-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-0016 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.11.08일자 |
| 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, 전처리 정리 |
| 정기호: Encoding |
| 송승한: Encoding |
| 최지호: 웹 기반 공부, 일지작성 |
|  |
| <회의내용> |
| -flag\_mobile drop안함, ability 기입>> scaling 안함. Days\_birth안함>>, 양수 0 대치 occu\_type 0처리, id 그대로, card 처리 안 함. Family\_size로 대치  ->>child\_num y/n, catboost해서 log loss가 0.73~ 시간은 별로 안걸림. (1분이하)  -catboost를 사용해서 진행이 낮게 나왔을 수도 있음. 위 데이터를 stacking emsemble로 실험  (RF, 서포트 벡터 머신, svm, lightgbm, catboost. logistic 최종모델(최종모델)) 사용  ->>결과 0/75-0.73.  -잔처리 중 베이스라인 코드 안 본것이 실점이긴 했으나 공부가 됨  -begin\_month 이용해 처리시((days\_month 등등) 포함한 상태로 처리하면 결과가 변질되어 나오니 begin\_month는 제외됨으로 예상.  - days\_employed days\_month 에서 begin\_month를 drop하고 나머지는 absolute 절대값 진행.  0값 이상 치환 days\_employed days\_birth를 days\_birth제외 후 결과 관찰중  ->>0.73나옴      -Income\_mean 실행 유무 차 결과 나옴  -Begin\_month 제외 안하기로 |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자: 최지호 | 점검자 | 점검자 |
| 일자 2021년 11월 8일 | 일자 2021년 11월 8일 | 일자 2021년 11월 8일 |

|  |  |
| --- | --- |
| 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 |