|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-002 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.10.19일자 |
| 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:private 1등 코드 리뷰(catboost), EDA 진행  <https://www.dacon.io/competitions/official/235713/codeshare/2768?page=1&dtype=recent> |
| 정기호:private 2등 코드 리뷰(stacking ensemble), EDA 진행  <https://www.dacon.io/competitions/official/235713/codeshare/2757?page=2&dtype=recent> |
| 송승한private 4등 코드 리뷰(XGBoost, LightGBM), EDA 진행  <https://www.dacon.io/competitions/official/235713/codeshare/2755?page=2&dtype=recent> |
| 최지호: private 5등 코드 리뷰(catboost), catboost복습, 일지 작성, 깃허브 공부  <https://www.dacon.io/competitions/official/235713/codeshare/2755?page=2&dtype=recent> |
|  |
| <회의내용> |
| 강: 하등 60퍼가 90퍼가 아님. income변수 자체가 credit이라 상관관계가 크진 않다  데이터 credt컬럼의 0 1 2 3가지를 다 dataframe으로 뽑았는데 income total을 히스토그램으로 표현함  0~2등급 히스토그램으로 나오나 유의미한 차이가 없음  신: 소등급 높게 나와야 하나 다 동등하게 나오나?  강:0등급은 약간 소득이 높긴 하나 1/2등급 차이는 많이 없다->>상관성 없어보임  직업별로 어떻게 다른가 확인중임 시각화는 아직 14~27정도라 지금 당장은 확인하기 힘듬  동일한 사람인데 다중소유 카드 제외해야 히스토그램 할수 있다 썰 지우고 평균?으로 해야할 듯  송: 업무폰을 갖고 있음 필수인력이라 등급이나 돈을 잘 벌지 않을까 했는데 결론적으로 상관없음  강: 업무폰 소유유무 플래그 다 1이라 삭제  강: family\_size는 봐야할 듯 titanic처럼 자식 수/형제자매 수를 family number로 묶었는데 이처럼 묶을 수 있지 않을까  송:근데 children\_size 볼거면 이거도 8000명만 뽑아서 진행해야 하나  강:8000명 어떻게 뽑아야 하느냐는 1등 코드에서 나온것처럼 개인정보 1~4로 표현, 소득표현 해서 진행해야 할 필요 있음  신: EDA 진행 시 전체데이터는 진행하고 웬만한 데이터는 중복 값 빼고 가족 수는 1등코드 관련으로 파생변수 만들 시 가족 별 소득(가족 별로 소득 나뉨)  children\_number을 가족에 포함됐다 하며 자녀수, 살아있는 일수(days\_birth) 일한 날자(work\_days) 삭제함  강: employed children\_size+children\_number days\_employeed days\_birth  flag mobile 삭제(값이 다 1이라 불필요 ) work\_phone phone, email 고려 카테고리 변수로 income total을 1~10으로 분류 후 8000명 진행  신: 이부분에 추가-income total 숫자 데이터 numeric column중 income total은 로그 스케일링으로 진행한 후 나머지는 standard scaling으로 변환  강:add type family type house type 0~5로 변환 해야한다?  신: catboost 확인했는데 boost쓰는 tree관련은 다 category 쓰나 1등코드에서 카드변화를 0~10으로 하나 예시 볼때 sunny cloudy처럼 변환해 credit이 0 1로 예측함. 예네는(1등)변환 안했으나 우리는 변환했다 이에관련 이해 필요함  강: data는 object니 5개로 나누나 글자로 쓰면 낭비 아닌가 생각해서 숫자로 변환 생각함 숫자가 1만개도 안되니 필요성 잘 모르겠지만  강: 그리고 대략적인 계획 필요함. 모델링은 마지막 주에 진행 전처리는 2주차부터 EDA랑 같이 하면 최적이나 계획대로 안될 가능성 높다 |
|  |
|  |
|  |
|  |
|  |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자최지호 | 점검자 | 점검자 |
| 일자 2021년 10월18일 | 일자 2021년 10월18일 | 일자 2021년 10월18일 |

|  |  |
| --- | --- |
| 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 | 2021.10.20까지 각자 담당 코드리뷰 한 후 회의를 통해 차후 프로젝트 방향성 조정 |