

FastCampus Presentation

# 신사숙녀 W Team 3

# CONTENTS

|   |   |
|---|---|
| 1 | 프로젝트 배경   |
| 2 | 머신러닝 이론   |
| 3 | 개발 진행 방식 및 데이터 설명   |
| 4 | 개인 맞춤 상품 추천 모델 (1), (2)<br>[K-Means]                      |
| 5 | 신체 치수를 이용한 체형 분류 모델 - 머신러닝<br>[RandomForest, K-Means]     |
| 6 | 신체 치수를 이용한 체형 분류 모델 - 딥러닝<br>[Multilayer Perceptron, MLP] |
| 7 | 긍정, 부정 후기 판단 모델<br>[DistilBert]                           |
| 8 | 프로젝트 회고   |

신사속녀

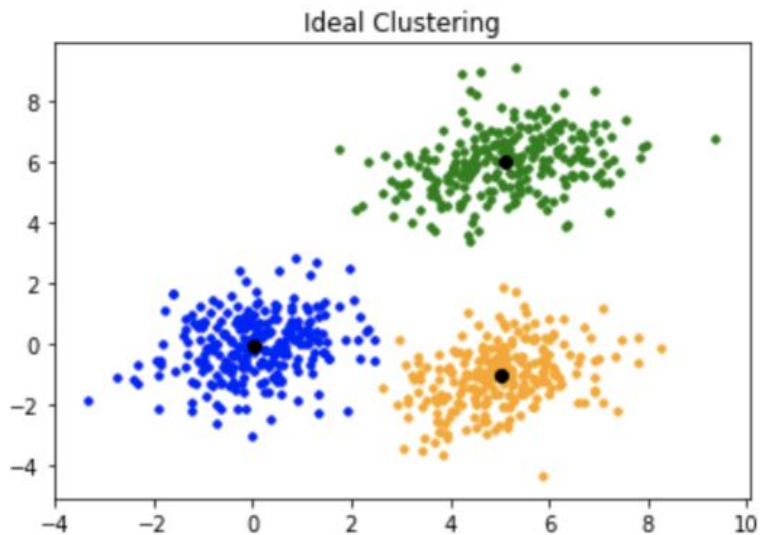
## 프로젝트 배경

MUSINSA

신사숙녀  
W

유행보다 개인의 개성과 취향이 중요해진 시대,  
AI와 함께 2030 고객에게 맞춤 상품을 추천합니다.

# 머신러닝 이론



## K-Means ?

- 비지도 학습
- 클래스 레이블
- 거리 기반(유클리디안)
- 클러스터링
- K개의 군집
- 엘보우 기법 & 실루엣 기법

## 개발 진행 방식 및 데이터 설명

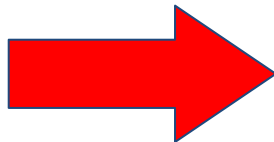
1. 구현한 이커머스 사이트에서 추가적으로 필요한 서비스 추출
2. Kaggle에서 상품 데이터(여성 의류 쇼핑몰 후기) 수집
3. 상품 데이터 개별 전처리
4. 개별 모델 학습 및 모델 튜닝
5. 여러 결과 중 근거 있는 2개 모델을 선정하여 최종 분석 및 시각화 진행

## 개발 진행 방식 및 데이터 설명

```
총 행 개수 (rows): 192544  
총 컬럼 수 (columns): 15
```

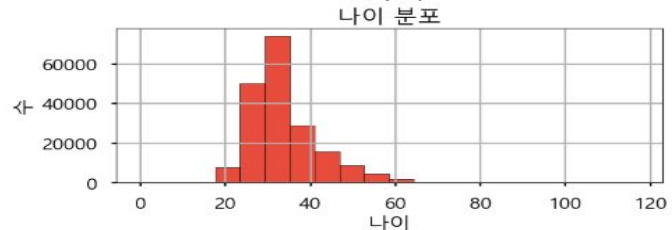
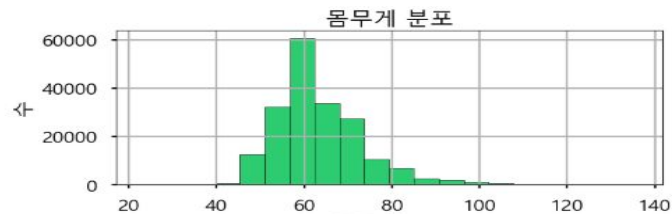
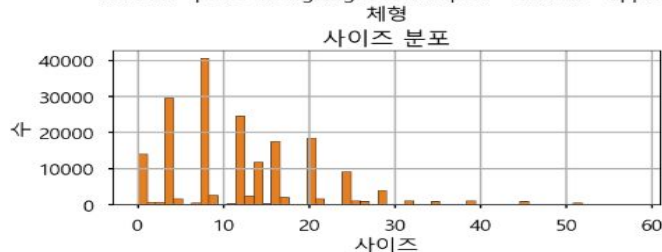
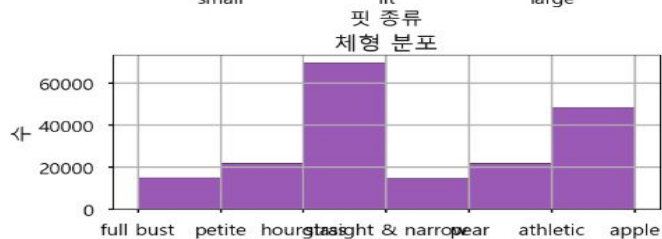
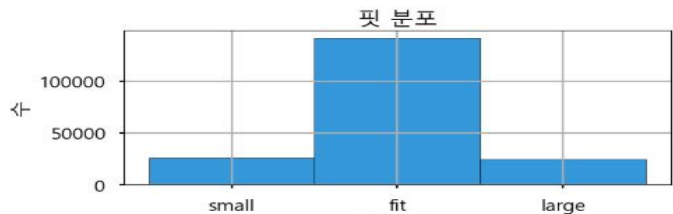
각 컬럼의 데이터 타입:

```
fit            object  
user_id        int64  
bust size      object  
item_id        int64  
weight         object  
rating         float64  
rented for     object  
review_text    object  
body type      object  
review_summary object  
category       object  
height         object  
size           int64  
age            float64  
review_date    object  
dtype: object
```



```
fit            object  
weight         object  
rating         float64  
body type      object  
height         object  
size           int64  
age            float64  
dtype: object
```

# 개발 진행 방식 및 데이터 설명



# 개인 맞춤 상품 추천 모델 [K-Means]

(1)

## 주요 feature

- 수치형 : bmi, size, age
- 범주형 : fit, body type (원 핫 인코딩)

## 결측치 처리

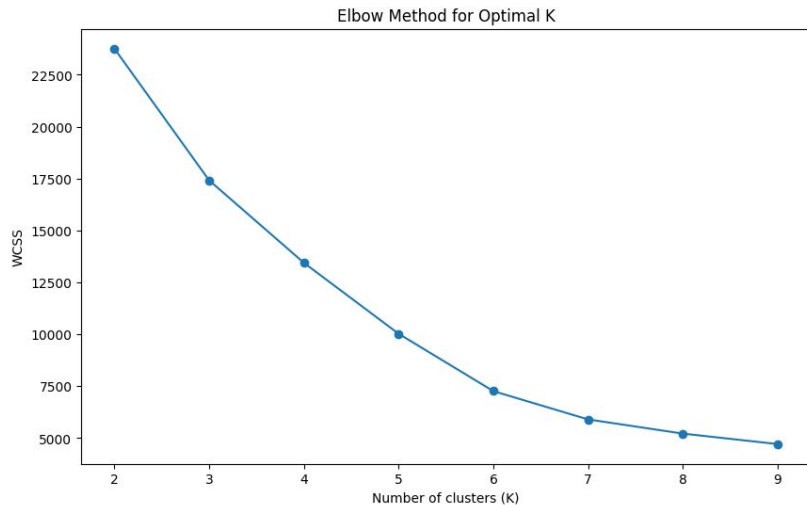
- 수치형 : 평균값 대체
- 범주형 : 최빈값 대체

## 범주형 컬럼 가중치 처리

- $(\text{원 핫 인코딩 후 생긴 컬럼 수}) / (\text{전체 컬럼 수})$   
-> 범주형 컬럼이 너무 많은 영향을 주기 때문에

스케일링 방법 : MinMaxScaler (수치형)

