3차 빅데이터 프로젝트 - 개발 계획서

김준회(조장), 문성만, 정해준, 정일호, 임도균

1. 프로젝트 주제명 : PUBG Finish Placement Prediction

2. 프로젝트 개요

Kaggle에서 진행 중인 Competition 문제 해결을 기반으로 한 프로젝트이다. 위의 Competition에서는 배틀그라운드 플레이어들의 게임 플레잉 데이터를 제공한다. 참가자들은 주어진 테스트 데이터셋의 각 플레이어 데이터에 대하여 1등할 확률(이하 ‘등수’)을 예측하는 모델 생성한다. 생성한 모델을 통해 실제 등수와 예측 등수의 오차를 최소화하는 것을 목적으로 한다.

본 프로젝트는 위의 Competition 참여를 통해 생성한 등수 예측 모델을 기반으로 한 “배틀그라운드 플레이어 등수 예측” 서비스를 개발한다.

3. 프로젝트 주요 기능

- 플레이어 등수 예측 서비스

- 플레이어 게임 결과 데이터 시각화 서비스

- 실 유저 데이터 입력 시 결과 예측 서비스

4. 주요 데이터셋

- 플레이어 비식별화 유저 플레이 데이터(학습 데이터) - 약 440만건

- Kaggle Competition 제출을 위한 submission data - 약 100만건

- 서비스 제공을 위한 실유저 게임 데이터

5. 주요 활용 기술(인프라 구성) - \*도식화

- Python 기반 Tensorflow : 모델 생성 및 플레이어 등수 예측

- Python 기반 numpy, pandas, matplotlib, seaborn 등 EDA

- R 기반 EDA

- Elastick-search or Hadoop : 실 유저 데이터 수집 및 학습 데이터 관리

- 플라스크 or 장고 : 웹 서비스 벡엔드 구현

- D3.js 및 시각화 언어 툴 : 실 유저 데이터를 활용한 시각화

6. 계획서 종합 및 목표

- Kaggle 대회 참여를 목표로 하는 만큼 우수한 성적을 내는 것이 목표

- Kaggle 대회는 Scoring을 MAE(Mean Absolute Error)로 하는 만큼 머신러닝 혹은 딥러닝 기반 모델을 활용할 때 MAE를 최소화 하는 방향으로 진행

- MAE를 최소화 하기 위해 예측이 가능한 모델(다중선형회귀분석, 의사결정나무, 랜덤포레스트, RNN, CNN 등)을 모두 적용해보면서 공부하는 것이 목표

- 차순위 목표로, 단순히 분석 뿐 아니라 실 유저 데이터를 활용한 웹 기반의 서비스를 통해 구현하는 것이 목표

참고 : 배틀그라운드란 블루홀 회사에서 개발한 생존게임으로 steam, kakao등의 배포회사를 통해 전 세계의 유저를 보유한 big scale의 게임. 솔로 게임의 경우 100명의 플레이어가 참가하여 1등부터 100등까지 순위를 가지게 되는데, 생존게임인 만큼 순위를 예측할 수 있는 변수가 다양하고(28개 변수), 변수간의 관계도 복잡하여 다양한 머신러닝 기법을 활용하는데 효과적일 것이라고 생각합니다.