

3 조 커리큘럼

진행방식

1. 각 주차별 스터디 주제를 바탕으로 개인 공부 진행
2. 스터디 주제에 대한 개인 공부 설명 자료를 작성하여 공유

- 설명 자료는 '[커리큘럼 참고사항](#)'을 토대로 작성
- 작성 후 dropbox 등 링크로 공유
- 공유 기한은 추후 협의 예정

3. 스터디 진행

- 2 에서 공유한 개인 공부 설명 자료를 바탕으로 온라인 화상회의로 3 조 스터디 진행
- 화상회의 날짜는 추후 협의 예정

4. 전체 온라인 화상회의(목)에서 조별 진행사항 공유

2 주차 목표 (3/21 수정)

- 스터디는 상시 자료 공유 방식
- 목요일 19:00~19:30 조별 정기 회의 (전체회의 20:00~)
- 필요시 추가 회의 진행 가능

측정지표

1. logloss 의 수학적 의미
 - Numpy 에서 로그로스 함수
2. 다른 측정지표와의 차이

우승자 코드 분석

1. 개념
 - 모델 : Random Forest, Xgboost, LightGBM
 - 하이퍼 파라미터 튜닝 : K-fold & Random seed
 - 예측 방법 : Hard-voting 방식, Soft-voting
2. 분석
 - 전처리 코드 분석

참고자료

- [데이터 설명 및 데이터 불러오기 - \[데이콘 원자력 상태판단 알고리즘 대회\]](#)
 - [log loss 에 대해 알아보자 - \[데이콘 평가산식\]](#)
-

3 주차 목표(3/26 수정)

~토(3/28, 21:00)

1. 데이터 전처리
 1. 불필요한 피처를 어떻게 판단할것인가?
 2. 결손값에 대해 대체할 대표값은 어떻게 선정할 것인가?(최빈값, 평균 등)
2. 분류기 (모델)
 1. 소프트 보팅 사용을 전제
 2. 어떤 분류기를 사용할 것인가? (XGBoost, LGBM 등)

~목(3/26, 21:00. 이후 목표는 28 일 회의 때 구체화 할 예정)

1. 앙상블
 2. 검증
 3. (Bayesian Optimization)
-

4 주차

주제 : 1. 모델 고도화 ([Level 4](#))

참고자료

1. [우승자 코드](#)
2. [우자 코드 설명자료](#)