

# 국민체력 증진을 위한 인공지능 운동 추천서비스

[국민체력100] 지구력을 선택하세요.

보통

[국민체력100] 근력을 선택하세요.

보통

[국민체력100] 운동 강도를 선택하세요.

중급자

[국민체력100] 선호하는 운동 환경을 선택하세요.

실내 운동

[국민체력100] 주로 운동하는 시간을 선택하세요.

저녁

[국민체력100] 운동 목적을 선택하세요.

근력 증진

제출

당신의 성별은 남성이며, 연령대는 10대입니다.

운동 추천 챗봇 버튼을 클릭하세요. 챗봇 페이지로 이동합니다.

운동 추천 챗봇

주변 지도

국민체력100\_Youtube

# 01 분석 배경

## 제안 배경

### 생활체육 관련 정보 수집 및 비참여 이유

그림 2-2-6 생활체육 관련 정보 수집 경로

(전체, 단위: %)



그림 3-5-4 체육활동 비참여 이유

(비규칙적 체육활동 참여자, 단위: %)



### OECD '방과 후 운동 학생' 비율

순위 국가 비율(%)

1	아일랜드	77.6	체육활동을 안하는 이유 단위: %, 자료: 한국청소년정책연구원
2	네덜란드	76.1	문화
3	슬로바키아	75.1	공부하느라 바빠서, 시간이 없어서
8	미국	71.0	계물러서
18	독일	68.6	몸이 약해서
34	일본	49.8	같이 할 사람이 없어서
35(골짜)	한국	42.9	주위에 운동할 장소, 도구, 시설이 없어서
	OECD 평균	66.0	



경제적 여유(돈)이 없어서  
가르쳐 주는 사람이 없어서  
관심이 없어서  
부모님이 반대해서  
운동을 싫어해서

그렇다	아니다
52.4	43.3
81.3	15.3
9.3	86.1
37.4	58.4
17.3	78.8
8.8	87.3
20.4	75.4
54.4	41.6
3.4	92.4
44.5	51.6

- WHO 전세계 146개국 11~17세 청소년 160만명 대상으로 운동 상태 조사
- 한국 남자 청소년의 운동부족 비율도 91.4%로 필리핀 (약 92.8%) 다음으로 심각
- 여자 청소년은 한국(97.2%)의 운동량이 가장 부족

WHO 연구팀이 꼽은 세 가지 요인

- 1) 지나친 학업에 따른 시간 부족
- 2) 스마트폰 과용
- 3) 청소년을 위한 스포츠레저시설이 부족한 현실

# 01 분석 배경

## 제안 배경

### 공공기관의 체력시험 평가제



- 국민의 체력 증진을 위한 복지서비스가 최초 목적과는 달리 활용되면서 국민이 이용하기 어려워짐
- 국민체력 100 인증 민간항공 채용 평가 둔갑
- 공공기관 등에서 '국민체력 100'으로 체력시험 평가를 대체
- 일부 민간항공사들마저 '국민체력 100'을 객실승무원 채용에 이용하기 시작하면서 예약은 더 어려워짐

### 국민체력100 인증센터의 저조한 참여율

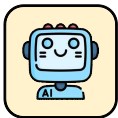


총인원	30만3772명	비율(%)
10대	15만1475명	49.9
20대	4만3517명	14.3
30대	1만4811명	4.9
40대	1만4512명	4.8
50대	1만9360명	6.4
60대	2만7599명	9.1
70대	2만6115명	8.6
80대 이상	6383명	0.8

- '국민체력 100' 인증센터 적고, 참가자는 전체 국민 대비 1% 미만(50%가 10대)
- 생애주기별 '평생 체육' 효과적 연결이 주요 요소
- "측정부터 상담까지 1인당 최소 30분이 소요"
- "연간 측정인원 35만명이 사실상 최대치"

# 01 분석 배경

분석 목표



인공지능을 활용한  
맞춤형 운동 추천



챗봇을 활용한  
맞춤형 운동 추천



온라인으로 누구나  
쉽고 간편한 이용



국민체력 증진



국민체력100 활성화

“

국민체력 증진을 위해 **인공지능으로 맞춤형**  
**운동 추천과 챗봇**을 온라인 웹앱으로 제공

”



국민체력 증진  
운동 추천 챗봇



국민체력 증진



국민체력 100 향상

# 02 데이터 전처리 분석 프로세스

## 데이터 수집

 문화 빅데이터 플랫폼

 국민체육진흥공단  
Korea Sports Promotion Foundation

체력측정 및 운동처방 종합데이터  
('19년 4월 ~ '23년 4월 까지)

서울시 체육시설인접 대중교통 정보

 공공데이터포털  
DATA .GO .KR

 국민체력100  
NATIONAL FITNESS INDEX

국민체력100 운동처방 동영상주소 정보

 서울  
서울특별시 도시

서울시 공공체육시설정보

서울시 생활체육포털 우리 동네 프로그램

서울시 체육시설 공공서비스예약 정보

서울시 주요 공원현황

## 데이터 전처리&EDA

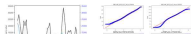
< 추천 데이터 >

결측치가 90% 이상인 컬럼 삭제

도메인 지식을 활용하여 이상치 삭제

EDA를 통한 데이터 편향 분석

< EDA >



시간 유사도 분석  
(DTW)

정규성 검정  
(QQ Plot)

< 군집분석 >



K-평균

실루엣 스코어

Elbow method

< 공간 데이터 >

결측치 삭제 및 전처리



주소



위도, 경도 생성

## 모델링

< 추천 시스템 모델 >



교사인 유사도

라벨 인코딩



< 운동 처방 모델 >

체력 측정 값을 넣으면 운동 추천

< 챗봇 모델 >

사용자의 입력 값

Persona

Few shot learning

(의사, 헬스 트레이너)

