-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-009 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.10.29일자 |
| 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: ensemble 공부자료 공유 및 강의 |
| 정기호: 전처리, |
| 송승한: EDA 진행, 전처리 |
| 최지호: 일지작성 |
|  |
| <회의내용> |
| 신: Stacking fold 주변으로 좀더 리서치함. 중간발표 진행 어떻게 해야 할까?  송: nan값 어떻게 해야 할까? 난 그냥 no data로 진행해야 할 듯  정: 오늘은 전처리 진행 후 정리해야 함  정: 카드개수 대치 완료  정: work\_year 말고 work\_month\_total 새로 생성(age\_month\_total이랑 같음)  Income\_mean(개인당 소득) card\_num과 관련 있는 문항들 대처 (begin\_month/credt) begin\_month/credit 같은 항목들은 대치함수로 표현 19초과하는(20 이상) 목록들 실현  -stacking emsemble 관련 논의  전체적인 흐름은 catboosing/stacking emsemble을 실행한 한 후 마지막에 catboositng 사용할 것으로 예상함  정: 모델링 알고있으니 내일 발표후에 돌려보고 같이 이해하는 방향으로 더이상은 전처리 진행할 필요 있나 싶음(대부분 다 대치 완료) |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자:최지호 | 점검자 | 점검자 |
| 일자 2021년 10월28일 | 일자 2021년 10월28일 | 일자 2021년 10월28일 |

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