1. 전처리

train, test 데이터를 df로 concat하여 전처리 적용

**결측치**

occyp\_type에만 있는 결측치는 no job으로 대치

**피처클렌징/생성**

1인당 수익 : income\_total / family size

Age : days\_birth를 365로 나누고 절댓값

Employed : 양수는 0으로 취급 나머지는 그대로 -> 이 값을 365로 나누고 절댓값

month : begin\_month의 절댓값

ID : 고유값 파악 위한 코드 각 컬럼의 값을 str 형태로 바꾼 후 더해서 고유한 사람을 파악

2. 인코딩

**피처 구분**

데이터타입이 object인 것들은 categorical\_feats로, 아닌 것들은 numeric\_feats로 저장

**로그 변환**

income\_total 피처를 로그변환

**ordinary encoding**

카테고리컬 데이터를 ordinary encoding

**StandardScale**

income\_total을 제외한 나머지 뉴메릭 데이터를 정규화

3. 모델링

stratified 0 to 10 fold로 교차검증 -> Catboostclassifier로 모델학습/평가

\*캣부스트는 파라미터 튜닝을 일체 하지 않음

=> 이 부분은 공유된 코드들을 많이 참고하여 시도해 보았습니다. 그러나 완전히 똑같이 베끼기보다는 교차검증법도 바꿔보았고 캣부스트 파라미터 외에 다른 파라미터들도 변경해보았습니다. 독창성은 떨어져도 한도 내에서 노력했습니다!!!