## Elbow Method 시각화



최적의 k=7

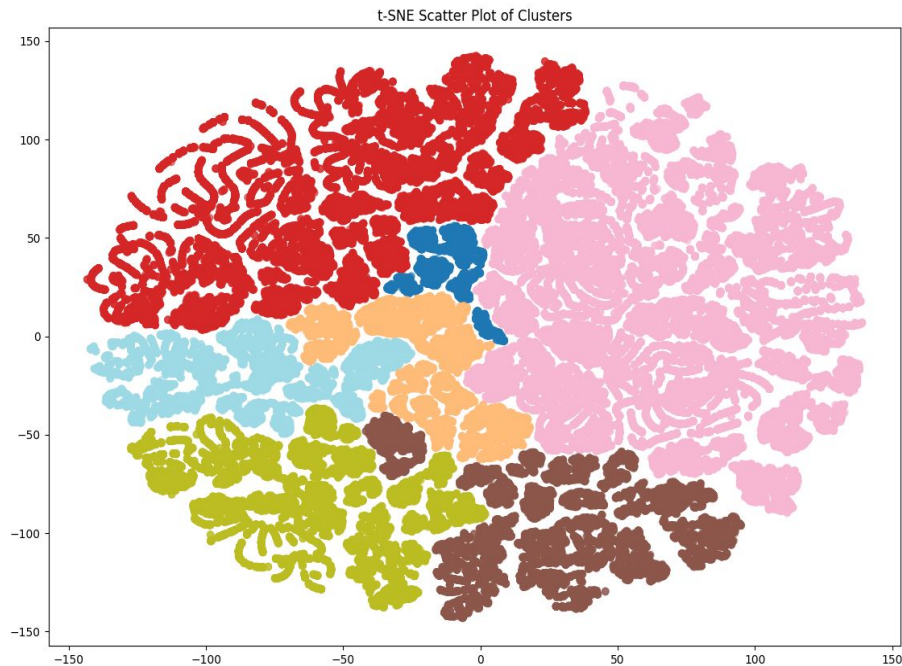
실루엣 점수 : 0.665



# 개인 맞춤 상품 추천 모델 [K-Means]

(1)

클러스터 간 시각적 경계가  
명확히 형성되며, **body type**  
중심의 구분이 시각적으로 드러남



t-SNE 시각화

# 개인 맞춤 상품 추천 모델 [K-Means]

(1)

cluster 데이터 평균값 시각화

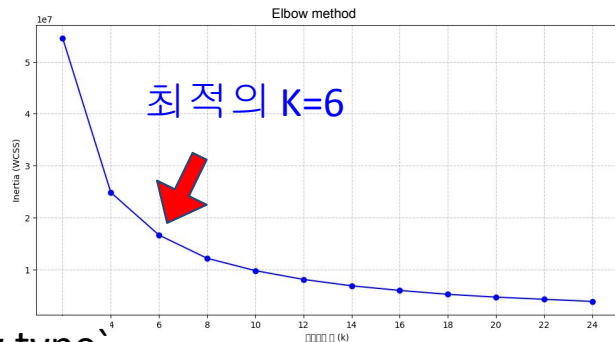
| cluster | bmi   | size  | age   | fit_fit | fit_large | fit_small | body type_apple | body type_athletic | body type_full bust | body type_hourglass | body type_pear | body type_petite | body type_straight & narrow |
|---------|-------|-------|-------|---------|-----------|-----------|-----------------|--------------------|---------------------|---------------------|----------------|------------------|-----------------------------|
| 0       | 25.65 | 21.22 | 36.63 | 0.16    | 0.03      | 0.04      | 0.54            | 0.00               | 0.00                | 0.00                | 0.00           | 0.00             | 0.00                        |
| 1       | 24.26 | 17.88 | 35.02 | 0.17    | 0.03      | 0.03      | 0.00            | 0.00               | 0.54                | 0.00                | 0.00           | 0.00             | 0.00                        |
| 2       | 21.98 | 10.68 | 33.72 | 0.17    | 0.03      | 0.03      | 0.00            | 0.54               | 0.00                | 0.00                | 0.00           | 0.00             | 0.00                        |
| 3       | 23.46 | 14.04 | 33.77 | 0.17    | 0.03      | 0.03      | 0.00            | 0.00               | 0.00                | 0.00                | 0.54           | 0.00             | 0.00                        |
| 4       | 23.28 | 14.05 | 33.87 | 0.17    | 0.03      | 0.03      | 0.00            | 0.00               | 0.00                | 0.54                | 0.00           | 0.00             | 0.00                        |
| 5       | 21.38 | 5.54  | 33.20 | 0.17    | 0.03      | 0.03      | 0.00            | 0.00               | 0.00                | 0.00                | 0.00           | 0.54             | 0.00                        |
| 6       | 19.94 | 7.28  | 33.36 | 0.17    | 0.03      | 0.03      | 0.00            | 0.00               | 0.00                | 0.00                | 0.00           | 0.00             | 0.54                        |

각 cluster는 고객의 평균 bmi, size, age와 body type, fit에 따라 구분  
-> 이러한 클러스터링 결과는 고객 체형에 따른 맞춤 상품 추천에 활용 가능

## 개인 맞춤 상품 추천 모델 [K-Means] (2)

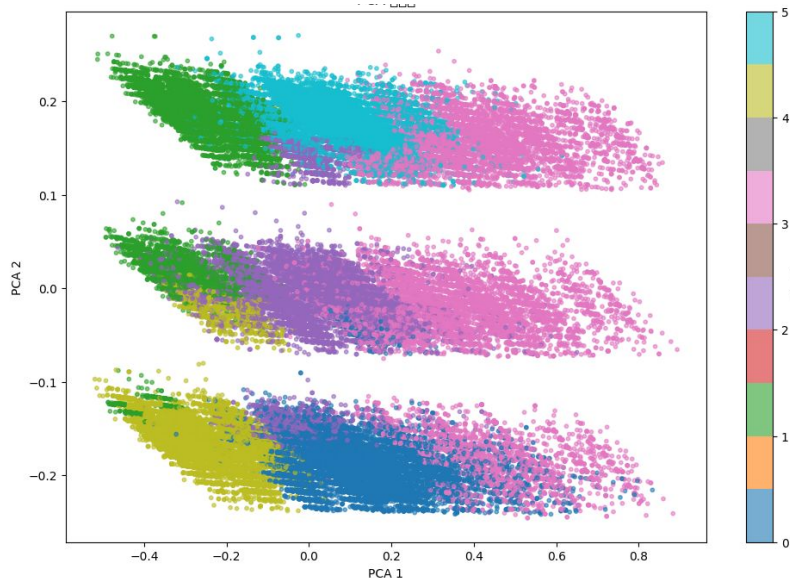
### 1. 도메인 지식 반영 및 데이터 구조

- 쇼핑몰 타겟: **20-30대 여성**
  - 사용 특성 6개: `weight`, `height`, `size`, `age`, `body type`
  - 추천 지표: `fit = fit`, `rating >= 8점 이상만`
  - 2만 샘플 · **15만개** 전체 데이터 모두 유사한 군집 구조 확인
  - 실루엣 점수 **0.3 ~ 0.4** (대규모 데이터에서 준수)
  - 대용량 데이터 -> 실루엣 점수 다소 낮음
- > 실제 해석력, 시각화, 도메인 지식 등 다양한 지표로 종합 평가 <sup>1)</sup>



