



Code States AIB(AI Bootcamp)

수강생 프로젝트 기획서

수강생 이름 : 백송하
시작일 : 2021. 12. 10
프로젝트 : 프로젝트1

1. 프로젝트 주제

스타크래프트2 승패 정량적 분석

2. 프로젝트 목적

실시간 전략 게임의 특성상 게임내 특정 활동의 횟수보다
그 활동의 방식이 더욱 승패를 좌우한다고 할 수 있지만,
정량적 분석이 필요 없다고 할 수도 없다.

게임내 자원관리, 유닛 생산량, 특정 활동의 횟수, 이런 것들은
플레이어가 실시간으로 세우는 전략을 얼마나 어떻게 실행하고 있는냐,
즉 플레이어의 '피지컬' 을 판단할 수 있는 지표가 되기도 한다.

따라서 이런 지표들을 선정하여 플레이어들 스스로
실력을 상승시키는데 선택과 집중을 할 수 있게 돕는다.

3. 프로젝트 가설, 예상결과

일꾼유닛 생산수가 승패에 미치는 영향이 가장 클것이다
건설된 건물수가 승패에 영향을 미칠것이다
일반 유닛 생산수가 승패에 영향을 미칠것이다
플레이어가 유닛에 내린 공격명령이 승패에 영향을 미칠것이다

4. 프로젝트 분석방법

Data wrangling을 통해 관심 특성별 승률차이를 살펴본다
Gradient Boosting 모델과 Random Forest Classifier 모델을 통해 승패를 예측해본다
Feature importance, Permutation Importance를 통해 승패에 영향을 끼치는 특성을 확인한다
머신러닝 분류모델 평가방법인 confusion matrix와 이의 평가지표를 통해 분류모델의 성능을 확인한다
K-fold Cross Validation, Randomized Search Cross Validation 을 통해 모델의 성능을 향상해본다

5. 프로젝트 언어, 환경

Python ver. 3.8
Pandas ver. 1.1.5
Numpy ver. 1.19.5
Scikit-learn(sklearn) ver. 1.0.1
Matplotlib ver. 3.2.2
ELI5 ver. 0.11.0
xgboost ver. 0.90

6. 프로젝트 결과물

발표 영상
분석 및 모델링을 담은 Jupyter Notebook 파일
Test 데이터 예측 결과 데이터

7. 프로젝트 데이터

데이터 이름 (파일명)	데이터 설명 (어떠한 데이터 인지 최대한 설명해 주시기 바랍니다)
데이터 1	원본 데이터를 데이터 내 game_id 칼럼의 10000까지 자른 데이터

train10000.csv	
데이터 2 train20000.csv	원본 데이터를 데이터 내 game_id 칼럼의 10001부터 20000까지 자른 데이터
데이터 3 train30000.csv	원본 데이터를 데이터 내 game_id 칼럼의 20001부터 30000까지 자른 데이터
데이터 4 train38871.csv	원본 데이터를 데이터 내 game_id 칼럼의 30001부터 마지막까지 자른 데이터
데이터 5 train_preprocessed.csv	위 원본 데이터를 분석하고자 하는 방식으로 전처리된 데이터
데이터 6 test.csv	원본 test 데이터
데이터 7 test_preprocessed.csv	원본 test 데이터를 분석하고자 하는 방식으로 전처리된 데이터
데이터 8 result_submission.csv	test 데이터 예측결과 데이터