**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.  or  과거의 판매량 데이터를 통해 모델을 훈련 후 미래의 판매량을 예측해야 하는 문제이다. |
| 프로젝트 수행 방향  수행 방법/도구 | 파이썬 툴: 주피터 노트북  1.팀원들과의 상의로 평가 지표 파악 및 문제 이해  2. EDA 분석 과정  pandas의 info 등을 이용한 데이터의 타입 및 결측값 확인 matplotlib의 boxplot이나, seaborn의 pairplot 등을 이용하여 변수간의 상관관계, 이상치, 주기성을 파악  3.위의 EDA 분석을 기초로 Feature engineering, 중복 데이터 제거, Label Encoding을 이용해 문자 데이터 변환 등 데이터 전처리 과정 진행  4. 여러가지 예측 모델(xgboost, light gbm,ARIMA등)을 테스트 데이터와 검증 데이터를 EDA에 따라 분활하여 테스트 데이터를 훈련 및 적용 후 적절한 방법으로 검증 데이터를 이용하여 검증(홀드아웃 , 교차 검증)  5. Hyperparameter optimization를 하여 성능 향상, 최종 결과 예측 후 보고서 제출 |
| 프로젝트 추진 일정 | 7 / 5 ~ 7 / 6 문제 이해 (평가 지표 파악)  7 / 7 ~ 7 / 11 데이터 분석 및 데이터 전처리  7 / 6 부터 격일로 대면 스터디  7 / 12 ~ 7 / 17 적합한 모델 찾기, 적용 및 검증  7 / 17 ~ 7 / 20 하이퍼 파라미터 최적화 , 발표 준비 |