## 개인 맞춤 상품 추천 모델 [K-Means] (2)

### 2. 시각화(1) - PCA (전역 구조)



### 체형 기반 계층 구조 파악

- PCA1(가로)  
size, weight, height 의 조합 → 3층 구조
- PCA2(세로)  
age 그룹 → 층 내에서 20-30대 나이대 혼합
- 결론: 체형(사이즈, 몸무게, 키)끼리  
한 층 형성, 각 층 내 다양한 나이대 혼합
- 실무적 방안: 재고/사이즈 기획  
→ **3개 체형 그룹별 생산·재고 전략**

## 개인 맞춤 상품 추천 모델 [K-Means] (2)

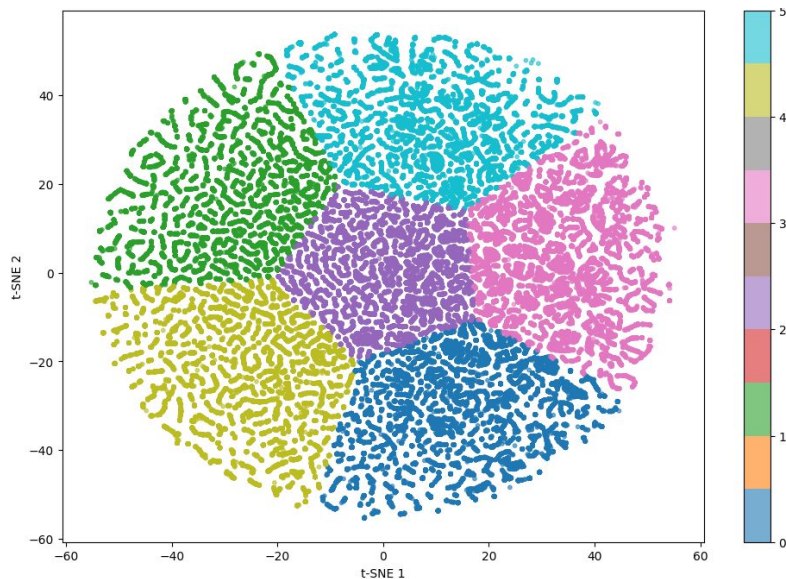
|                | PC1(41.3) | PC2(17.2) |
|----------------|-----------|-----------|
| height         | 0.273179  | -0.153569 |
| weight         | 0.453205  | -0.028093 |
| size           | 0.839412  | 0.007897  |
| age_20-24      | -0.015037 | -0.007373 |
| age_25-29      | -0.043259 | -0.687245 |
| age_30-34      | 0.028013  | 0.708798  |
| age_35-39      | 0.030283  | -0.01418  |
| body_apple     | 0.010802  | 0.002661  |
| body_athletic  | -0.009848 | -0.009955 |
| body_full bust | 0.025717  | 0.009163  |
| body_hourglass | 0.059217  | -0.006853 |
| body_pear      | 0.017661  | -0.002897 |
| body_petite    | -0.079272 | 0.01684   |
| body_straight  | -0.024276 | -0.008959 |

### (1) PCA (전역 구조) : 체형 기반 계층 구조 파악

- PC1+PC2 = 누적 분산 비율 58.45%  
→ 데이터의 절반 이상 설명, 유의미한 지표
- PC1(41.27%)  
size(0.839), weight(0.453), height(0.273)  
→ 체형 크기 관련 변수
- PC2(17.18%)  
age\_30-34(0.709), age\_25-29(-0.687)  
→ 나이 관련 변수

# 개인 맞춤 상품 추천 모델 [K-Means] (2)

## 2. 시각화(2) - t-SNE (국소 구조)



### 세부 세그먼트별 타겟팅 가능

- 6개의 뚜렷한 고객 세그먼트가 식별됨
- 결론: 각 세그먼트는 고유한 체형·사이즈·나이대 조합을 가짐
- 실무적 방안: 마케팅 / 제품 기획  
→ **6개 세그먼트별 타겟팅**

## 개인 맞춤 상품 추천 모델 [K-Means] (2)

### (2) t-SNE (국소 구조): 세부 세그먼트별 타겟팅 가능

- 클러스터 0: 모래시계/운동형/배 체형, 25-29세(98%), 평균 사이즈 13.84, 몸무게 65.71kg
- 클러스터 1: 작은체형/운동형, 30-34세(71%), 평균 사이즈 3.84, 몸무게 53.72kg
- 클러스터 2: 모래시계/운동형, 35-39세 비중 높음, 평균 사이즈 10.38, 몸무게 61.29kg
- 클러스터 3: 모래시계/풀버스트 체형, 평균 사이즈 21.67, 몸무게 74.30kg
- 클러스터 4: 작은체형 다수, 25-29세(91%), 평균 사이즈 4.60, 몸무게 55.19kg
- 클러스터 5: 모래시계/배형, 100% 30-34세, 평균 사이즈 12.67, 몸무게 63.32kg

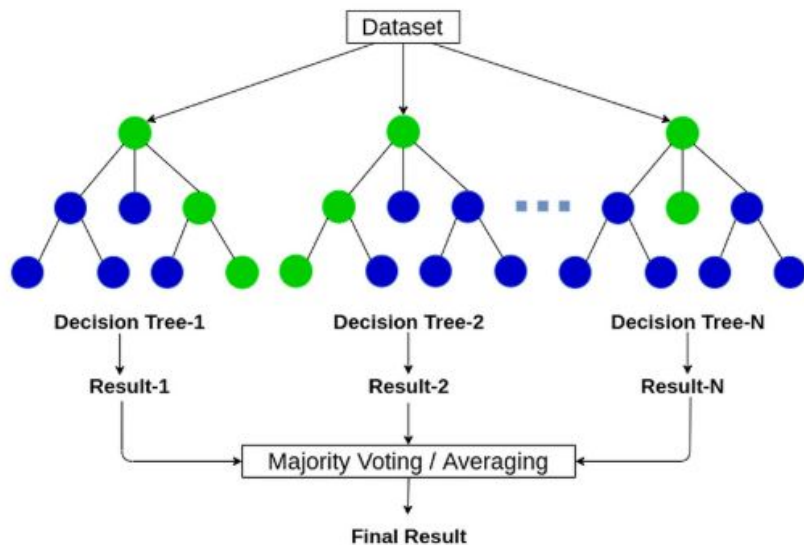
# NEXT

|   |   |
|---|---|
| 1 | 프로젝트 배경   |
| 2 | 머신러닝 이론   |
| 3 | 개발 진행 방식 및 데이터 설명   |
| 4 | 개인 맞춤 상품 추천 모델 (1), (2)<br>[K-Means]                      |
| 5 | 신체 치수를 이용한 체형 분류 모델 - 머신러닝<br>[RandomForest, K-Means]     |
| 6 | 신체 치수를 이용한 체형 분류 모델 - 딥러닝<br>[Multilayer Perceptron, MLP] |
| 7 | 긍정, 부정 후기 판단 모델<br>[DistilBert]                           |
| 8 | 프로젝트 회고   |

