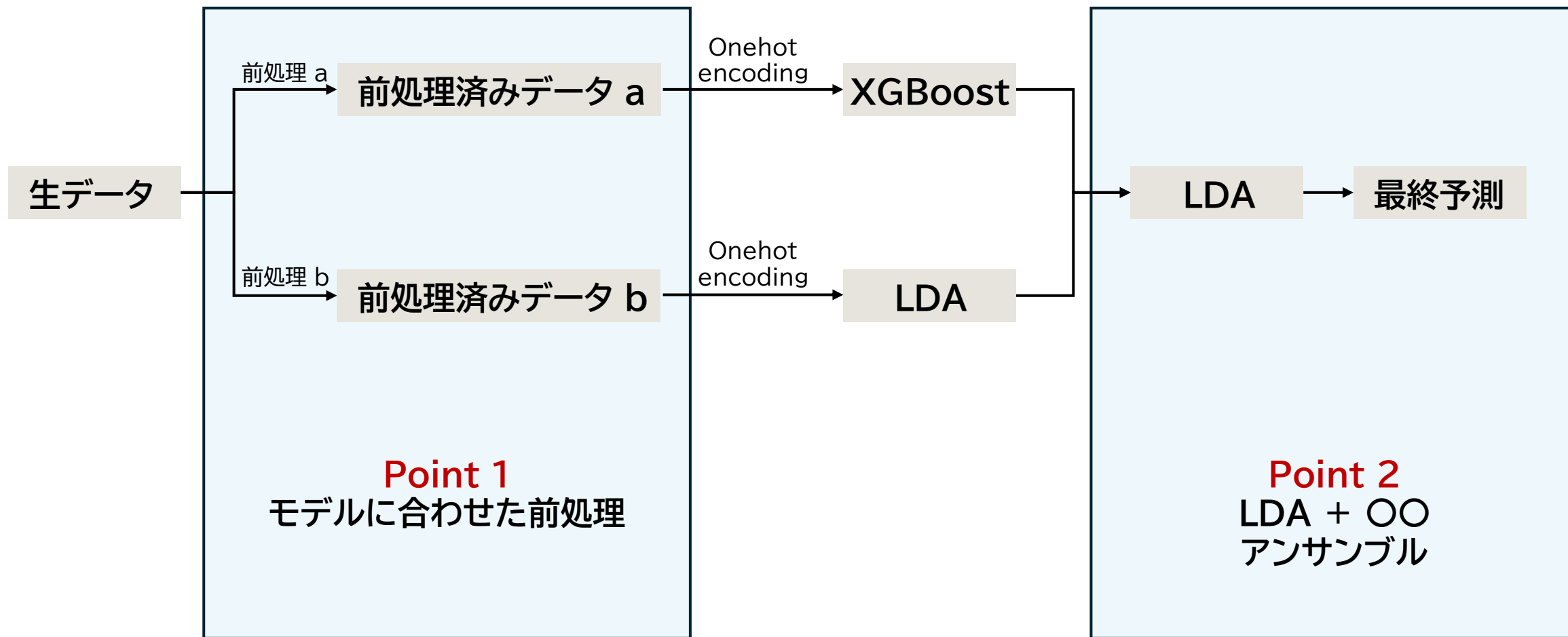
### 評価指標

AUC  
Area Under the Curve

02

解法について

## 全体像



# ▼ 前处理

## 名寄せ

ProductPitched

Basic



basic

MonthlyIncome

六十一歳



61

## 異常値処理

NumberOfFollowups

4000



4

## 単位統一

DurationOfPitch

2160秒



36分

## 欠損値補完

Age  
DurationOfPitch  
MonthlyIncome

平均値

TypeofContact

Self Enquiry

NumberOfFollowups  
NumberOfTrips

Missing

## ▼ 特徴量作成 (XGBoost)

customer\_info の分割

customer\_info → Marriage, Car, Children

未婚 車未所持 子供なし



## ▼ 特徴量作成 (LDA)

### customer\_info の分割

customer\_info → Marriage, Car, Children  
未婚 車未所持 子供なし

### 文字による特徴量作成

$\alpha$  は ProductPitched のデータに含まれるか?  
...

Basic → 0

basic → 1

### 四則演算

Age \* DurationOfPitch, Age \* MonthlyIncome, DurationOfPitch \* MonthlyIncome

$\frac{\text{DurationOfPitch}}{\text{Age}}, \frac{\text{MonthlyIncome}}{\text{Age}}$

# ▼ アンサンブル

## 1段目

train					test
学習用	学習用	学習用	学習用	検証用	

XGB, LDAで予測

## 2段目

train	test
1段目の予測を結合	1段目の予測の平均

LDAで予測      CV 0.8444623  
LB 0.8447259 (public)  
0.8445007 (private)

04

コンペを終えて

## ▼ まとめ

LDAは単体でも一定の精度を示し、  
アンサンブルすることによりさらに精度が向上した。

コンペ期間中にもっとやりたかったこと

- 文字ごと(キリル文字, ギリシャ文字等)の特徴量作成
- NN導入