

목표

01

CatBoost tutorials

02

Imbalance Data 처리

03

Categorical Data처리

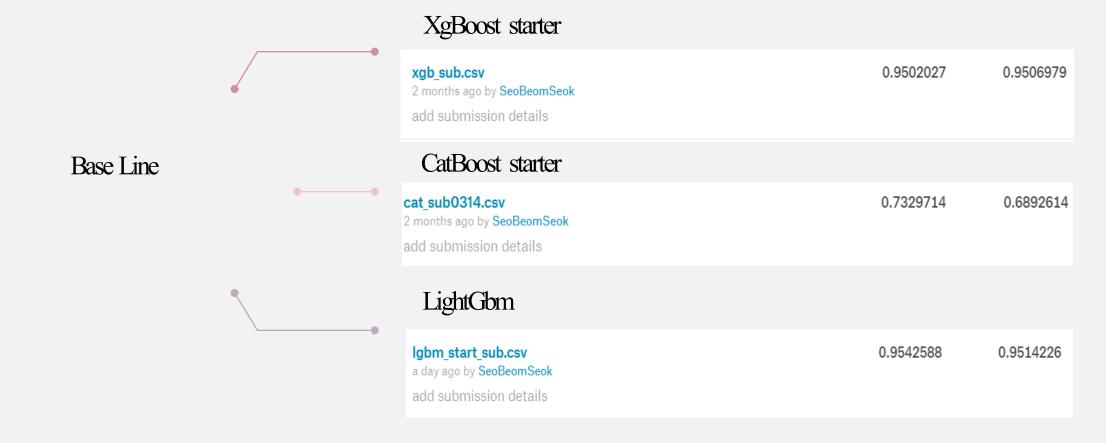
#### CatBoost tutorials

1. BaseLine

2. Classifier? Regreessor?

3. CPU? GPU?

#### Base Line



CatBoost Model

01 Classifier? Regressor? 02 CPU? GPU?

Parameter Tuning?

#### Imblance Data

UnderSampling? OverSampling?

#### Inblanced Data Sampling

01 Undersampling 02 Oversampling

- 1. Counter({0: 24842471, 1: 61419})
- 2. Counter({0: 24842471, 1: 24842471}): OverSampling(SMOTE)
- 3. Counter({0:61419, 1:61419}): UnderSampling(RandomSampling)

### Inblanced Data Sampling

# LightGbm basic

Igbm_start_sub.csv a day ago by SeoBeomSeok	0.9542588	0.9514226
LightGbm OverSampling		
labm_resample_sub.csv a day ago by SeoBeomSeok	0.9431151	0.9429953
LightGbm OverSampling		
Igbm_undersample_sub.csv a day ago by SeoBeomSeok	0.9472244	0.9462172

### Inblanced Data Sampling

#### Xgboost basic

xgb_start_sub.csv a day ago by SeoBeomSeok	0.9495495	0.9505427
Xgboost OverSampling		
xgb_resample_sub.csv a day ago by SeoBeomSeok	0.9322828	0.9345201
Xgboost OverSampling		
xgb_undersample_sub.csv 8 hours ago by SeoBeomSeok	0.9486490	0.9471463

# Categorical Data

Groupby?

# 이번 Kaggle 후기...

- 1. 평가함수를 작 확인하자
- 2. Pandas보다는 numpyanay 를 선호하자
- 3. Data 특성에 <sup>또</sup>·라 sampling 기법을 다르게적용
- 4. Categorical Data의 변수생성에 대한 경험

옵션



