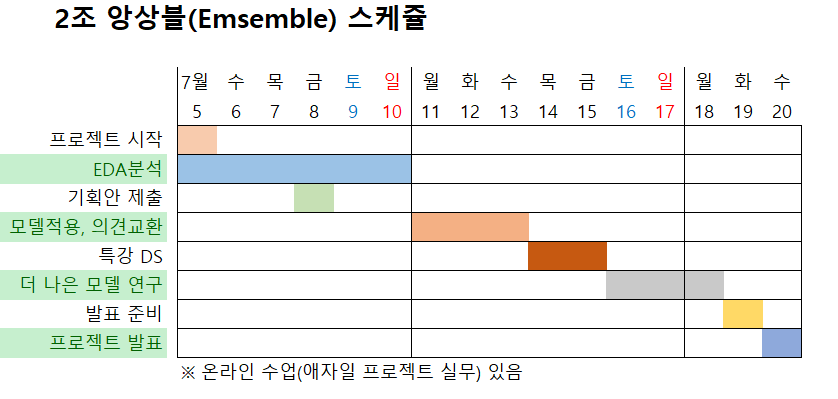
**세미프로젝트 기획안**

**과목명: 실전 빅데이터 분석 프로젝트**

**기획안 작성일자 : 2022-07-07**

|  |  |
| --- | --- |
| 팀 명 | 2조 : 앙상블 (Ensemble) |
| 프로젝트 조직  (구성원 및 역할) | 팀 장 : 노도영 LightGBM  부팀장 : 권희연 Random Forest  팀 원 : 신상연 Linear Regression  : 신상연 KNN  : 박주영 XGboost |
| 프로젝트 주제 및 개요  해결하고자 하는 문제  최종 산출물의 청사진 | 캐글 대회의 실제 현실데이터를 통해  데이터 분석, 가공 능력을 높이고,  최적의 모델링을 만드는 과정을 체득 |
| 프로젝트 수행 방향  수행 방법/도구 | 1. 실제 데이터의 EDA분석을 통해 데이터 구조를 파악, 시각화 2. 예측에 필요한 변수 추출 및 생성 - threshold 설정의 논거 준비 3. 기초 모델(다중회귀분석) 훈련 및 검증 - train/test data 분류 방법 준수 4. 예측값의 정확도가 다소 떨어질 것으로 예상되는 바,   더 나은 학습 모델을 찾아보고 연구 및 적용 - 모델 성능 평가 바탕의 threshold 피드백 반영   1. 각자 맡은 모델을 학습하고 앙상블 기법을 통해 최적의 모델링을 도출 2. RMSE 를 통한 모델 성능 평가   프로젝트는 Jupyter Notebook을 주로 이용하며,  경우에 따라 Kaggle Notebook을 이용 할 수 있음 |
| 프로젝트 추진 일정 | **07 / 05** 프로젝트 일정 수립, 방법 논의  **07 / 06 ~ 08** 데이터 EDA분석  **07 / 09 ~ 13** 첫 모델 적용 및 결과 검증  **07 / 14 ~ 18** 더 나은 모델 연구 및 적용  **07 / 19 ~ 20** 프로젝트 발표, 종료 |



〇 버전이력

# v1.4 – 박경민님 자진탈퇴