빅데이터 분석 기획서

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 분석 과제명 | Blizzard 스타크래프트2 경기 행동 데이터를 활용한 승패 예측 | | | | |
| 분석 기간 | 2020.10.19 ~ 2020.11.13 | | | | |
| 입주자(팀) | 소속기관 | KIC캠퍼스 | | 소재지 | 서울특별시 강남구 역삼동 819-10 세경빌딩3층 |
| 구분 | 성명 | 이메일 | 연락처 | 역할 |
| 팀장 | 곽영상 | ysk1188@naver.com | 01072306006 | 데이터 수집,  데이터 전처리,  데이터 분석,  발표자료 작성,  발표 |
| 팀원1 | 안소연 | ggj03250@naver.com | 01029536491 |
| 팀원2 | 이상연 | juliano\_yeon@naver.com | 01020321565 |
| 팀원3 | 이종희 | registation@naver.com | 01064844761 |
| 팀원4 | 장민정 | sparklestory@naver.com | 01064702218 |
| 연구 배경 및 목적 | 한국콘텐츠진흥원이 발간한 ‘2019 대한민국 게임백서’에 따르면 e스포츠의 위상이 증가함에 따라 2018년 8월 18일부터 9월 2일까지 진행된 ‘2018 자카르타 팔렘방 아시안 게임(2018 Jakarta-Palembang Asian Games)’에서는 아시안 게임 역사상 최초로 e스포츠를 시범 종목으로 채택하였다. 포함된 종목은 리그 오브 레전드, 아레나 오브 발러, 프로 에볼루션 사커, 스타크래프트2, 클래스 로얄, 하스 스톤으로 총 6개였다. 한국은 스타크래프트에서 금메달을, 리그 오브 레전드에서 은메달을 각각 차지했다. 스타크래프트2는 3개의 종족으로 플레이하는 실시간 전략 게임으로 초반 빌드, 전략으로 인해 승패가 정해진다.  본 연구에서는 스타크래프트2 경기의 행동 데이터를 토대로 승패를 예측하고 이를 통해 게임 승리를 위한 요소들을 확인하여 전략을 발전시키고자 한다. | | | | |
| 연구 및 분석 계획 | 1. 자료 조사 및 데이터 수집을 통한 변수 의미 파악  * Youtube, Blizzard 사이트 활용  1. 데이터 분석을 위한 데이터 Remodeling  * 새로운 DataFrame 생성  1. 추가적인 데이터 분석을 통한 추가 변수 생성  * Train\_worker, 정찰 column 추가  1. 데이터 분석 기법(Random Forest)을 위한 데이터 정형화  * 종족 = Terran -> 1, Protoss -> 2, Zerg -> 3  1. Random Forest 적용 및 결과 확인 2. 결론 도출 | | | | |
| 결론 및 한계점 | · 결론   * Timegap(시간차) 변수에서 예측력이 가장 많이 향상 * 승패 예측의 핵심 요인은 초반 유닛의 시간 효율성과 유의미한 행동의 수로 추정   · 한계점   * 프로 대회 경기 내용이 아닌 일반인들의 게임 내용이기 때문에 이상치 관측 * 게임 시간이 일정하지 않아 시간 개념에서 공정한 기준을 못 만듬 * Attack 횟수는 알 수 있지만 거기서 누가 승리했는지 알 수 없음 * Recon 컬럼 세분화 부족 * 늘어난 변수들 사이에 존재하는 다중공선성을 해결 못 함 * 다중 중첩 for문 사용으로 데이터 전처리 속도가 낮음 | | | | |

데이터

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **번호** | **데이터셋명** | **주요필드설명** | **데이터의 종류** | **원천** |
| 1 | Blizzard 스타 2 행동 데이터 셋 | Species, Event, Event\_contents | CSV | DACON |
| 2 | Blizzard 스타 2 행동 세분화 데이터 셋 | Species, Train\_worker, Train\_unit, Build, Upgrade, Selection, Camera, ControlGroup, Right Click, Result | CSV | 자체 제작 |

일정관리

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **10월 4주** | **10월 5주** | **11월 1주** | **11월 2주** |
| **프로젝트 기획** | 데이터 탐색 | 데이콘에서 프로젝트 선정 |  |  |
| **자료 조사 및 데이터 수집** |  | 유튜브와 블리자드 사이트 경로 통해 수집 |  |  |
| **데이터 Remodeling** |  |  | 유닛의 효율성 분류 |  |
| **추가 변수 생성** |  | counter 변수와 recon 변수 생성 | Time weight, Time gap 변수 생성 | Attack 변수 생성 |
| **데이터 정형화** |  |  |  |  |
| **Random Forest 기법 적용** |  |  |  | 53.9에서 56.1로 상승 |
| **결론 도출** |  |  |  | 초반에 유닛의 시간 효율성과 행동의 빈도 수가 승패에 영향을 끼침 |