-

|  |  |  |
| --- | --- | --- |
| 실습제목 | 신용카드 사용자 연체 예측 AI 경진대회 | 일련번호4-0012 |
| 실습목적 | 신용카드 사용자들의 개인 신상정보 데이터로 사용자의 신용카드 대금 연체 정도를 예측 | |

|  |
| --- |
| 21.11.02일자 |
| 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: 일지작성 |
|  |
| <회의내용> |
| 송: 기호씨가 보내주신 함수로 데이터 정재예정  내일 모든 데이터 진행해야 함  송: 정기호씨가 준 코드를 인코딩만 하고 scaling 하지 않은 체 진행해보니 0.716 나옴 (public 0.716 private 0.713 모델링은 랜덤 포레스트, log loss는1.xx(낮을수록 좋음) 운이 좋은 것 일수도 있으니 scaling다시 해봄  정: training set과 test set 다르다. Begin\_month에 있는 요소들 업데이트 된 시점  보통 시간이 지날수록 credit이 좋아하야 한다는데 이런 상태면 age\_plot 수정해야 할 필요 있음. begin\_month 기준으로 -3개월 -46개월 진행해야 함 (age\_total work\_total에 age\_month를 전부 빼고 -4를 다 0으로 치환 필요함)  신:  -begin\_month 시점별로 다 바꿈 이유는 한 번도 악화될 수 있음.  credit이 else: 에서 걸리면 +1해야함 (전체 format\_change에서 한번이라도 악화되면 1) begin\_month가 저하되는게 좋음.  -다른 걸로 바꾸게 되면 work\_year/age 둘 다 해보고 정확도 올라가면 기호씨 모델대로 진행하고 값이 잘 나오면 다른 정보를 찾아내야 함  정: 중복 값이 생각보다 크다(8000명) 나와야 하는데 1천명 줄음  Python 파일 저장 안하고 진행하면 안 되니 파일 새로 열때마다 다른 효과 있음.  -path는 위에 있음으로 절대경로를 입력하면 되니 삭제금지  정: begin\_month 문제 해결해주고 승헌씨 code 업로드 요청  송: 속상한소식: 아까 상태에서 path 오류가 있어서 고치고 진행시켜보니 점수가 더 나쁘게 나옴  송: 내일 ensemble 모델 무조건 돌려봐야 할 듯  최대한 많은 데이터셋 두고 돌려야 한다  정: 내가 만든 데이터셋으로 random forest 돌려보니 안됐다. (점수가 1.14)  신: test를 저장한 게 안 뜸. 그래서 주석처리 하고 for문은 df에서 뽑히니까 if로 train 저장/test저장 따로 함  신: 어차피 days\_birth는 시작 시 begin\_month\*30 한 후 빼면 뽑히니 influenza는 문제니까 양수 값을 영어로 변환하고 day\_influ는 계산해야 하니 mode에서 바꿔서 진행시키면 될 듯.  송: loss 얼마나 나옴? (scale 안 넣어서 0.86)  1등코드는 base 0.9 나왔는데 이쪽은 1.0대 나와서 다시 진행해야 하는지 고민중 (random forest가 문제인가 고민중) |
| <대회 진행 일정> |

|  |  |  |
| --- | --- | --- |
| 기록자:최지호 | 점검자 | 점검자 |
| 일자 2021년 11월 2일 | 일자 2021년 11월 2일 | 일자 2021년 11월 2일 |

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