신사속녀



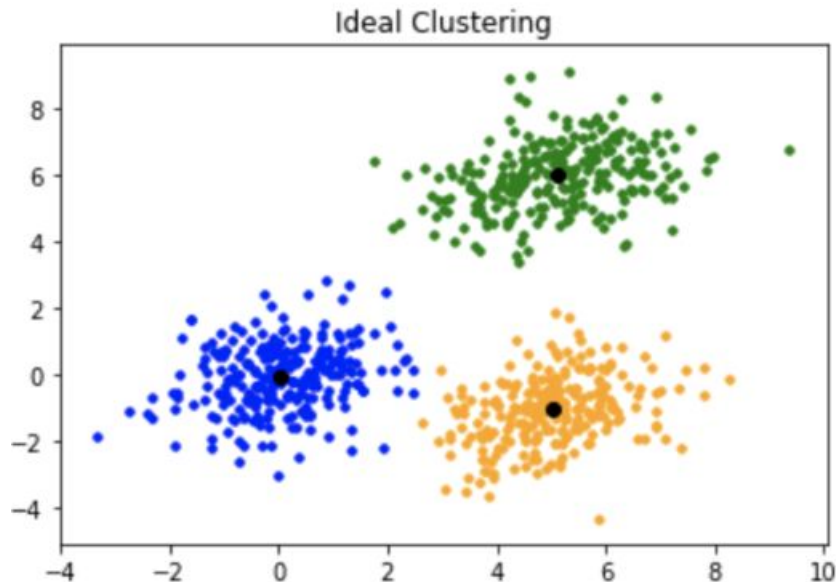
## 신체 치수를 이용한 체형 분류 모델 (1) - [RandomForest, K-Means]



### 1. 랜덤 포레스트 (Random Forest)

- 알고리즘 선택 이유: 다양한 신체 지표들의 복합적 관계를 잘 학습하면서, 과대적합 없이 바디타입을 안정적으로 분류할 수 있어 이 프로젝트에 적합하다고 판단

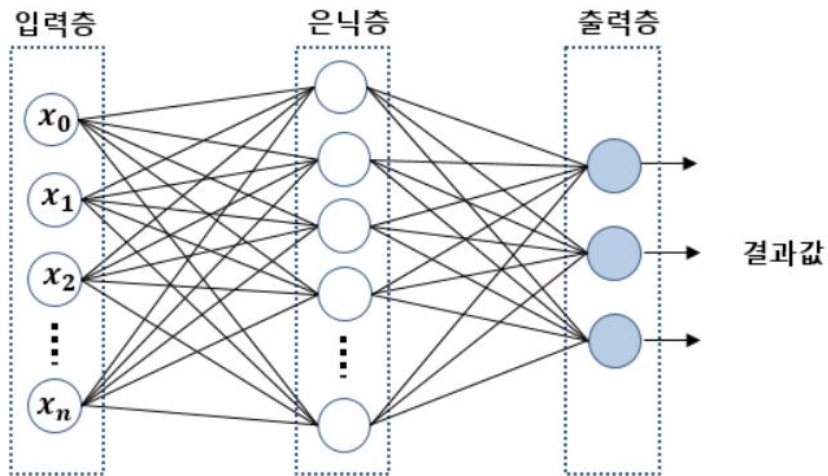
## 신체 치수를 이용한 체형 분류 모델 (2) - [RandomForest, K-Means]



### 2. K-평균 (K-Means)

- 알고리즘 선택 이유: 신체 측정치만으로 자연스럽게 사람들을 그룹화하고, 군집 대표 체형을 직관적으로 분석할 수 있어 이 프로젝트에 적합하다고 판단

## 신체 치수를 이용한 체형 분류 모델 (3) - [Multilayer Perceptron, MLP]



### 3. 다층 퍼셉트론

#### (Multilayer Perceptron, MLP)

- 알고리즘 선택 이유: 복잡한 수치형 피쳐들 사이의 비선형 관계를 스스로 학습하기 위해

## 신체 치수를 이용한 체형 분류 모델 (4) - 기본 특성 / 파생 특성

샘플 개수: 20000

특성 정보(컬럼명/타입)

Gender object

Weight 사용 X int64

Body Shape Index int64

Waist float64

Hips float64

Bust/Chest float64

Height float64

Cup Size 파생 특성 object

BMI float64

Body Shape Index

0 1823

1 7370

2 1585

3 3085

4 6137

우리 쇼핑몰의 바디  
타입 종류는 총 7개,  
즉 기존에 존재하는  
타겟과는 맞지 않음  
→ 새로운 타겟 생성

# 12. BMI 계산해서 추가

```
data['BMI'] = data['Weight'] / ((data['Height'] / 100) ** 2)
```

기존에 존재하는 '몸무게'와 '키'를 이용해서 계산

## 신체 치수를 이용한 체형 분류 모델 (5) - 특성 전처리 / 타겟 생성

샘플 개수: 20000

특성 정보(컬럼명/타입)

Gender int64 ← object

Weight int64

Waist float64

Hips float64

Bust/Chest float64

Height float64

Cup Size int64 ← object

BMI 새로운 타겟 생성 float64

body\_type object

인코딩

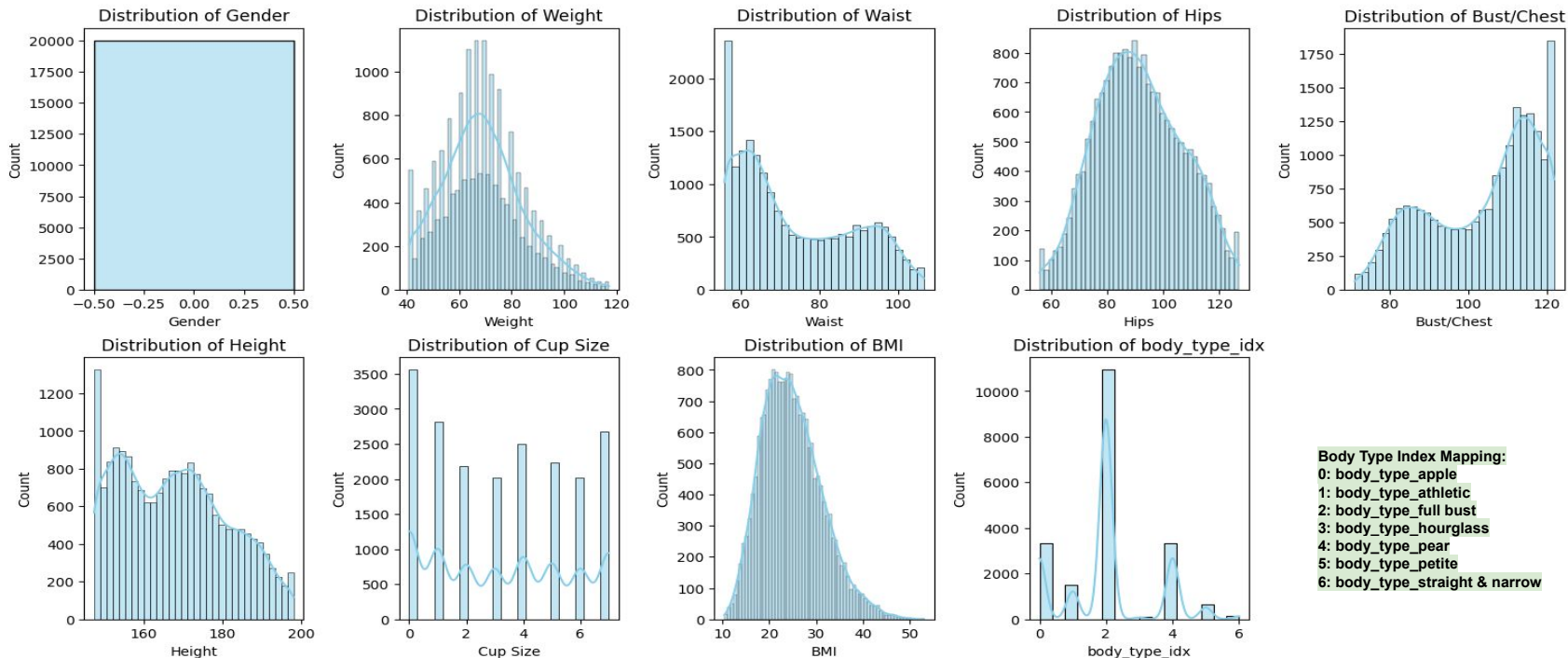
```
def classify_body_type(row): 1개의 사용 위치 kimyelin0506
    waist = row['Waist']
    hips = row['Hips']
    bust = row['Bust/Chest']
    height = row['Height']

    if bust > hips and waist / bust > 0.8:
        return 'body_type_apple'
    elif bust > hips and waist / bust <= 0.8:
        return 'body_type_full bust'
    elif hips > bust and waist / hips <= 0.75:
        return 'body_type_pear'
    elif abs(bust - hips) < 3 and waist / hips < 0.8:
        return 'body_type_hourglass'
    elif height < 157: # 157cm 이하
        return 'body_type_petite'
    elif abs(bust - hips) < 2 and waist / hips >= 0.8:
        return 'body_type_straight & narrow'
    else:
        return 'body_type_athletic'
```