(본운동 목록)

↓

 Bard

↓

생성형 LLM 모델 기반  
맞춤형 운동 추천

## 웹앱(Streamlit)

 Streamlit

< 맞춤형 운동 추천 >

사용자에게 맞춤형 운동 추천

(추천 시스템 모델, 챗봇 모델)

< 운동 추천 챗봇 >

바드 API를 활용한 서비스 제공

< 주변 지도 >

서울시 체육시설 위치 및 정보 제공

< 국민체력100 유튜브 >

 국민체력100

국민체력100 유튜브 연동

< XAI&Profiling >

추천 데이터를 사용한 XAI 제공  
(운동 처방 모델)

## 02 데이터 전처리

### 데이터 전처리(결측치, 이상치)

### 데이터 범위

#### 고려사항

##### 장기간 체력 검증

- 시간이 갈수록 체력 측정 결과가 좋아진 사람들의 데이터가 필요

##### 코로나 기간

- 코로나 기간인 20~21년 동안 거리두기로 인해 다른 연도에 비해 데이터가 적음



19년도 데이터 사용



4년치 데이터 사용  
(19~23.04 활용)

### 예시 데이터

	MBER_SEQ_NO_VALUE	MESURE_SEQ_NO	CNTER_NM	AGRDE_FLAG_NM	MESURE_PLACE_FLAG_NM	MESURE_AGE_CO	INPT_FLAG_NM	CRTFC_FLAG_NM
0	AAGcV7mrSzMrZDRgRnl4dsr	1	나주	청소년	물장	15.0	관리자	3등급
1	AAEx9WqQ1IThUlypWchgyYTL	4	종구(광주)	노인	일반	69.0	관리자	참가중
2	AAHIGFYeEJpaL2duLdfWR7m	1	남구(울산)	청소년	물장	15.0	관리자	참가중
3	AAHyVsNhXusNBqK-YQb398c8	1	남원	성인	물장	30.0	관리자	2등급

데이터 : 약 83만



#### 범주형 데이터 카테고리 타입으로 변환

영문명	한글명	Before	After
AGRDE_FLAG_NM	연령대구분명	VARCHAR	category
CNTER_NM	센터명	VARCHAR	category

#### 측정 등급을 정수로 매핑

- 데이터 분석과 등급이 높을 수록 더 높은 값을 부여  
(참가중 : 0, 3등급 : 1, 2등급 : 2, 1등급 : 3)

#### 날짜 데이터를 Datetime 형식으로 변환

```
df_cleaned['MESURE_DE'] = pd.to_datetime(df_cleaned['MESURE_DE'], format='%Y%m%d')
df_cleaned.info()
```

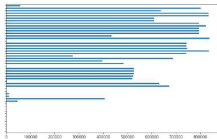
## 데이터 결측치

절반 이상의 컬럼에서  
결측치 발견

컬럼	결측치 수
MESURE_IEM_04_VALUE	48.512331
MESURE_IEM_09_VALUE	80.271101

...

MESURE_IEM_041_VALUE	95.6
----------------------	------



## 데이터 컬럼 제거

결측치가 90%가 넘는 데이터

영문명	한글명	결측치 수
MESURE_IEM_024_VALUE	6분걸기(m)	99.7
MESURE_IEM_029_VALUE	피부두겹합	100
MESURE_IEM_031_VALUE	트레드밀_안정시(bpm)	94.7
MESURE_IEM_032_VALUE	트레드밀_3분(bpm)	94.8

...

MESURE_IEM_041_VALUE	성인체공시간(초)	95.6
----------------------	-----------	------

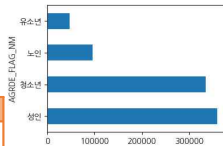
국민체력100 연령대별 측정기준과 맞지 않는 데이터

실제 데이터 연령대 비율

성인 : 42.9%, 청소년 : 40.1%  
노인 : 11.5%, 유소년 : 6%

유아기 인증기준 데이터

컬럼	한글명
MESURE_IEM_021_VALUE	10M 4회 왕복달리기(초)





## 02 데이터 전처리

데이터 전처리(결측치, 이상치)

### 데이터 이상치

#### 왕복오래달리기 속성

- 왕복 거리에 따라 구별됨
  - ▶ 왕복오래달리기 20(m)
    - 성인과 청소년만 포함되며, 유소년과 노인은 포함되지 않음
  - ▶ 왕복오래달리기 15(m)
    - 유소년만 포함됨

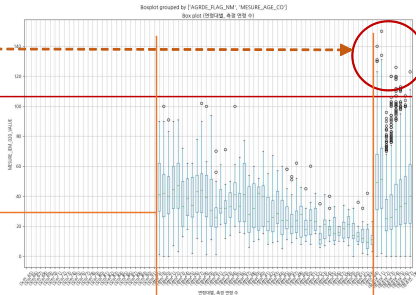
↓  
데이터 확인 결과  
성인과 청소년, 유소년  
모두 포함되어 있음

✓ 데이터 분석을 통해  
이상치 여부 판단

유소년(11,12)세  
데이터 이상치 발견

성인들의 최대  
오래달리기 횟수

노년층 데이터  
(결측값)



