Github로 정리

1. 기획
2. 정리 -> 최지호 (일지)
3. PPT 및 발표
4. 코딩
   1. <https://www.dacon.io/competitions/official/235713/codeshare/2768?page=1&dtype=recent> 코드 리뷰 (수요일 저희 회의시간까지 진행)
      1. 어떤 코드를 리뷰할껀가? (모델을 얘들이 어떤것을 선택했는지?)
         1. 오늘 점심먹고나서 정하는 것으로 (링크, 몇위?,어떤 모델 썻는지?)
   2. EDA (탐색적 데이터 분석) : 4명 구성원이 EDA 주제를 설정하는 것이 목표
      1. Label (신용도)를 결정하는 feature가 무엇이 있는지? -> feature 결정
         1. 상관성 (heatmap)
         2. Pie, line, bar plot
   3. 전처리 (Nan, missing value)
      1. 소득, 교육, 가족관계(이혼, 미혼, 재혼,…) -> 전처리
   4. 모델링
      1. Classification (랜덤포레스트, 의사결정나무,….)
      2. 여러 모델 중에서 결정해야 함 (최종 모형을)

**: EDA – model -전처리**

**모델에 맞는 전처리를 진행하는 것.**

신문혁

실질적으로는 1번이 좋음. (모델링 지식이 많이 있지 않지만 코드가 존재함. 코드를 리뷰한 다음 모델에 대한 지식을 쌓고 전처리할 수 있다. 그리고 모델링을 여러개 설정하면 보완이 될 것 같다)

1등 코드 같은 참고 코드가 있기 떄문에 이게 유리할 수 있음.

1등 : catboost

<https://www.dacon.io/competitions/official/235713/codeshare/2768?page=1&dtype=recent>

2등 : stacking ensemble

<https://www.dacon.io/competitions/official/235713/codeshare/2757?page=2&dtype=recent>

3등 : stacking ensemble

<https://www.dacon.io/competitions/official/235713/codeshare/2788?page=1&dtype=recent>

4등 : XGBoost, LightGBM

<https://www.dacon.io/competitions/official/235713/codeshare/2755?page=2&dtype=recent>

5등 : catboost

<https://www.dacon.io/competitions/official/235713/codeshare/2755?page=2&dtype=recent>

6등 : stacking ensemble

<https://www.dacon.io/competitions/official/235713/codeshare/2763?page=3&dtype=recent>