-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-0013 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.11.03일자 |
| 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> |
| -카드별 최신 결제 날짜가 다른데 가장 최근 갱신된 결제일로 날짜 변경 중(최신일자(credit)예측) 함수 성능향상 위해서 타 column 추가  -mode(days\_employed, months\_total 등등) 새 column 제작한 것 중 (days\_total/year\_total 등등) 추가  -log써서 양 수 추출 후 unique, 다시 log함수 사용해 for문 돌림. 값을 추출하지 않고 log 사용시 변화할 수 있는 방법이 있지 않을까 고민함  -jobless 계산 시 일을 한번도 하지 않은 사람(과거~지금(파일내)이 증가함 (days\_employed)  -log 진행되기가 더딤. 점점 추가해야 할 항목이 증가 중. 하지만 이 항복들이 주요 문제해결에 관계 없을 가능성이 크다.  -jobless=은퇴한 사람? nan/data= 은퇴한 사람들  -employed mode 이용해서 직업 column에서 (days\_emoplyed=begin\_month)로 지금도 일하고 있는지 알아봄  -jobless 안에서 은퇴한 사람들을 확인할 필요 있음. pensioner이 있는 사람들 = 은퇴한 노인들(retired)  -주부= (jobless에서 gender value.count 로 분류)  -jobless =retired로 판명됨. 수정 개시  -현재로는 다른 파생변수 생성시키는 것이 제일 중요함  -날짜 관련된 파생변수 설정이 중요함. 이 근거는 현재까지 모델링(random forest, catboost)을 진행 시 feature importance의 수치(begin\_month, days\_birth, days\_employed)가 제일 높음 |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자:최지호 | 점검자 | 점검자 |
| 일자 2021년 11월 3일 | 일자 2021년 11월 3일 | 일자 2021년 11월 3일 |

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