-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-0014 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.11.04일자 |
| 팀원:신문혁,정기호,송승한,최지호 |
| <프로젝트 개요>   1. 신용카드 사용자 데이터 EDA 2. 신용카드 사용자 데이터 전처리 3. 모델을 통해 사용자 Credit 평가/분류 4. Credit 분류 값 결과분석 |
| <프로젝트 목적>  신용카드 신청자가 제출한 개인정보 데이터와 대금 연체 정도를 예측할 수 있는 인공지능 알고리즘 활용해 신용 점수를 산정함. 이 신용 점수를 활용해 신청자의 향후 채무 불이행과 신용카드 대급 연체 가능성을 예측 가능할 것으로 예상 중. |
| <실습 내용> |
| 신문혁: Encoding |
| 정기호: Encoding |
| 송승한: Encoding |
| 최지호: 웹기반 공부, 일지작성 |
|  |
| <회의내용> |
| 신: 어제 진행한 lgbm scaling과 오늘 진행한 ordinal scaling의 결과가 큰 차이가 없음.  송: clustering 진행해서 직업(occu\_type) 줄이고 카테고리화 해서 진행했으나 이 역시 큰 차이가 없음.  다른 팀과 유사하게 clustering 진행함. (유사한 직업류끼리 통합, 19>>4개로 축소). occu\_type family\_type , family\_size 등등 융합 및 분류함. numaric 을 scaling진행 예정.  신. child는 냅뜨고 decision tree 진행 시 변경 필요함 family\_type(couple/single)과 family\_num중에서 결혼 유무가 차이는 있는지 의문이 듬 (이혼은 동양 문화상 좋지 않게 보니 credit에 악역향 준다고 가설제시)  송: 모델링안에 전처리까지 다 있음  정: 모델링 계속 진행중이나 애로사항이 있음.  신: lgbm으로 가정 2개 진행중. ( 1.카테고리별 변수의 깊이(컬럼수 과부 방지/파생변수 너무 크지 않게) 1등변수에 파생변수 넣어 확인 중임. 2개 사이에는 현재로선 큰 차이 없음  신: ordinal 인코딩 시 둘 다 카테고리별 변수 인코딩 하면 변수들 사이에 크기에 대한 의미가 없음. (가족타입/직업 등)  정: get the me가 더 간단함. 이대로 계속 모델링 진행 예정.  정: 코드 리뷰중 catboost 사용한 1등 코드가 있음으로 참고하며 이를 토대로 모델링 진행하는 것이 좋을 듯함. |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자: 최지호 | 점검자 | 점검자 |
| 일자 2021년 11월 4일 | 일자 2021년 11월 4일 | 일자 2021년 11월 4일 |

|  |  |
| --- | --- |
| 실습 후기 | * 수상 코드 리뷰 및 분석 후 적용 시 유용할지 확인 필요. * 모르는 코드 있을 시 검색(구글)및 학습/복습 필요함 |
| 향후 계획 | * 모델링 진행 |