

제1회 산학연계 공모전

: 머신 러닝 모델을 통한 분석

용돈별조

송민규, 백찬진

# Index

Round 1. Feature Engineering

Round 2. Modeling

Round 3. Ensemble - submissions

# Round 1 - Feature Engineering

## 공통 처리

### 1. Train column 제거

- 매니저최초가입일, 매니저최초서비스일, 매니저주소 제거.

### 2. 결측치 처리

- (청소교육이수여부, CS교육이수여부, 부재중여부) -> 0으로 채움.
- (매니저이동방법, 매니저사용휴대폰, 고객가입일, 결재형태, 접수시각) -> 최빈값으로 채움.
- (반려동물, 우선청소) -> 없음으로 채움.
- (평수) -> 30평대로 채움.

### 3. 매니저 생년월일 전처리

- train과 test에서의 데이터 형태가 달라 test 형태로 맞춤.

# Round 1 - Feature Engineering

## Pipeline 1 - catboost

Numeric Feature

Categorical Feature  
- One-hot Encoding

## Pipeline 2 - ExtraTree

Make Feature

- 매니저나이
- 매니저나이대
- 서비스시간

Numeric Feature

- PowerTransformer scaling

Categorical Feature

- Label Encoding

Feature Selection

- Percentile Selection(23)

# Round 1 - Feature Engineering

## Pipeline 3 - Kneighbors

Make Feature

- 고객가입개월수(str)
- 나이\*고객가입개월수

Numeric Feature

- PowerTransformer scaling

Categorical Feature

- Label Encoding

Feature Selection

- Percentile Selection(23)
- ABTest(13)

## Pipeline 4 - RandomForest

Make Feature

- 매니저나이
- 매니저나이대

Numeric Feature

- PowerTransformer scaling

Categorical Feature

- Label Encoding

Feature Selection

- Permutation Importance / ABTest(25)

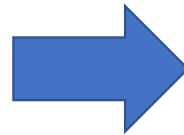
# Round2 - Modeling

## Pycaret

AutoML 방식의 Pycaret을 사용하여 모델링을 진행.

모델링 뿐만 아니라 하이퍼파라미터 튜닝 또한 Pycaret에 내장된 기능을 사용.

각각의 모델링 진행

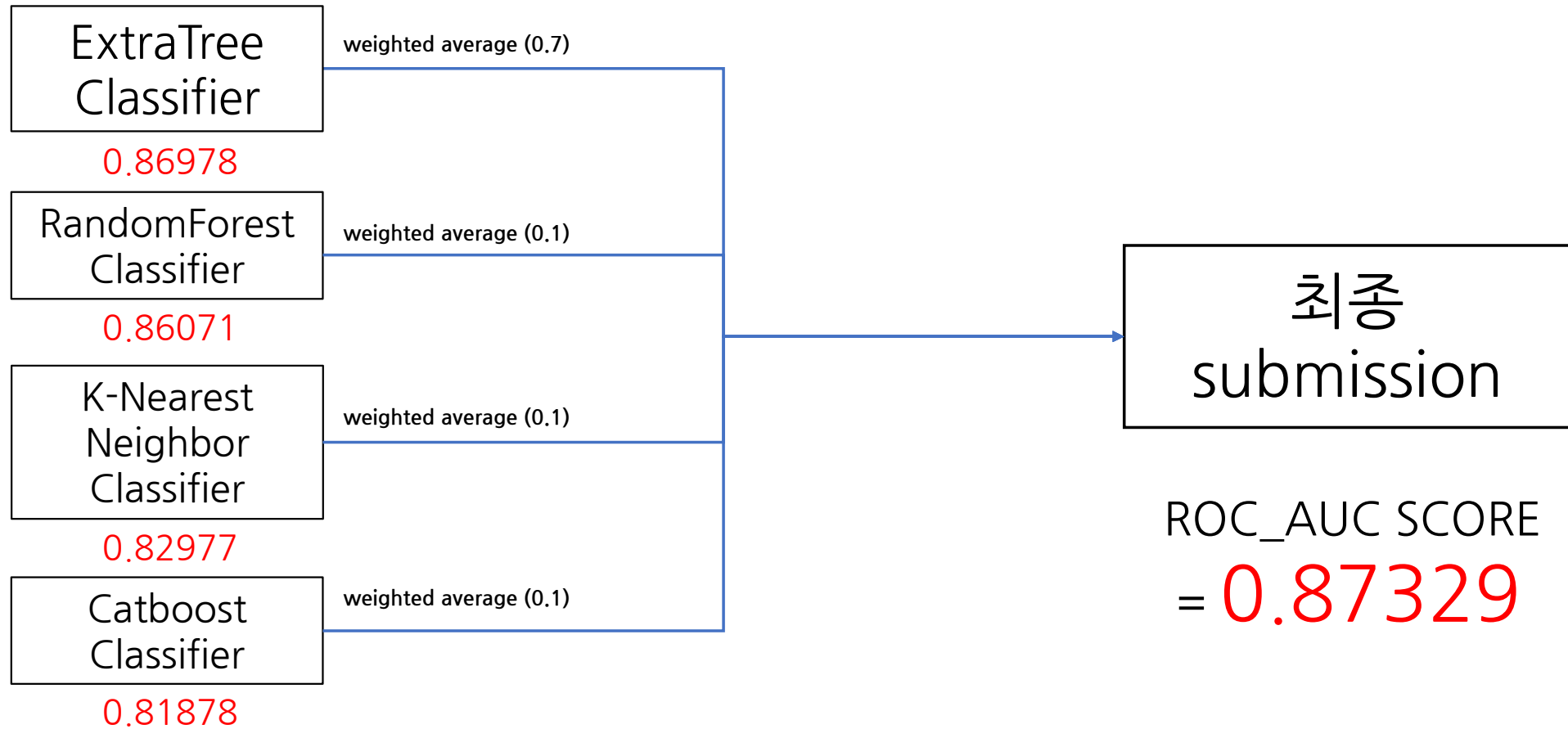


## Tuning

K-nearest neighbors Classifier,  
Catboost Classifier

두 가지 모델에 대해서 Pycaret 내장 기능인 `tune_model`을 사용.

# Round 3 - Ensemble all submission



Thank You