#### MESURE\_IEM\_020\_VALUE : 왕복오래달리기 (회)

- 유소년 데이터
  - ▶ 유소년 데이터는 국민체력100에서 성인과 동일한 왕복거리(20m)를 달리지 않고 15(m)를 달리므로 국민체력100 기준과 맞지 않음
  - ▶ 성인들의 최대 오래달리기 횟수보다 많이 달렸을 뿐만 아니라, 장거리 달리기 선수(마라토너 제외) 횟수인 141.65 보다 더 많이 달린 데이터가 존재

✓ 데이터 전산 및 통합 과정에서 생긴 오류로 판단  
왕복오래달리기에 포함된 모든 유소년 데이터를 제거

## 데이터 고유 특징 분석

## 성인, 청소년 데이터

- 청년, 청소년 데이터
  - 타성인 데이터에서 추출된 청년 데이터와 청소년 데이터의 빈도가 가장 큼
- 측정일자 데이터
  - ▶ 월별로 분리
    - 월 정보만 따로 추출하여 추세를 가지고 있는지 분석
  - ▶ '연도-월' 데이터 활용 (2023-04)
    - 연도와 월을 활용해서 시각화

✓ 학교 일정과 연관성 확인

✓ 추운 겨울보다 따뜻한 봄에 주로 측정

## 학교 일정 정보

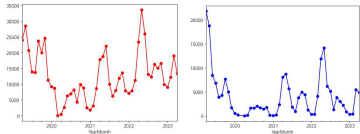
- 7~8월 : 여름방학
- 1~2월 : 겨울방학
- 7월 : 1학기 기말고사
- 12월 : 2학기 기말고사
- 3월 : 입학식 (학교 적응 기간)
- 12월 : 한 해 마무리

## 측정일자 (월) 데이터

	측정 월	데이터 수
봄	4	129191
	5	128927
	6	99130
가을	11	78475
	10	75275
여름	9	67355
	7	60607
겨울	3	51967
	8	49172
	12	36063
	2	33236
	1	28585
	합계	837983

## 추세 확인

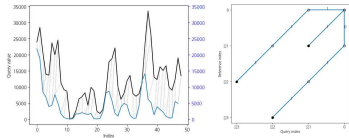
## ■ 청소년 '반복점프' line 그래프



▲ 결측값

▲ 결측치가 아닌 값

청소년만 측정하는 데이터를 기준으로 데이터 분석  
두 데이터의 추세가 비슷해 보이므로 DTW를  
사용해서 유사도 분석



스피어만 상관계수 결과 : 0.86  
강한 양의 상관관계가 있음을 확인

## 02

### 데이터 전처리

탐색적 데이터 분석(EDA)

#### 데이터 정규성 검정

##### 노인 데이터

##### ■ 노인 속성의 고유 데이터

- 노인 데이터의 다른 연령대는 측정을 하지 못하는 데이터로 이루어져 있음

##### ■ EDA를 통한 정규성 검정

- ▶ 데이터 시각화를 통해 정규성 총족
- ▶ Q-Q Plot와 히스토그램 활용
- 데이터 시각화로 정규성 판단에 도움됨

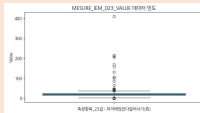


노인 데이터 추가 분석



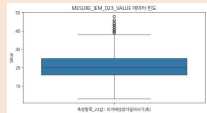
Q-Q Plot, 히스토그램 활용

#### IQR 이상치 제거



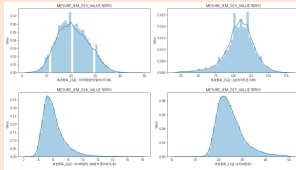
##### 이상치 제거 전

- 각 데이터의 사분위 수를 구한 후 IQR을 계산
- 상단 경계, 하단 경계를 기준으로 해당 값을 벗어난 데이터 이상치로 판단



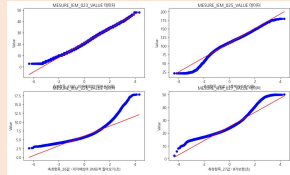
##### 이상치 제거 후

#### 정규성 검정



##### 히스토그램

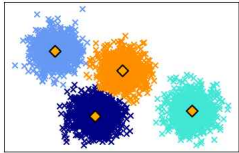
- 시각화를 통해 데이터 분포가 정규성을 보이는지 분석
- 왜도를 확인할 수 있으며, log 변환 혹은 Box-cox 변환을 통해 데이터를 정규화 한다.



##### Q-Q Plot

# 03 데이터 분석 모델링 - 군집분석

## K-means

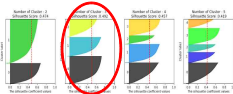


### K-means란?

? 데이터 간의 거리의 차이가 최소가 되도록 하는 중심점을 생성하여 데이터들을 K개의 클러스터에 할당하는 군집화 알고리즘

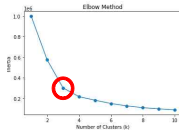
성별에 따른 1차 분류 진행 후 나이, BMI 수치를 통한 2차 군집화 진행

## Silhouette Method



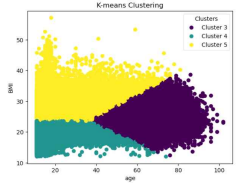
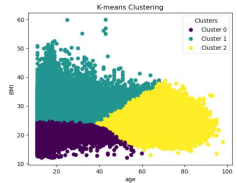
Silhouette score가 가장 높고  
데이터의 분포도 가장 고른  
군집의 개수(K)를 3으로 결정

## Elbow Method



군집의 개수(k)가 2->3, 3->4에서의  
변동성이 크게 감소한다고 판단하여  
군집의 개수(k)는 3으로 결정

## K-means 결과



cluster	0	1	2	3	4	5
성별	M	M	M	F	F	F
나이	높음	낮음	낮음	높음	낮음	낮음
BMI	중간	낮음	높음	중간	낮음	높음

## 03

## 데이터 분석

## ANOVA 분석 - 군집분석

## ANOVA 데이터셋

## ANOVA 란 ?

여러 그룹 간의 평균 차이를 비교하는 통계적인 방법이다. 운동처방 그룹 간의 평균 차이를 비교한다.

