|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-005 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.10.22일자 |
| 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:Catboost/Stacking Ensemble 공부/정리 |
| 정기호: EDA 진행 + 전처리 |
| 송승한: EDA 진행 + 전처리 |
| 최지호: Stacking ensemble 공부, 일지 작성 |
|  |
| <회의내용> |
| 정:  -현제 begin\_month로 처리 중임. data user\_list, begin\_month를 최신화 된 걸로 뽑고 train해서 다시 뽑음.  -같은 작업으로 family\_size, credit, begin\_month가 다 같은 상태라 확인이 안되니 다시 train해야 할 듯  신:  -train 결과가 id=a고 계산 후 begin\_month의 mean값을 구한 후 변환한 결과를 절대값으로 넣은 결과가 a임. a/b값 구한 후 새롭게 형성한 user\_list에 넣음. 이를 begin\_month에 b로 넣으면 모든 8700명 중 user\_list 에begin\_month를 진행해보니 결과값이 3700명이 다수의 중복 값이 나옴. 그래서 begin\_motnh, mean\_month로 대체함. Last\_month를 추출해서 다시 넣을 수도 있음  -abs(절대값)사용이유도 원래 -(음수)이기에 대체했음. 오후에는 이를 통해 더욱 EDA 진행 예정.    정:  -begin\_month 확인후 신규카드개설(begin\_month=0인사람) 숫자fmf 보니 214명임. 여기서 user\_list를 in값으로 대체함.  -추가로 신규개설한 사람이 1번째인 사람들 수가 29명. Begin\_month 분포도에서 봐야 할 건 credit의 begin\_month랑 연관성 있는가 (begin\_month(=in)  -대체적으로 신용등급 좋은 사람(0/1등급)은 분포가 많지 않다.  Category(begin\_month를 10개월 순으로 (0-10, 10-20, 20-30, 30-40개월 순) 카테고리 5개로 분류)하니 부여됨  unnamed 0인사람을 drop 하니 null값이 없음.  -value.count 통해 category\_ month로 credit을 진행 categorical\_months>>credit(0) 외 등급을 정규화 한 결과 등급은 큰 변화는 없음… 의미가 있다하면 0-1등급 애들이 상승함. 사용도 낮음으로 고민중  송:  ***for i in tqdm(range(len(train))):***  ***a = train.iloc[i]['ID']***  ***b = abs(int(train[train['ID']==a]['begin\_month'].mean()))***  ***user\_list.loc[(user\_list['ID']==a), 'begin\_month\_mean'] = b*** 로 실행  -credit 0인애들/아닌 애들 EDA한 결과가 비슷함    신:  -gradient\_bust 진행하다 강사님에 질문해 미니 강의 들어 부스팅은 (random\_forset light gboost) 어느정도 이해 완료.  -월요일쯤 설명회 실행 예정. |
|  |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자: 최지호 | 점검자 | 점검자 |
| 일자 2021년 10월21일 | 일자 2021년 10월21일 | 일자 2021년 10월21일 |

|  |  |
| --- | --- |
| 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 | * 2021.10.29 중간점검 전까지 발표자료 및 자료정리 * XGboost, Catboost, S.Emsemble 공부 후 설명 및 자료공유 * EDA 및 전처리 진행 * 2021.10.25 회의 진행 |