-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-007 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.10.26일자 |
| 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: StackingEnsemble 공부/정리 |
| 정기호: EDA 진행+ 전처리 |
| 송승한: EDA 진행+ 전처리 |
| 최지호: Stacking ensemble공부, 일지 작성 |
|  |
| <회의내용> |
| 정:    어제 작업했던 함수를 도입해 진행하면 이렇게 나옴  정: 함수작업 계속 진행중이며 현재 함수 6개 제작완료.  송: EDA 진행중이나 크게 눈에 띄는 게 없어서 걱정임. Occupy를 어떻게 진행해야 할지 확인 중.  정: 코드 정리 중(삭제, 나이처리 등)이며 깃허브에 이후 업로드 예정.  boxplot 이상치 보려 data전처리 하기 전에 생성함. 성별차(gender) income\_total(data(userlist)활용)  formatting이용 해서 목록 변환하면 바뀌면서 나옴.      occupy된 nan(type) 없는 사람들(무직) 된사람들을 확인을 위한 boxplot(일종의 EDA)  feat1. feat2. feat3 def 확인위해 만든 graph        그 다음 column\_type 제작했으나 항목이 너무 많아서(def) dataset 생성함.    오늘은 함수만 제작 예정  신: 부스팅 공부중  교차검증 맞는지 확인(<https://m.blog.naver.com/ckdgus1433/221599517834> 참고)  정: heatmap 진행중이라 시간 좀 걸림  Incomes\_total이 14개 정도 있어서 14\*14로 나옴 여기서 credit만 뽑고 싶었는데 sort\_value가 떠서 dataset입력후    barplot을 mean값으로 도출    송: 별책지를 어떻게 처리할 것인지 고민중. 무직을 분류했더니 1508명이나 나와서 정리중. 나머지는 정기호씨 함수로 시험 진행 예정.  신: 분류법을 3가지 확인됨.   1. Training set과 test set을 안 나누고 한꺼번에 검증 2. Training set과 test set을 나눈 뒤 training set으로 검증 3. Training set과 test set을 나눈 뒤 training set으로 실험한 후 test set으로 검증   진행은 3가지중 2번째로 진행예정. |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자: 최지호 | 점검자 | 점검자 |
| 일자 2021년 10월26일 | 일자 2021년 10월26일 | 일자 2021년 10월26일 |

|  |  |
| --- | --- |
| 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.Emsemble공부 후 설명 및 자료공유 * EDA 및 전처리 진행 |