cluster	0	1	2	3	4	5
성별	M	M	M	F	F	F
나이	높음	낮음	낮음	높음	낮음	낮음
BMI	중간	낮음	높음	중간	낮음	높음

클러스터 내의 'MVM\_PRSCRPTN\_CN'

운동처방 컴벌 데이터를 운동별로 split 하여  
운동처방 전 후 데이터의 몸무게, 체지방율,  
허리둘레, 혈압 데이터를 사용

## Kruskal-Wails 분석 결과

Shapiro-Wilk test(정규성 검정) 결과:  
statistic: 0.9111748933792114  
p-value: 3.413299784858087e-16

Kruskal-Wails 결과  
Kruskal-Wallis Statistic: 964.5978664932767  
P-value (Kruskal-Wallis): 8.643983463638803e-183

클러스터 5에서 실시한 운동별 정규성 검정 결과  
데이터는 정규분포를 따르지 않는다.  
Kruskal-wails 분석에서 통계량은 964로, 그룹 간  
변동이 그룹 내 변동에 비해 크다는 것을 나타낸다.  
또한, 귀무가설인 '그룹 간의 평균 차이가 없다'를  
p-value 0.05보다 작은 값으로 기각하여, 대립가설 '  
운동 그룹 간의 평균 차이가 있다'를 채택한다.

## 운동 추천

MVM_PRSCRPTN_CN	MESURE_1_MESURE_1	MESURE_1EM_002_VALUE_DIFF
무릎 높여 제자리 걸리기	1.00315405	0.055836
벽의 테스트	1.00504797	0.1210536
몸무게 줄이기	1.00974359	0.038034
팔굽혀 펴기	1.0130293	0.057003
맨몸로 주먹으로 그룹	1.00772205	0.2123552
턱걸이	1.01944444	0.0130555
상의 자전거각기	1.0087209	0.047093
제단 뒤로 오르기	1.0155763	-0.04704
누워서 다리 들어올리기	1.0572289	0.1036144
달리기	1.0167882	-0.0003
몸집기 운동	1.02307095	0.1776923
누워서 엉덩이 들어올리기	1.0365239	0.0227959
제단 올라갔다 내려오기	1.0193842	-0.02828
트레드밀에서 걷기	1.0090361	0.00753
일드리 벡지기	1.0339233	0.036726

위의 결과는 cluster5의 운동처방 후  
몸무게 변화량을 오름차순으로 정렬한  
데이터의 일부이다. 위의 데이터를 기  
준으로 현재 몸무게, 키, 체지방율, 혈  
압과 운동 목적(체중감량, 체지방율감  
소 등)에 맞는 n개의 운동을 추천한다.

# 03 데이터 분석 모델링 - 추천 시스템

## 추천시스템 데이터셋

KSPD 국민체육진흥공단

체력측정 및 운동처방 종합데이터

(2019년 4월 ~ 2023년 4월)

각각의 2019년 4월 ~ 2023년 4월  
csv파일을 하나의 데이터셋으로 병합

선택한 운동(체중으로 가정)이 없는 행 drop

837989, 51 (Shape) → 837916

사용자가 선택한 운동의  
이상치를 IQR(1.2)기반으로 제거 (ex: 키)

837918, 51 (Shape) → 814857, 51 (Shape)



MVM\_PRSCRIPTN\_CN\_P 컬럼에서 본문동만 추출

본운동이 없는 행 삭제

아이디값이 겹치는 사용자만 다시  
필터링 (전 후 데이터 생성 목적)

835419, 51 (Shape)

257346, 51 (Shape)

MVM_PRSCRIPTN_CN		본운동	
1	본운동이 없는 행 삭제	1	본운동이 없는 행 삭제
2	본운동이 없는 행 삭제	2	본운동이 없는 행 삭제
3	본운동이 없는 행 삭제	3	본운동이 없는 행 삭제
4	본운동이 없는 행 삭제	4	본운동이 없는 행 삭제
5	본운동이 없는 행 삭제	5	본운동이 없는 행 삭제
6	본운동이 없는 행 삭제	6	본운동이 없는 행 삭제
7	본운동이 없는 행 삭제	7	본운동이 없는 행 삭제
8	본운동이 없는 행 삭제	8	본운동이 없는 행 삭제
9	본운동이 없는 행 삭제	9	본운동이 없는 행 삭제
10	본운동이 없는 행 삭제	10	본운동이 없는 행 삭제

동일한 개체라고 판단되는 동일한 아이디의 중복 값에 (n) 네이밍

(가장 최신 날짜)의 행 저장 후 삭제  
(Ex: AAEx9WqQ11ThUfyp\_1, AAEx9WqQ11ThUfyp\_2, AAEx9WqQ11ThUfyp\_3)

신체데이터가 아닌, 불필요한 컬럼들 삭제  
(측정일련번호, 센터명, 연령대구분명, 측정장소구분명, 입력구분명, 측정일자, 본문동)

최종 데이터셋 Shape  
266116, 51 (Shape)

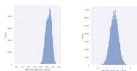
155779, 43 (Shape)

스케일링 및 인코딩

범주형 컬럼 : Label Encoder

연속형 컬럼 : QuantileTransformer

(대부분의 데이터가 비정규분포, 이상치가 많음)