지도 학습을  
위한 타겟 생성

계산식:  
현업/논문 참고<sup>2)</sup>

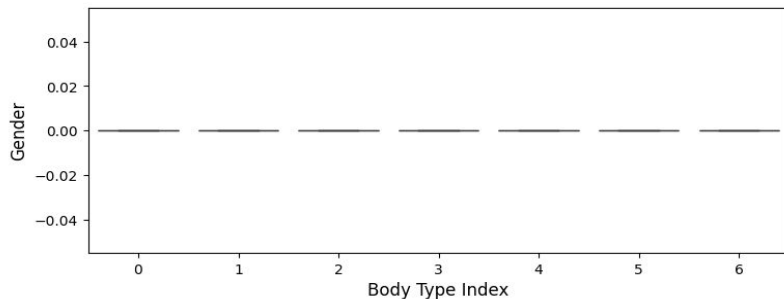
## 신체 치수를 이용한 체형 분류 모델 (6) - 데이터 분포 시각화



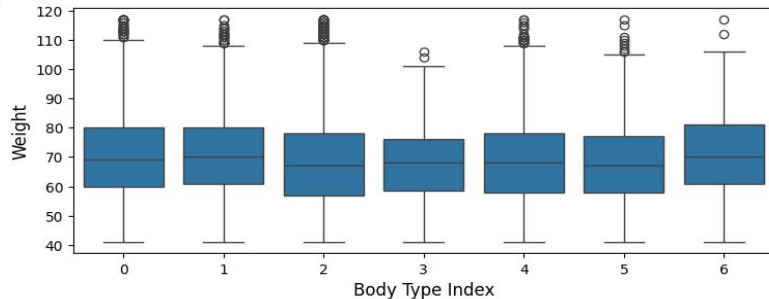
# 신체 치수를 이용한 체형 분류 모델 (7) - 이상치 시각화

Body Type Index Mapping:  
0: body\_type\_apple  
1: body\_type\_athletic  
2: body\_type\_full bust  
3: body\_type\_hourglass  
4: body\_type\_pear  
5: body\_type\_petite  
6: body\_type\_straight & narrow

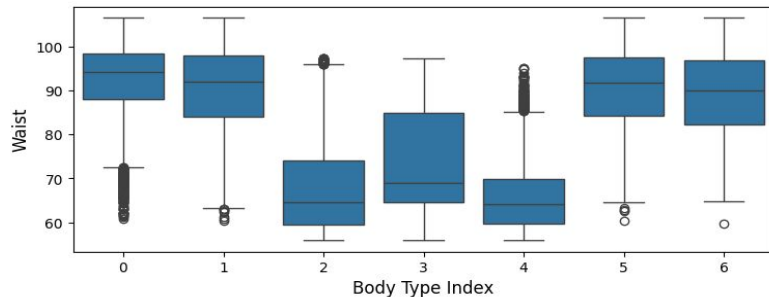
Gender Distribution by Body Type Index



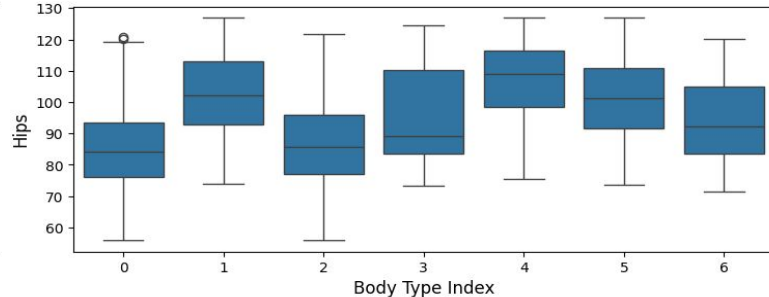
Weight Distribution by Body Type Index



Waist Distribution by Body Type Index

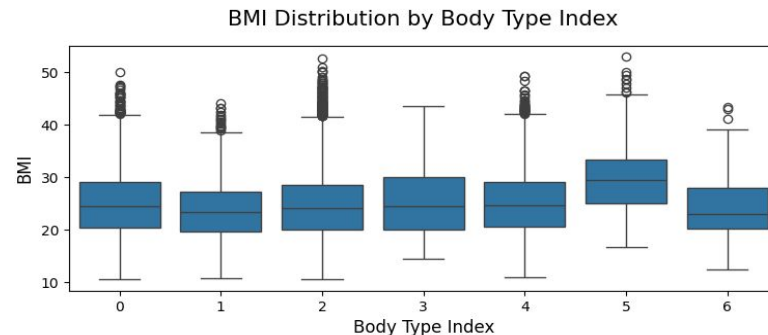
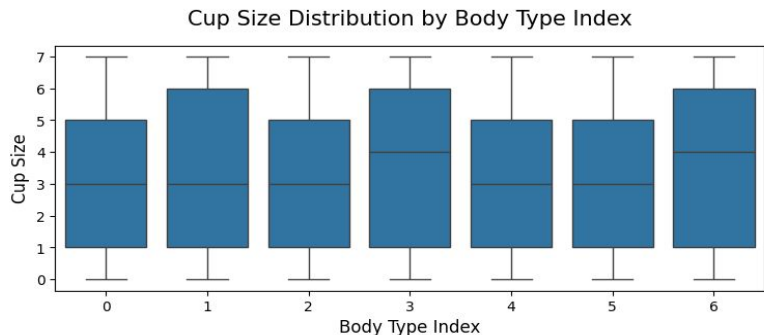
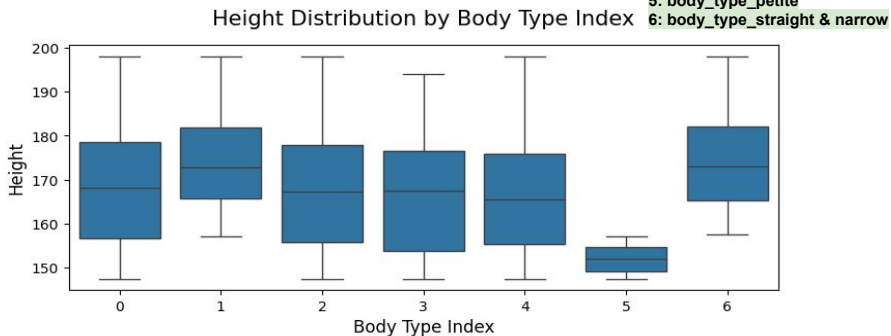
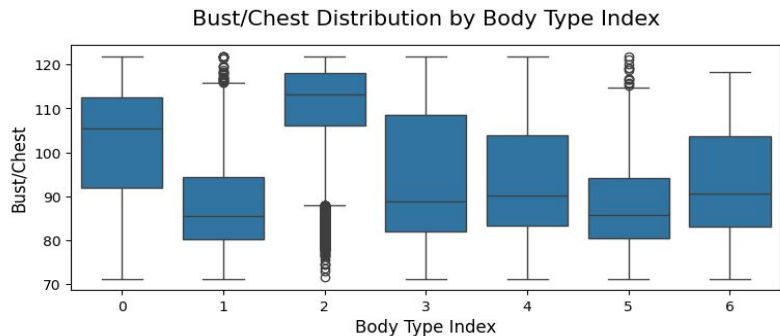


