|  |  |  |  |
| --- | --- | --- | --- |
| **2021 혁신성장 빅데이터 분석 프로젝트 요약서** | | | |
| **팀 명** | ADAM | **팀 장** | 김기영 |
| **팀 원** | 김민송, 박소연, 이동렬 | | |
| **주 제 명** | NBA 경기데이터 분석 | | |

|  |  |  |
| --- | --- | --- |
| 구분 | 프로젝트 상세내용 | |
| 프로젝트명 | NBA 경기데이터 분석 | |
| 구현 목표 | 1. NBA 경기데이터를 통해 각 포지션별 선수들의 스탯 분석 후 전성기 나이대를 알아보기 2. 역대MVP 스탯을 통해 2020~21년도의 MVP 예측해보기 | |
| 구현 내용 | 1. NBA 경기 실적 분석해 각 포지션별 선수들의 스탯 분석 후 이를 바탕으로 전성기 나이 뽑아 보기 | 2. MVP 예측모델  역대MVP스탯을 학습시켜 2020~2021 MVP 예측모델 만들기 |
| -NBA 경기 데이터 수집 및 정제  -정제한 데이터를 이용하여 각 포지션별 스탯을 분석하고 ggplot2를 통한 시각화하기  -시각화한 결과를 해석하여 전성기 나이 구하기 | -모델링 과정  (Decision Tree, Random Forest, SVM)  -모델링 과정을 거친 후 모델 검증 (교차 검증)  -모델 검증을 한 후 최적 모델 선택 및 알고리즘 시각화 |
| 상세 설명 | **1. 주제 선정/개요**  1-1) 선정 배경  우선 데이터를 살펴보니 데이터가 다루기 좋아 보였고 우리가 수업시간에 배운 내용을 적용해보기에 좋을 것이라고 생각해서 선정하게 되었습니다.  1-2) 솔루션  데이터를 정제하고 시각화하는 과정을 거쳐서 결과를 해석함  모델링 작업을 통해 모델 검증과정을 거치고 앞으로 예상되는 결과에 대한 예측모델을 만듬  **2. 분석 및 기능 설계/개발**  2-1) 데이터 수집 및 전처리  -주사용 데이터: NBA Player stats since 1950  -데이터수집사이트:  -Kaggle  (<https://www.kaggle.com/drgilermo/nba-players-stats>)  -추가 수집 데이터: 2020-2021 시즌 NBA 경기 데이터 수집 (https://www.basketball-reference.com)  2-2) 데이터 전처리 및 시각화  2-3) 데이터 분류 및 모델링  2-4) MVP예측 모델을 만들고 예측하기  **3. 서비스 소개**  -NBA 경기 실적데이터를 기반으로 각 포지션 선수들의 스탯을 분석하여 전성기 나이를 알 수 있다.  -MVP스탯을 통하여 앞으로 누가 MVP로 선정될지를 예측할 수 있다.  **4. 활용방안 및 기대효과**  -각 포지션별 선수들의 전성기 나이가 언제가 될지를 알아볼 수 있다.  -선수들의 경기 실적을 통해 어떤 선수가MVP선수가 될지를 예측해 볼 수 있다  **5. 개선점**  **-**연구 결과에서 결과값에 너무 큰 영향을 주는 데이터 값이 예측되어 결과의 정확도가 떨어지는 것 같다.  **6. 분석도구(R)**  -분석에 사용되는 모델: Decision Tree, Random Forest, SVM | **1.분석 개요(설계)**  1) 현 상황과 문제점  정확도가 너무 높다,  학습 데이터와 테스트 데이터로의 구분이 어렵다  2) 분석설계: 데이터 키워드  채널  - Kaggle(NBA player)  - 2020-2021 시즌 NBA 경기 데이터  3) 분석내용  -각 포지션별 스탯 분석  -각 포지션별 전성기 나이 대 파악 및 시각화하기  -역대MVP선수들과 MVP가 아닌 선수들의 기량 차이  -MVP선수 예측하기  4) 분석도구:  -R  -분석 모델의 종류: Decision Tree, Random Forest, SVM,  3. 프로젝트 결과 요약  연구1. 각 포지션별 선수들의 스탯 분석을 통한  전성기 나이대는   1. PG: 다른 포지션에 비해 전성기가 늦으며 전성기 기간이 길다 2. SG: 20대 중반 득점력이 만개하나 슈팅력은 점차 하락한다. 3. SF: 20대 운동능력 중요, 30대 중반 부상 관리 4. PF: 25~30, 상대적으로 전성기가 2년 길다 5. C: 리바운드와 블락이 두드러져야 한다  * 결과를 보면 기량이 뛰어난 한두 선수가 전체 통계에 영향을 다소 끼치는 경우가 있었다   연구2. 2020~21년도의 MVP 예측 모델은 랜덤 포레스트의 결과로는 Nikola Jokić이 니왔고 서포트 벡터 머신의 결과로는 Giannis Antetokounmpo, Nikola Jokić 2명이 나왔다. 두 모델이 결과값에 공통적으로 2021 MVP로 예측한다  3.프로젝트 보완점  데이터의 단위 통일이 이루어지지 않아서 데이터 전처리작업이 복잡했으며, 데이터 처리방법이 미흡한 부분이 있었다 |

|  |  |  |
| --- | --- | --- |
| 각 팀원 역할 |  |  |
|  |  |
| 이름-담당역할-업무내용-분석프로그램 등등  김기영(조장) 발표  김민송(팀원)  박소연(팀원)  이동렬(팀원)  각자 역할을 구분지어 하기 보다는 서로서로 도와가며 데이터 추출에서부터 데이터 처리과정을 함꼐 수행했습니다. | |

양식은 팀 프로젝트에 따라 세부항목 및 내용 변경가능하오니, 수정하여 작성해 주세요.

PPT 사진 첨부 및 상세 작성