EX) 신장(cm) 컬럼의 스케일링 이후 분포

## 결측치 처리

결측치는 '999999' 값으로 대체

최종 전처리 완료된 데이터셋

ID	SEX	AGE	HEIGHT	WEIGHT	WEIGHT_RATIO	WEIGHT_RATIO_RATIO	WEIGHT_RATIO_RATIO_RATIO
1	1	20	170	60	0.35	0.35	0.35
2	1	20	170	60	0.35	0.35	0.35
3	1	20	170	60	0.35	0.35	0.35
4	1	20	170	60	0.35	0.35	0.35
5	1	20	170	60	0.35	0.35	0.35
6	1	20	170	60	0.35	0.35	0.35
7	1	20	170	60	0.35	0.35	0.35
8	1	20	170	60	0.35	0.35	0.35
9	1	20	170	60	0.35	0.35	0.35
10	1	20	170	60	0.35	0.35	0.35

## 유사도 베이스의 운동 추천 시스템



타 이용자와의 신체데이터 유사도  
베이스의 운동 추천시스템  
(자신이 원하는 신체 데이터 선택)

유사도 베이스의 다양한 운동 추천 최적화 모델  
(여러가지 원하는 신체 데이터  
선택을 통한 효과워주)

전체 신체데이터를 보고 운동처방을 하는 모델

# 03

## 데이터 분석

### 모델링 - 추천 시스템

#### 추천 모델1

##### 코사인 유사도 기반의 추천 시스템 Flow

변화량이 가장 큰 사용자의 운동 추천

1. 신체 데이터의 값과 자신이 관심있는 신체 데이터 컬럼을 입력



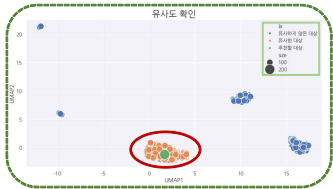
2. 최종 전처리된 데이터셋에서 코사인 유사도를 기반으로 가장 비슷하다고 판단되는 신체 데이터셋의 사용자들 상위 150 표본 선택



3. 선택된 표본의 운동 처방을 받기 이전과 이후의 변화량 수치를 구함



4. 사용자가 선택한 신체 데이터의 변화량이 가장 높은 사용자의 변화량, 운동, 처방 전후의 날짜 차이를 보여줌



실제 사용자가 선택한 상위 유사도 표본과 하위 유사도 표본의 비교 그래프



(UMAP으로 차원 축소)

#### 추천 모델2

변화량이 가장 큰 사용자의 운동 추천

본운동 컬럼에서 구분된  
각 운동의 변화량 평균을 구함



평균 변화량이 높은 운동 5개를 추천

#### 추천 모델3

가장 많이 처방 받은 운동 추천  
(선택한 신체 데이터와 무관)

본운동 컬럼에서 구분된  
각 운동의 카운트를 구함



가장 카운트가 높은 운동 5개를 추천





## 데이터 분석

## 모델링 - 추천 시스템(군집분석)

## 일반 및 변화량 기반 추천 머신러닝 모델

1. 각각 분리되어 있는 2023.04~2019.04 기간의 데이터셋을 하나로 병합

2.아래와 같은 예측에 불필요한 컬럼들을 제거

```
['MBER_SEQ_NO_VALUE','MESURE_SEQ_NO','CNTNR_NM','AGRDE_FLAG_NM',  
 'MESURE_PLACE_FLAG_NM','INPT_FLAG_NM','MESURE_DE']
```

3. 준비운동, 본운동, 마무리운동으로 되어있는 컬럼을 본운동만 필터링하여 "."를 구분자로 하여 아래와 같이 전처리

```

graph LR
    User[MEM.PRESIDENT.CHN] --> Server[MEM.PRESIDENT.CHN]
    Server --> Database[MEM.PRESIDENT.CHN]
  
```

The diagram illustrates the data flow for a specific user. It starts with a user labeled 'MEM.PRESIDENT.CHN' on the left. An arrow points from the user to a server, also labeled 'MEM.PRESIDENT.CHN'. From the server, another arrow points to a database, also labeled 'MEM.PRESIDENT.CHN'. The database is represented as a cylinder with a grid of data points.

4. 이후 모든 이상치가 많은, 아래와 같은 5개 컬럼에 대해 IQR(1.2)기반으로 이상치를 제거

```
['MESURE_IEM_001_VALUE','MESURE_IEM_002_VALUE',
'MESURE_IEM_003_VALUE','MESURE_IEM_004_VALUE','MESURE_AGE_CO']
```

5. 범주형 컬럼에 대해서는 **Labelencoding**을,  
연속형 컬럼에 대해서는 **StandardScarlring**을 적용

6. X데이터셋은 신체 측정 데이터 컬럼 (818, 43)  
Y데이터셋은 전처리한 각각의 본운동 컬럼 으로 구성 (818,100)

7. Y데이터컬럼이 하나가 아니라 여러 개로 되어있기 때문에, **LightGBM+MultiOutputClassifier**을 사용하여 모델링을 진행

## 모델1

이 모델은 사용자가 입력한 신체데이터셋에 따른, 추천 운동이 출력으로 나옴,  
테스트셋(30%)의 평균 정확도는 약 0.94로 매우 좋은 성능을 가지는 것을 볼 수 있음

```
from sklearn.metrics import accuracy_score

# 예측 결과와 실제값의 각 출력행에 있는 지표 계산
accuracy_scores = []

for i in range(y_test.shape[1]): # 각 출력 행에 대해 (반복)
    accuracy = accuracy_score(y_test.iloc[:, i], predictions[:, i])
    accuracy_scores.append(accuracy)

# 전체 평균을 얻는 지표 계산
overall_accuracy = np.mean(accuracy_scores)

# 결과 출력
print("평균 정확도:", overall_accuracy)

평균 정확도: 0.9492276427704228
```