Hips Distribution by Body Type Index



## 신체 치수를 이용한 체형 분류 모델 (8) - 이상치 시각화

Body Type Index Mapping:  
0: body\_type\_apple  
1: body\_type\_athletic  
2: body\_type\_full bust  
3: body\_type\_hourglass  
4: body\_type\_pear  
5: body\_type\_petite  
6: body\_type\_straight & narrow





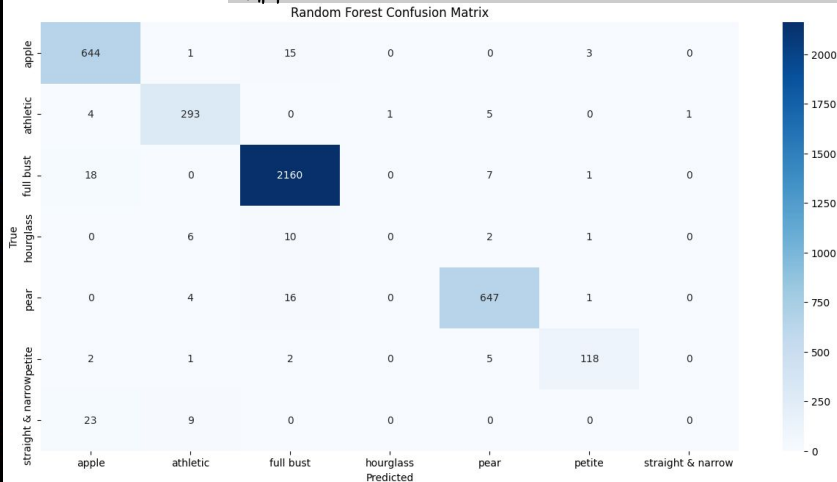
# 신체 치수를 이용한 체형 분류 모델 (9) - RandomForest 학습 결과

[결과]

|                             | precision<br>정밀도 | recall<br>재현율 | f1-score | support<br>테스트 데이터 개수 |
|-----------------------------|------------------|---------------|----------|-----------------------|
| body_type_apple             | 0.93             | 0.97          | 0.95     | 663                   |
| body_type_athletic          | 0.93             | 0.96          | 0.95     | 304                   |
| body_type_full bust         | 0.98             | 0.99          | 0.98     | 2186                  |
| body_type_hourglass         | 0.00             | 0.00          | 0.00     | 19                    |
| body_type_pear              | 0.97             | 0.97          | 0.97     | 668                   |
| body_type_petite            | 0.95             | 0.92          | 0.94     | 128                   |
| body_type_straight & narrow | 0.00             | 0.00          | 0.00     | 32                    |
| accuracy                    |                  |               | 0.97     | 4000                  |
| macro avg                   | 0.68             | 0.69          | 0.68     | 4000                  |
| weighted avg                | 0.95             | 0.97          | 0.96     | 4000                  |

[결과 시각화]

x축: 예측된 클래스  
y축: 실제 정답 클래스  
y, x 위치의 값: 실제 y클래스를 x클래스로 예측한 개수



결론: 데이터 imbalance 심함

→ 일부 소수 클래스들은 예측이 거의 안 됨. 즉, 샘플 수가 매우 적거나 구분이 어려운 케이스 존재  
But 다수 클래스를 중심으로는 예측 성능이 매우 좋음

# 신체 지수를 이용한 체형 분류 모델 (10) - K-Means 학습 결과

## [특성 조합]

RandomForest와 비교하기 위해  
모든 특성 사용

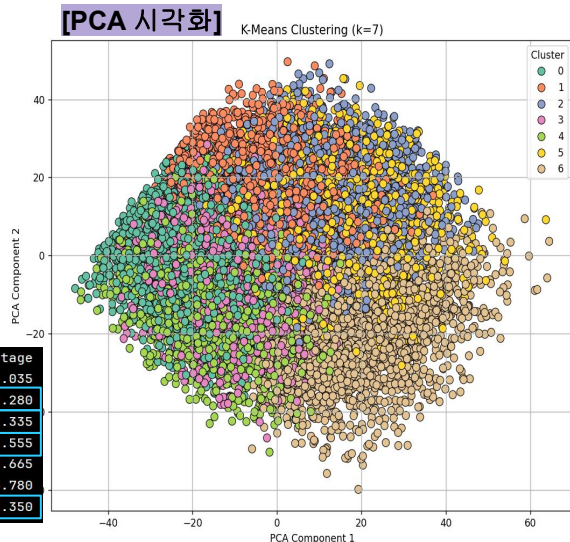
## [결과]

Silhouette Score: **0.1584**

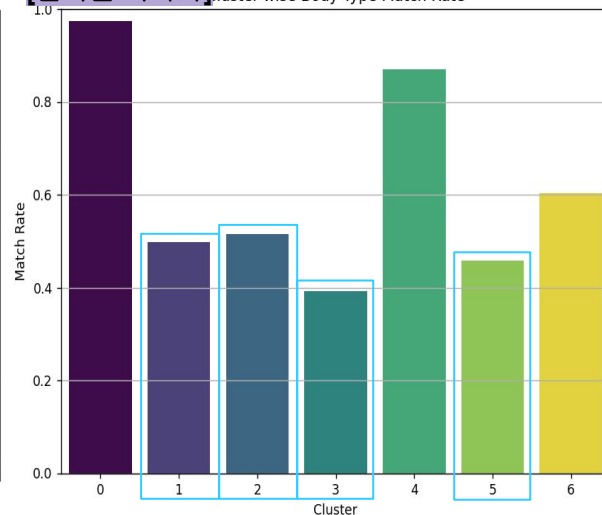
전체 데이터 기준 일치율 (Accuracy): **0.6262**

## [군집별 대표 체형과 개인 체형 일치율]

|   | representative_body_type | match_rate | cluster_size | cluster_percentage |
|---|--------------------------|------------|--------------|--------------------|
| 0 | body_type_full bust      | 0.974743   | 3207         | 16.035             |
| 1 | body_type_full bust      | 0.497709   | 3056         | 15.280             |
| 2 | body_type_apple          | 0.515813   | 3067         | 15.335             |
| 3 | body_type_pear           | 0.392305   | 2911         | 14.555             |
| 4 | body_type_full bust      | 0.870781   | 2933         | 14.665             |
| 5 | body_type_athletic       | 0.457792   | 2156         | 10.780             |
| 6 | body_type_full bust      | 0.603745   | 2670         | 13.350             |



## [일치율 시각화] Cluster-wise Body Type Match Rate



결론: 모든 특성을 RandomForest와 같이 구성할 경우, **RandomForest** 보다 성능이 떨어짐

1. **0.1584** → 군집이 어느 정도 형성되긴 했지만, **겹침이 꽤 많다**는 의미
2. **62.62% 일치** → 완벽하진 않지만, **절반 이상은 기존 바디타입 분류와 비슷하게 묶였다**는 의미

