**데이터 수집 및 전처리 정의서**

1. 데이터 수집 및 설명
   1. Football Manager 2023 dataset
      1. <https://www.kaggle.com/datasets/platinum22/foot-ball-manager-2023-dataset>
   2. 축구 시뮬레이션 게임인 Football Manger에서는 실제 축구 선수들의 데이터를 수치화 하여 저장 중
   3. 수치화된 데이터를 이용하여 선수 검색
2. 데이터 전처리 – 번수 명 한글 변환
   1. 영어로 되어있는 변수들의 이름을 한글로 변환하는 작업
   2. 예시

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 영어 | 한글 | 영어 | 한글 | 영어 | 한글 |
| Name | 이름 | Position | 포지션 | Age | 나이 |
| ca | 현재 능력 | pa | 잠재력 | Passing | 패스 |
| Composure | 침착성 | Tackling | 태클 | Values | 시장 가치 |
| Pace | 스피드 | Dribbling | 드리블 | Finishing | 슛 |

1. 데이터 전처리 – 몸무게 결측치 보완
   1. 몸무게가 0으로 입력된 선수 일부 존재
   2. KNN ( K-최근접이웃 ) 알고리즘을 이용하여 선수들의 키로 몸무게 예측
   3. 단순 평균 보간법보다 신뢰성 확보 가능
2. 데이터 전처리 – 잠재 능력 변환
   1. 나이가 어린 선수들 중 잠재능력이 음수를 가지는 선수 일부 존재
   2. 해당 게임에서 제공하는 기준을 이용하여 0~200 사이의 수치로 변환

|  |  |
| --- | --- |
| 원본 잠재력 | 변환 잠재력 |
| -10 | 185 |
| -95 | 175 |
| -9 | 165 |
| -85 | 155 |

1. 데이터 전처리 – 대표 포지션 설정
   1. 선수들을 포지션 별로 구분하여 관리하기 위해 대표 포지션 설정
   2. 원본 데이터에서 제공하는 각 포지션 점수를 이용하여
   3. GK(골키퍼) / D(수비수) / M(미드필더) / AM(공격형 미드필더) / ST(공격수) 로 구분
2. 데이터 전처리 – 데이터베이스 구축
   1. 선수의 기본 정보와 포지션 별 주요 스탯을 기반으로 테이블 생성
   2. 포지션 별 총 5개의 테이블을 하나의 데이터베이스로 관리
   3. 예시) ST(공격수) 테이블 스키마

|  |  |
| --- | --- |
| Field | Type |
| 이름 | text |
| 나이 | int |
| 포지션 | text |
| 현재\_능력 | int |
| 잠재력 | int |
| 국적 | text |
| 소속팀 | text |
| 키 | int |
| 몸무게 | int |
| 왼발\_능력 | int |
| 오른발\_능력 | int |
| 시장\_가치 | int |
| 슛 | int |
| 드리블 | int |
| 퍼스트\_터치 | int |
| 헤딩 | int |
| 예측력 | int |
| 오프\_더\_볼 | int |
| 침착성 | int |
| 가속력 | int |
| 체력 | int |
| 민첩성 | int |