## 모델2

즉, 여기서는 모델의 INPUT값을 변경하여 사용자가 원하는 신체 변화량 조합을 입력하게 되면, 출력으로 '본문등' 조합이 나오게 설계하였음

```
from sklearn.metrics import accuracy_score

# 예측 결과를 사용하여 각 출력별로 성능 지표 계산
accuracy_scores = []

for i in range(y_test.shape[1]):
    accuracy = accuracy_score(y_test.iloc[:, i], predictions[:, i])
    accuracy_scores.append(accuracy)

# 전체 평균 성능 지표 계산
average_accuracy = np.mean(accuracy_scores)

# 결과 출력
print("평균 정확도: ", average_accuracy)

평균 정확도: 0.9378103916866507
```

150, 43  
(X: 각 신체 데이터 변화량)

150, 139  
(Y: 보은동)

# 03 데이터 분석 모델링 – XAI 모델



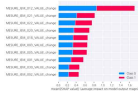
## SHAP(SHapley Additive exPlanations) ?

기법은 기계 학습 모델의 예측에 대한 해석을 제공하는 도구다.

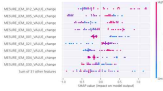
- 1.정확성과 일관성: SHAP는 기여도를 정확하고 일관되게 계산함. 즉, 모델 예측에 대한 각 특성의 기여도를 신뢰할 수 있게 평가 가능함
- 2.해석 용이성: 특성의 중요도와 영향 방향을 명확하게 이해할 수 있어, 비전문가도 결과를 쉽게 해석할 수 있음
- 3.모델 무관성: SHAP는 다양한 종류의 기계학습 모델에 적용 가능함. 따라서 복잡한 모델이나 알고리즘에도 유연하게 사용될 수 있음
- 4.로컬 해석과 글로벌 해석: 개별 예측 뿐만 아니라 모델 전체의 행동을 해석하는 데에도 유용함. 이를 통해 모델의 전반적인 동작을 이해할 수 있음

## 변화량 기반의 운동 추천 모델의 줄넘기 모델(Testset 1번 인덱스)

Summary\_plot



Beeswarm Plot



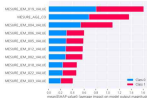
Waterfall Plot



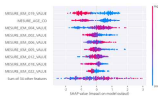
1. 변화량 데이터 기반 줄넘기 운동 예측 모델의 경우에는, 전체 피쳐의 중요도를 확인해봤을 때, “앞아랫몸앞으로굽히기” 피쳐가 예측에 가장 큰 영향을 주었음. (Summary\_plot)
2. 해당 피쳐는 줄넘기 처방과 대략적으로 양의 관계를 가지고 있어서, “윗몸일으키기”운동을 많이 했을 때, “줄넘기” 운동의 추천 확률도 올라감(Beeswarm\_Plot)
3. 실제 건강센터와 학습된 모델에서 줄넘기 운동을 추천 받은 1번 사용자 데이터를 보면, “앞아랫몸앞으로굽히기”, “허리둘레”, “이완기최저혈압” 신체지수가 모델의 출력값에 유의미한 영향을 주었다고 판단됨 (Waterfall Plot)

## 일반 신체 데이터로 기반의 추천 모델의 줄넘기 모델(TestSet 99번 인덱스)

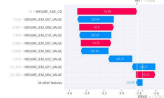
Summary\_plot



Beeswarm Plot



Waterfall Plot



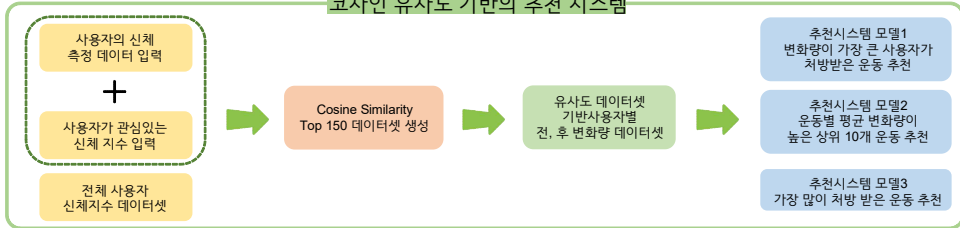
1. 신체 데이터 기반 줄넘기 운동 예측 모델의 경우에는, 전체 피쳐의 중요도를 확인해봤을 때, “교차윗몸일으키기(회)” 피쳐가 예측에 가장 큰 영향을 주었음 (Summary\_plot)
2. 해당 피쳐는 줄넘기 처방과 대략적으로 음의 관계를 가지고 있어서 윗몸일으키기를 많이 하면, 줄넘기 운동의 추천 확률은 낮아진다고 할 수 있음 (Beeswarm\_Plot)
3. 실제 건강센터와 학습된 모델에서 줄넘기 운동을 추천 받은 99번 사용자 데이터를 보면, “나이”, “신장”, “악력\_우” 신체지수가 모델의 출력값에 유의미한 영향을 주었다고 판단됨 (Waterfall Plot)

