**[데이터 시각화&분석 취업캠프]**

**최종 프로젝트 기획안**

기획안 작성일자 : 2023-05-04

|  |  |
| --- | --- |
| 조 명 | 2조 : 축구 쉽조 |
| 조 원 | 조장 : 정재승  조원 : 박용운, 김동욱 |
| 프로젝트 주제 및 개요 | **■ 주제**  축구 역할 분류 시스템  **■ 개요**      첫 번째 기사는 TV 프로그램 ’뭉쳐야 찬다‘에서 국가대표 출신인 안정환이 직접 방송인들의 포지션을 지정해 주는 기사입니다. 하지만, 조기축구 팀에서는 TV 프로그램과 같이 축구를 전문으로 배운 사람이 포함되어 있지 않은 경우가 많습니다. 이러한 경우 각 회원이 자신이 원하는 포지션의 역할인지, 경기를 소화할 때 자신이 어떤 역할을 수행해야 하는지 알기 어렵습니다. 두 번째 글을 보면, 포지션 설명을 해주는 블로그도 볼 수 있습니다. 그렇지만, 백문이 불여일견이라는 말이 있듯이 아마추어 축구 선수들이 직접 경기를 뛰면서 자신의 역할을 정확하게 이해하고, 최적의 역할을 수행할 수 있도록 도와주는 축구 역할 분류 시스템이 필요합니다.  이러한 시스템을 구현하기 위해 20년, 21년, 22년 K리그 선수들의 데이터를 수집하고, 이를 분석해 축구 선수의 능력치와 역할을 분류하는 모델을 개발했습니다. 이를 위해 다양한 머신 러닝 분류기와 데이터 분석 기술을 활용할 수 있으며, 이를 통해 정확하고 신뢰성 높은 시스템을 구축할 수 있습니다.  이러한 축구 역할 분류 모델은 아마추어 축구 선수들뿐만 아니라, 프로 축구 선수들에게도 유용하게 활용될 수 있습니다. 선수들은 역할 분류 모델을 통해 자신의 역할을 더욱 명확하게 이해하고, 경기에서 더욱 효과적인 역할을 수행할 수 있게 됩니다.  **■ 목적 및 기대효과**   1. 아마추어 축구 선수들이 최적의 역할을 수행할 수 있도록 도와준다. 2. 선수의 기술과 능력, 경기 상황 등을 고려해 해당 선수가 최적의 역할을 수행할 수 있는 포지션을 분류한다. 3. 이 시스템을 통해 아마추어 축구 선수들은 자신의 역할을 더욱 명확하게 이해하고, 경기에서 더욱 효과적인 역할을 수행할 수 있다.   **■ 대상**  아마추어 축구 선수  축구선수를 준비하는 미래 유망주  축구를 좋아하는 모든 사람들 |
| 프로젝트 수행 방향 | 0. 데이터 수집 : 2022년, 2021년, 2020년 K-리그 1 / K-리그 2 선수의 개인 스탯, 2022년, 2021년, 2020년 K-리그 1 / K-리그 2 선수의 히트맵, 2022년 K-리그 1 / K-리그 2 선수 프로필 이미지  1. 데이터 전처리 : 2020년 K-리그 1 / K-리그 2 선수 중 이적 선수 병합, 동명이인 분리, 출전시간이 180분 미만인 선수 제거, 2021년 K-리그 1 / K-리그 2 선수 중 이적 선수 병합, 동명이인 분리, 출전시간이 180분 미만인 선수 제거, 2022년 K-리그 1 / K-리그 2 선수 중 이적 선수 병합, 동명이인 분리, 출전시간이 180분 미만인 선수 제거, 2022년 기준 선수 추천을 위한 데이터 프레임(607개)과 군집화를 위한 20년, 21년, 22년 선수 기록 데이터(1701개) 분리.  2. 포지션 분류 : GK, DF, MF, FW 총 4가지로 나누기 위해 히트맵으로 분류 학습 진행  3. 군집화 모델링 - KMeans  3.1 DF는 측면 수비수, 중앙 수비수로 군집하기 위해 알맞은 속성값으로 군집화  3.2 측면 수비수는 Overlapper, Progressor, Safety로 군집화  3.3 중앙 수비수는 Anchor, Spreader, Aggressor로 군집화  3.4 MF는 수비형 미드필더, 공격형 미드필더로 군집하기 위해 알맞은 속성값으로 군집화  3.5 수비형 미드필더는 Builder, Distributor, Box to box로 군집화  3.6 공격형 미드필더는 Creator, Box crasher로 군집화  3.7 FW는 중앙 공격수, 측면 공격수로 군집화기 위해 알맞은 속성값으로 군집화  3.8 중앙 공격수는 Finisher, Target, Roamer로 군집화  3.9 측면 공격수는 Wide threat, Unlocker, Outlet으로 군집화  4. 결과 분석 및 라벨링  5. 라벨링 된 데이터 프레임과 분류 학습을 위한 속성값들로 데이터 프레임 재구성.  6. 사용자가 한 경기 스탯을 입력해 어떤 역할을 수행하는 결과 출력 및 그 역할을 수행하는 2022년 기준 K리그 선수 추천 시스템 모델링  7. 축구의 경험이 없는 사용자가 선호하는 역할을 추천하고, 그 역할을 수행하는 2022년 기준 K리그 선수를 추천해 주는 추천 시스템 모델링  8. 추천 선수 스탯 상세 페이지 구현(선수 이미지, 선수 역할, 스탯 시각화)  ● 수행도구  Python, Pandas, Numpy, Matplotlib, Seaborn, Streamlit, Selenium, Open CV, Sklearn, Tensorflow, Keras  ● 데이터  K-리그 데이터 센터(선수 20년, 21년, 22년 기록 데이터) - [www.data.kleague.com](http://www.data.kleague.com)  K-리그 데이터 센터(선수 이미지 데이터) - www.data.kleague.com  K-리그 선수 히트맵 데이터 – [www.sofacore.com](http://www.sofacore.com) |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  **조장 : 정재승** - 2022년 K-리그1 / K-리그2의 선수 명단, 선수 프로필 사진, 2022년 2021년 2020년 K-리그1 / K-리그2의 선수 개인 스탯 정보 수집, 군집화 모델링, 시각화 페이지 구현  **조원 : 박용운** - 2022년 K-리그1 / K-리그2의 선수 명단, 선수 프로필 사진, 2022년 2021년 2020년 K-리그1 / K-리그2의 선수 개인 스탯 정보 수집, 히트맵 이미지 학습  **조원 : 김동욱** - 2022년 K-리그1 / K-리그2의 선수 명단, 선수 프로필 사진, 2022년 2021년 2020년 K-리그1 / K-리그2의 선수 개인 스탯 정보 수집, 역할 분류 학습, 챗봇 |
| 프로젝트 추진 일정 | ● 일정  4/4 ~ 4/11 : 프로젝트 주제 선정 및 기획안 작성  4/12 ~ 4/19 : 프로젝트 데이터 수집 및 데이터 전처리 작업  4/20 ~ 4/28 : 프로젝트 파트별 작업 진행(각 페이지에 들어갈 모델 또는 결과물 구축)  5/1 ~ 5/9 : 스트림릿 구축 및 포트폴리오 작성  5/10 : 프로젝트 발표 및 평가 |

프로젝트 다이어그램