# 신체 치수를 이용한 체형 분류 모델 (11) - K-Means 학습 결과

## [특성 조합]

'BMI', 'Waist', 'Bust/Chest'

## [결과]

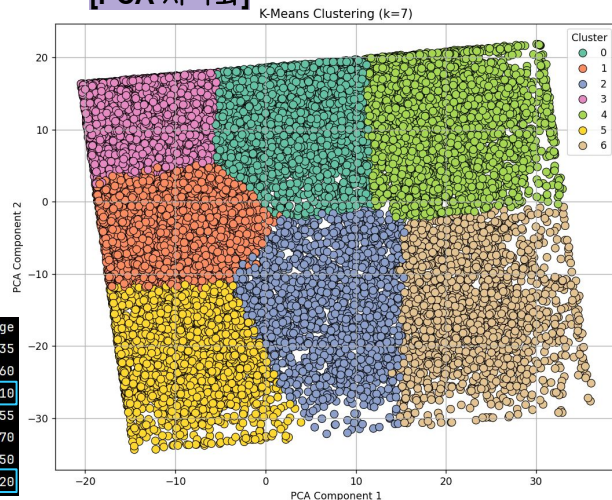
Silhouette Score: **0.2934**

전체 데이터 기준 일치율 (Accuracy): **0.6962**

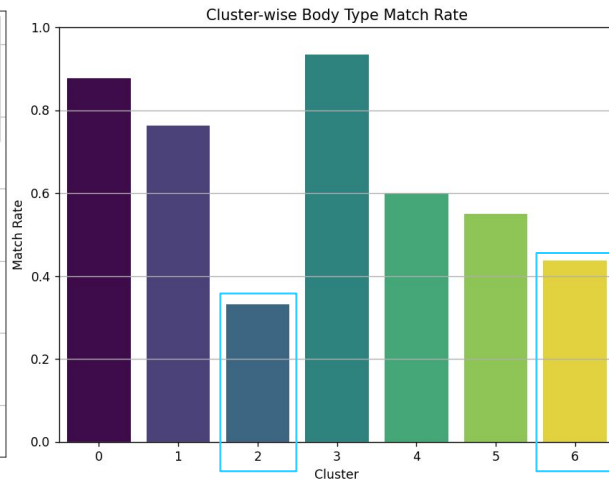
## [군집별 대표 체형과 개인 체형 일치율]

|   | representative_body_type | match_rate | cluster_size | cluster_percentage |
|---|--------------------------|------------|--------------|--------------------|
| 0 | body_type_full bust      | 0.876211   | 2787         | 13.935             |
| 1 | body_type_full bust      | 0.763610   | 2792         | 13.960             |
| 2 | body_type_apple          | 0.331528   | 1662         | 8.310              |
| 3 | body_type_full bust      | 0.933984   | 4711         | 23.555             |
| 4 | body_type_apple          | 0.600181   | 3314         | 16.570             |
| 5 | body_type_pear           | 0.550502   | 2990         | 14.950             |
| 6 | body_type_athletic       | 0.438073   | 1744         | 8.720              |

## [PCA 시각화]



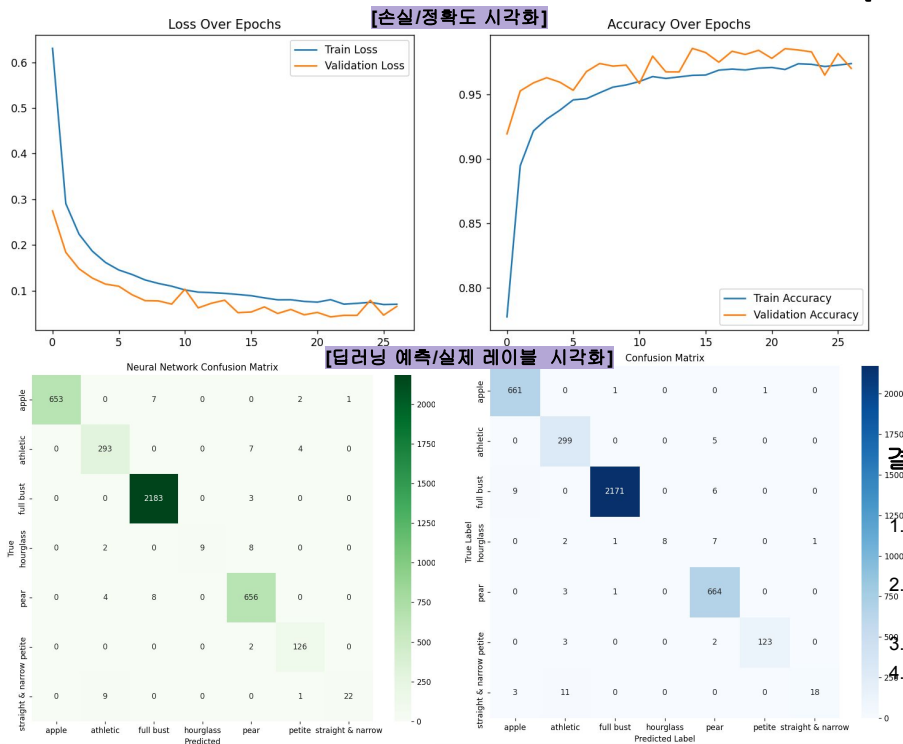
## [일치율 시각화]



결론: 주요 바디타입 패턴은 데이터 안에 분명히 존재함을 확인

1. **0.2934** → 군집들이 전보다 **덜 겹치고 더 명확하게 구분**
2. **69.62%** 일치 → 라벨 없이 군집을 나눴는데도 **약 70%**는 기존 **body type** 분류와 일치

# 신체 치수를 이용한 체형 분류 모델 (12) - MLP 학습 결과



정밀도, 재현율의 조화 평균

테스트 데이터 개수

| Classification Report: |           |        |          |         |
|------------------------|-----------|--------|----------|---------|
|                        | precision | recall | f1-score | support |
|                        | 정밀도       | 재현율    |          |         |
| 0                      | 1.00      | 0.98   | 0.99     | 663     |
| 1                      | 0.95      | 0.96   | 0.96     | 304     |
| 2                      | 0.99      | 1.00   | 1.00     | 2186    |
| 3                      | 1.00      | 0.47   | 0.64     | 19      |
| 4                      | 0.97      | 0.98   | 0.98     | 668     |
| 5                      | 0.95      | 0.98   | 0.97     | 128     |
| 6                      | 0.96      | 0.69   | 0.80     | 32      |
| accuracy               |           |        | 0.99     | 4000    |
| macro avg              | 0.97      | 0.87   | 0.90     | 4000    |
| weighted avg           | 0.99      | 0.99   | 0.98     | 4000    |

결론: 복잡한 패턴을 잘 학습했고, body type 분류 작업에 적합

Loss, Accuracy: 학습/검증 모두 부드럽게 감소,

accuracy 꾸준히 증가 → 오버피팅 없이 학습이 잘 된 것

Confusion Matrix: 거의 모든 클래스가 대각선에 몰려 있음

→ 모델이 정확하게 예측, 큰 클래스들은 잘 맞음, 소수 클래스(3, 6번)는 조금 어려움

99%에 가까운 높은 정확도와 F1-score를 기록

데이터가 적은 소수 클래스(hourglass, straight & narrow)에서는 recall이 상대적으로 낮음 → 데이터 증강이나 클래스 재분류 등의 보완이 필요