# 03

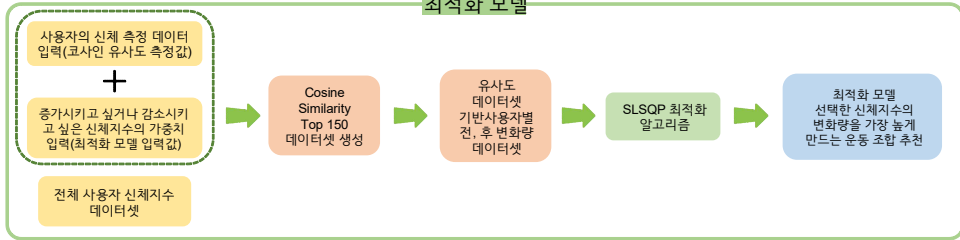
## 데이터 분석

모델링 - 추천 모델 순서도

### 코사인 유사도 기반의 추천 시스템



### 최적화 모델

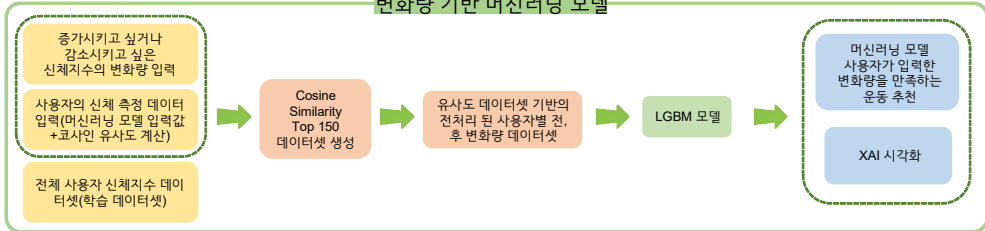


# 03

## 데이터 분석

모델링 - 추천 모델 순서도

### 변화량 기반 머신러닝 모델



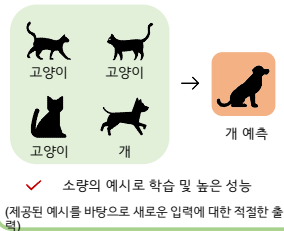
### 일반 추천 머신러닝 모델



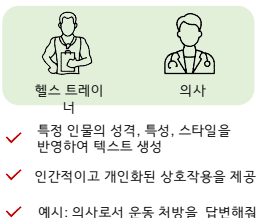
# 03 데이터 분석 모델링 - 챗봇 모델

Few shot learning과 Persona가 적용된 사용자 입력을 Bard API에 전달

## Few shot learning



## Persona(퍼소나)



## 사용자 입력(Input)

(목적항목1) 신장(cm)을 입력하세요.

0.00 - +

(목적항목2) 체중(kg)을 입력하세요.

0.00 - +

(목적항목3) bmi를 선택하세요.

비만 ▾

(목적항목4) 유언장을 선택하세요.

종류 ▾

(목적항목5) 지구력을 선택하세요.

종류 ▾

✓ 쉽고 간편한 입력

✓ 최대 추천 운동 개수 설정



- ✓ Few shot learning으로 사용자에게 개인화된 운동 추천
- ✓ Persona로 사용자의 요구와 선호에 맞는 맞춤형 운동 추천
- ✓ 사용자 입력으로 사용자에게 개인화된 운동 추천(성별, 연령대 등)
- ✓ 누구나 Bard API를 사용하면 무료로 온라인에서 맞춤형 운동 추천

웹앱 링크: <https://homepy-x9czsew7kciwjrxnbcbr.streamlit.app/>

Azure 서버 링크: <http://20.214.137.21:7777>

## 04

## 활용방안&amp;기대효과

Streamlit 웹앱(메인 페이지)

## 사용자 데이터 입력

## 데이터 분석을 통한 맞춤형 운동 처방

원하는 추천 방식을 선택하세요.

추천시스템 기반 맞춤형 운동 추천

[측정항목1] 신장(cm)을 입력하세요.

0.00

[측정항목2] 체중(kg)을 입력하세요.

0.00

[측정항목3] 체지방률(%)을 입력하세요.

0.00

[측정항목4] 허리둘레(cm)을 입력하세요.

0.00

[측정항목5] 이완기최저혈압(mmHg)을 입력하세요.

0.00

[측정항목6] 이완기최고혈압(mmHg)을 입력하세요.

0.00

[측정항목7] 약력\_좌(kg)을 입력하세요.

0.00

[측정항목8] 약력\_우(kg)을 입력하세요.

0.00

[측정항목9] 뒤통알아올리기(회)를 입력하세요.



제출 버튼 클릭

## 맞춤형 운동 추천

[측정항목36] 스태프관 산소맥( $\text{VO}_{2\text{max}}$ )을 입력하세요.

0.00

[측정항목37] 허벅지\_좌(cm)을 입력하세요.

0.00

[측정항목38] 허벅지\_우(cm)을 입력하세요.

0.00

[측정항목39] 전신반응(초)을 입력하세요.

0.00

[측정항목40] 성인체중시간(초)을 입력하세요.

0.00

[추천 시스템] 관심있는 신체 지수를 선택하세요.

허리둘레

[추천 시스템] 운동을 통해 어떤 효과를 얻길 원하시나요?

플러스(+)

제출

주변 지도

XAI&amp;Profiling

국민체육진흥공단

## 멀티 페이지 구성

✓ 맞춤형 운동 추천

운동 추천 챗봇

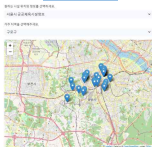
주변 지도

국민체력100 Youtube

XAI&Profiling

5페이지로 구성하여 다양한 정보 제공

## 주변 지도



다양한 플랫폼에서 수집한  
위치 정보 & 데이터프레임 제

공

## 맞춤형 운동 추천(Main)

데이터 분석을 통한 맞춤형 운동 처방

원하는 추천 방식을 선택하세요.

비드 API를 활용한 맞춤형 운동 추천

(국정항공) 신장(cm)을 입력하세요.

