

【モデル : LightGBM cv : 0.763 Omnicampus LB : 0.76679】

【特徴量について】

追加した特徴量

EXT_SOURCE_SUM_1 : $\text{EXT_SOURCE_1} + \text{EXT_SOURCE_2}$
EXT_SOURCE_SUM_2 : $\text{EXT_SOURCE_2} + \text{EXT_SOURCE_3}$
EXT_SOURCE_SUM_3 : $\text{EXT_SOURCE_1} + \text{EXT_SOURCE_3}$
EXT_SOURCE_DIFF_1 : $\text{EXT_SOURCE_1} - \text{EXT_SOURCE_2}$
EXT_SOURCE_DIFF_2 : $\text{EXT_SOURCE_2} - \text{EXT_SOURCE_3}$
EXT_SOURCE_DIFF_3 : $\text{EXT_SOURCE_3} - \text{EXT_SOURCE_1}$
EXT_SOURCE_AVG_123 : $(\text{EXT_SOURCE_1} + \text{EXT_SOURCE_2} + \text{EXT_SOURCE_3})/3$
EXT_SOURCE_AVG_1 : $(\text{EXT_SOURCE_1} + \text{EXT_SOURCE_2})/2$
EXT_SOURCE_AVG_2 : $(\text{EXT_SOURCE_2} + \text{EXT_SOURCE_3})/2$
EXT_SOURCE_AVG_3 : $(\text{EXT_SOURCE_3} + \text{EXT_SOURCE_1})/2$
EXT_SOURCE_MIN : EXT_SOURCE の最小値
EXT_SOURCE_MAX : EXT_SOURCE の最大値
DAYS_EMPLOYED_PREC : 'DAYS_EMPLOYED'/'DAYS_BIRTH'
INCOME_TOTAL_EMPLOYED_PREC : 'AMT_INCOME_TOTAL'/'DAYS_EMPLOYED'
CREDIT_GOODS_PRICE : 'AMT_CREDIT'/'AMT_GOODS_PRICE'
GOODS_PRICE_CREDIT_DIFF : 'AMT_GOODS_PRICE' - 'AMT_CREDIT'
ANNUITY_INCOME_PREC : 'AMT_ANNUITY'/'AMT_INCOME_TOTAL'
CREDIT_INCOME_PREC : 'AMT_CREDIT'/'AMT_INCOME_TOTAL'
PAYMENT_RATE : 'AMT_ANNUITY'/'AMT_CREDIT'
DAYS_BIRTH_OWN_CAR_AGE : $(\text{DAYS_BIRTH}/365) - \text{OWN_CAR_AGE}$
AGE : $\text{DAYS_BIRTH}/365$

削除した特徴量

FLAG_MOBIL, FLAG_CONT_MOBIL, REG_REGION_NOT_LIVE_REGION, LIVE_REGION_NOT_WORK_REGION, LIVE_CITY_NOT_WORK_CITY

全特徴量で学習した際、EXT_SOURCE が効いていることが分かったので、四則演算などで特徴量を増やしてみました。他の特徴量は「貸し倒れしにくそう」という観点で作成しました(年齢に占める就労期間が長い、総所得に占める年金支払額が少ないなど)。特徴量は合計 65 になりました。特に EXT_SOURCE_MIN や PAYMENT_RATE は精度向上に寄与しました。

【モデルについて】

モデルは LightGBM を使用し、optuna でハイパーパラメータチューニングしました。バリデーションは不均衡データのため StratifiedKfold を使用し、分割数は 8 としました。