**세미프로젝트 기획안**

**과목명: 실전 빅데이터 분석 프로젝트**

**기획안 작성일자 : 2023. 01. 12**

|  |  |
| --- | --- |
| 팀 명 | 2조 |
| 프로젝트 조직  (구성원 및 역할) | 조장 : 김의석  조원   * 김종선 * 차한솔 * 이상엽 |
| 프로젝트 주제 및 개요  해결하고자 하는 문제  최종 산출물의 청사진 | Kaggle : Predict Future Sales  상품의 미래 판매량 예측 |
| 프로젝트 수행 방향  수행 방법/도구 | 1. EDA   - 김의석 train  - 김종선 items  - 차한솔 cats  - 이상엽 shops  2. 데이터 전처리  - 이상치 및 문자열 정리  - 데이터 병합  3. Feature Engineering  - 기초 통계량 및 상관관계 확인  - feature 선택 및 lag data 생성  4. 모델링  - 모델 분담  1) 김의석XGBoost  2) 차한솔 lightGBM  3) 이상엽RandomForest  4) 김종선SVC  - 하이퍼파라미터 튜닝  5. 평가  1) RMSE  2) 1차 평가 후 4, 5번 반복 수행 |
| 프로젝트 추진 일정 | 1. EDA : 1/9~ 1/11 2. 데이터 전처리 : 1/12 ~ 1/13 3. Feature Engineering : 1/16 4. 모델링 및 평가 : 1/17 ~ 1/18 5. 최종 발표 준비 : 1/19 6. 최종 발표 및 제출 : 1/20 |