0.00

(국정항공) 체중(kg)을 입력하세요.

0.00

제출

사용자에게 값을 입력 받고  
다양한 모델(추천 시스템, 챗봇 등)로  
제출 버튼을 누르면 **맞춤형 운동 추천**

## XAI&amp;Profiling

Feature Importances

Classification Stats

Individual Predictions

What if...

Feature Dependence

Feature Interactions

Decision Trees

모델에 종속적이지 않은  
피쳐 임포턴스 산출기법 제공

## 국민체력100 유튜브

국민체력100 유튜브 채널을 선택해주세요.

국민체력100

국민체력100 유튜브 동영상을 선택해주세요.

백세만사 운동처방

국민체력100 유튜브 소문류를 선택해주세요.

김복문동

국민체력100 유튜브 채널을 선택해주세요.

진중혁 버프가 필요할 때? (ep1. 뇌를 자극시키는 집중력...



선택한 분류에 따라 원하는  
**국민체력100 유튜브** 시청 가능  
(국민체력100 운동처방 동영상주소 정보 데이터)

단순히 분석에서 끝나지 않고 프로토타입 웹앱에서 다양한 서비스 제공

\* 배포된 웹앱 우측 상단의 깃허브 링크에서 코드 확인과 다운로드하여 수정 및 활용 가능

# 04

## 활용방안&기대효과

Streamlit 웹앱(맞춤형 운동 추천 챗봇)

### 사용자 데이터 입력

### 운동 추천 결과

#### 사용 방법

#### 운동 추천 Chatbot

나에게 맞춤형 운동을 추천해줄 수 있어?



추천할 운동 개수를 알려주세요.(최대 5개)



대화 초기화

챗봇과 대화하기:

추천 받을 운동 개수 입력 (예시: 3개)

바드 API를 활용한 운동 추천  
(챗봇 실제 실행 영상)

#### 데이터 분석을 통한 맞춤형 운동 처방

원하는 추천 방식을 선택하세요.

바드 API를 활용한 맞춤형 운동 추천

[운동할 목적] 신장(km)을 입력하세요.

1000

[운동할 목적] 체중(kg)을 입력하세요.

0.00

[운동할 목적] bmi를 선택하세요.

비만

[운동할 목적] 유연성을 선택하세요.

중음

[운동할 목적] 지구력을 선택하세요.

중음

[운동할 목적] 근력을 선택하세요.

중음

[운동할 목적] 운동 강도를 선택하세요.

상급자

[운동할 목적] 선호하는 운동 환경을 선택하세요.

실내 운동

[운동할 목적] 주로 운동하는 시간을 선택하세요.

[운동할 목적] 지구력을 선택하세요.

보통

[운동할 목적] 근력을 선택하세요.

보통

[운동할 목적] 운동 강도를 선택하세요.

중급자

[운동할 목적] 선호하는 운동 환경을 선택하세요.

실내 운동

[운동할 목적] 주로 운동하는 시간을 선택하세요.

저녁

[운동할 목적] 운동 목적을 선택하세요.

근력 증진

제출

당신의 성별은 남성이며, 연령대는 10대입니다.

운동 추천 챗봇 버튼을 클릭하세요. 챗봇 페이지로 이동합니다.

운동 추천 챗봇

주변 지도

국민체육100\_YouTube

Made with Streamlit

사용자는 쉽고 간편하게 채팅 한 번으로 Few shot learning과 Persona가 적용된 맞춤형 운동을 추천 받음



# 04

## 활용방안&기대효과

Streamlit 웹앱(맞춤형 운동 설명 챗봇)

### 사용 방법

#### 운동 설명 Chatbot

운동하는 방법에 대해 자세하게 알고 싶어, 어떻게 하면 될까?



궁금한 운동을 입력해주세요.

대화 초기화

챗봇과 대화하기:

설명이 궁금한 운동 입력 (예시:스카프 저글링)



### 사용자 데이터 입력

원하는 챗봇을 선택하세요.

바드 API를 활용한 운동 설명 챗봇

#### 운동 설명 Chatbot

운동하는 방법에 대해 자세하게 알고 싶어, 어떻게 하면 될까?



궁금한 운동을 입력해주세요.

대화 초기화

챗봇과 대화하기:

Made with Streamlit



### 운동 설명 결과

원하는 챗봇을 선택하세요.

바드 API를 활용한 운동 설명 챗봇

#### 운동 설명 Chatbot

운동하는 방법에 대해 자세하게 알고 싶어, 어떻게 하면 될까?



궁금한 운동을 입력해주세요.

대화 초기화

챗봇과 대화하기:

Made with Streamlit

바드 API를 활용한 운동 설명  
(챗봇 실제 실행 영상)

사용자는 쉽고 간편하게 채팅 한 번으로 Few shot learning과 Persona가 적용된 맞춤형 운동을 추천 받음



## 국민체육진흥공단

- 국민체력100 유튜브 활성화에 기여
- 체력측정 및 운동 처방 종합데이터  
활용 및 개선 제안
- 국민체력100 프로젝트 지원



## 사용자

- 사용자(국민)의 체력 증진 효과
- 챗봇을 사용해 쉽게 운동에 대해 배울 수 있음
- 시공간 제약 없이 무료로 운동 추천 서비스 이용
- 기존에 운동 결과에서 운동 능력이 향상 된 후  
새로운 운동을 하고 싶을 때 활용 가능



## 맞춤형 운동 추천 서비스

- 누구나 쉽고 간편하게 무료로 이용
- 다양한 인공지능으로 운동 추천
- 운동에 대해 잘 모르는 사람들에게  
어떤 운동이 필요한 지 정보 제공