**세미프로젝트 기획안**

**과목명: 실전 빅데이터 분석 프로젝트**

**기획안 작성일자 :**

|  |  |
| --- | --- |
| 팀 명 | 3조 : 케세미 |
| 프로젝트 조직  (구성원 및 역할) | 팀장 : 정양섭  부팀장 : 박예원  팀원 : 주혜린, 백혜수, 송상민, 최준섭 |
| 프로젝트 주제 및 개요  해결하고자 하는 문제  최종 산출물의 청사진 | 케글대회를 통하여 실데이터 분석과정과 데이터 가공(데이터 전처리) 을 경험하고 최적의 모델링을 만드는 과정을 익히고자한다. |
| 프로젝트 수행 방향  수행 방법/도구 | # 주피터 사용  1. 평가 지표 파악을 통해 구체적으로 문제 이해  2. EDA 분석 과정을 통해 데이터를 분석하고 결과를 내는 과정에 있어 지속적으로 데이터에 대해 탐색 후 데이터 시각화를 통해 주요 변수를 선별  3. Feature engineering, 중복 데이터 제거, Label Encoding을 이용해 문자 데이터 변환, hold out의 기준점을 선정하는 데이터 전처리 과정 진행  4. 여러가지 예측 모델(xgboost, light gbm 등)을 훈련하고 정확도가 가장 높은 모델 선택  5. Hyperparameter optimization 후 성능 검증 진행, 최종 결과 예측 후 보고서 제출 |
| 프로젝트 추진 일정 | 7 / 5 ~ 7 / 6 문제 이해 (평가 지표 파악)  7 / 7 ~ 7 / 10 데이터 분석 및 데이터 전처리  7 / 6 부터 격일로 대면 스터디  7 / 11 ~ 7 / 17 적합한 모델 찾기, 적용 및 검증  7 / 17 ~ 7 / 20 하이퍼 파라미터 최적화 , 발표 준비 |