# 신체 치수를 이용한 체형 분류 모델 (13) - 사용한 알고리즘 특징 과정 / 결과 정리

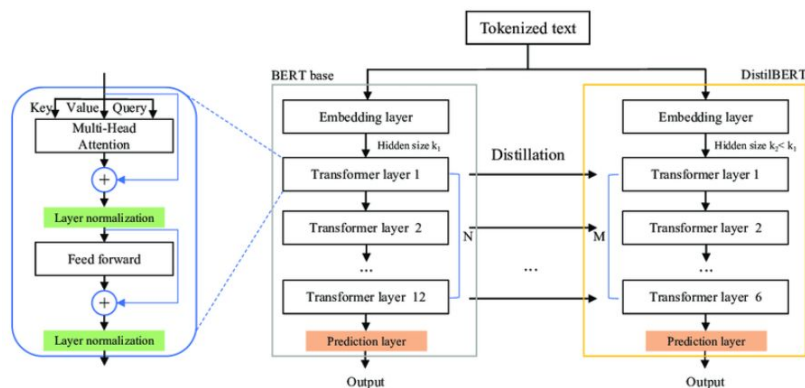
## [과정 비교]

| 항목        | 1번(Random Forest)                    | 2번(Neural Network)                   | 3번(K-Means Clustering)                |
|-----------|--------------------------------------|--------------------------------------|---------------------------------------|
| 기법        | 지도학습 (Random Forest)                 | 지도학습 (딥러닝 MLP)                       | 비지도학습 (K-Means)                       |
| 목표        | 체형(body type) 분류                     | 체형(body type) 분류                     | 체형 기반 군집화 및 대표 체형 찾기                  |
| 입력 데이터    | 전처리된 피쳐(X_train)                     | 전처리된 피쳐(X_train)                     | 원본 신체 치수(Bust, Waist, Hips, Height 등) |
| 타겟(label) | body_type (정답 존재)                    | body_type (정답 존재)                    | 없음 (label 없이 스스로 군집)                  |
| 모델 종류     | Random Forest Classifier             | Multi-layer Perceptron (Dense 층 3개)  | K-Means Clustering (n_clusters=5)     |
| 평가 방법     | Accuracy, F1 Score, Confusion Matrix | Accuracy, F1 Score, Confusion Matrix | 없음 (군집 해석)                            |
| 특징        | - 빠르고 튼튼<br>- 과적합 잘 버티               | - 복잡한 패턴 잘 학습<br>- 약간 오버핏 주의         | - 분류 없이 그룹화<br>- 대표 체형 분석 가능          |
| 시각화       | Confusion Matrix (Blues 컬러)          | Confusion Matrix + 학습 그래프 (loss/acc) | 없음(결과 출력)                             |
| 특이점       | 트리 기반 앙상블                            | 드롭아웃으로 과적합 방지                        | 군집별로 '대표 체형' 지정                       |

## [결과 비교]

| 모델            | 전체 정확도     | 소수 클래스 처리   | 특징                  | 주의점             |
|---------------|------------|-------------|---------------------|-----------------|
| Random Forest | ★ 97%      | ✗ 낮은 성능     | 이상치에 강함             | imbalance 보완 필요 |
| K-Means       | 👉 약 70% 일치 | -           | full bust 군집화 매우 우수 | 비지도, 라벨 없음      |
| MLP           | ☀ 99%      | ⚠ 일부 클래스 낮음 | 복잡한 관계 학습 가능        | 이상치 민감          |

# 긍정, 부정 후기 판단 모델 - [DistilBert]

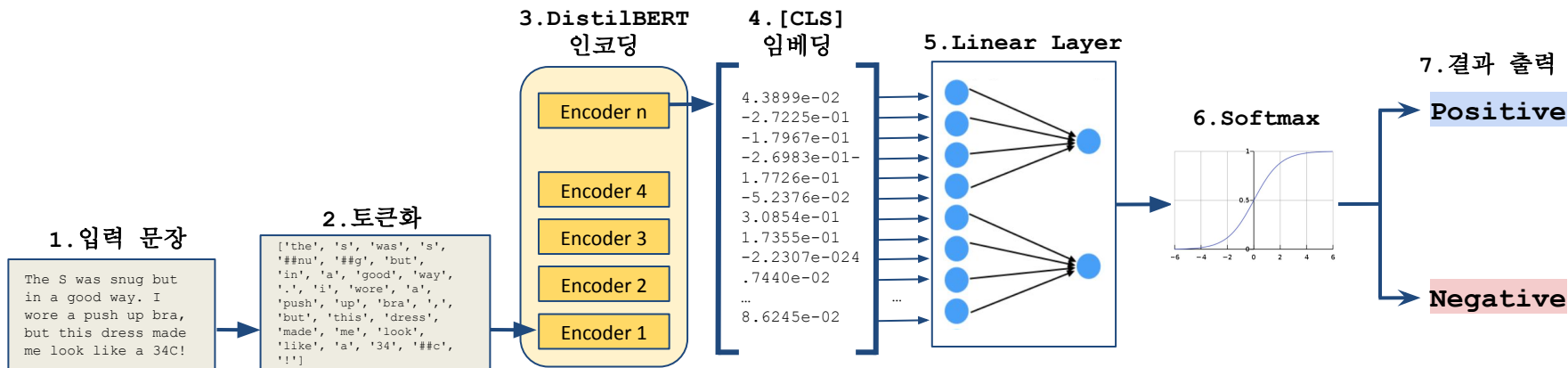


출처 : [https://www.researchgate.net/figure/The-DistilBERT-model-architecture-and-components\\_fig2\\_358239462/amp](https://www.researchgate.net/figure/The-DistilBERT-model-architecture-and-components_fig2_358239462/amp)

## 1. DistilBert

- 알고리즘 선택 이유: 사전학습 +  
경량화 + 실전 적용에 최적화된  
모델이기 때문에 리뷰 감성 분석에  
**빠르고 정확하게 적용**할 수 있어서  
선택

# 긍정, 부정 후기 판단 모델 - 흐름

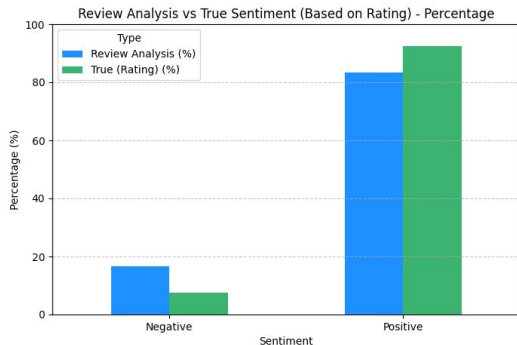


# 긍정, 부정 후기 판단 모델 - 학습 결과

## [결과]

별점과 감성 예측이 일치한 비율 (정확도): 0.8547

## [결과 시각화]



## 감성 분석 결과 요약 및 인사이트

- 총 20만 개의 리뷰 데이터 중, 리뷰에 포함된 문장의 감성과 사용자가 부여한 별점 사이의 일치 여부를 평가함.
- 분석 기준:
  - 긍정 감성 → 별점 8점 이상
  - 부정 감성 → 별점 8점 미만

## 인사이트

- 사용자의 리뷰 텍스트와 별점은 높은 상관관계를 보이며, 이를 기반으로 신뢰할 수 있는 상품 평가가 가능
- 별점 기준(예: 8점 이상)으로 상품을 추천하는 방식은 실제 사용자 만족도를 잘 반영하는 방법

## 분석 결과

- 감성 예측과 별점이 일치한 비율은 약 85.47%  
→ 전체 리뷰 중 약 17만 건이 리뷰 내용과 별점 부합

## 결론

- 리뷰 감성과 별점 간의 높은 일치도를 바탕으로, 8점 이상인 상품을 우선적으로 추천하는 전략이 타당하며, 이는 사용자 경험 향상 및 추천 시스템의 정확도를 높이는 데 기여



## 프로젝트 회고

A : 코드와 데이터로 옷장을 채웠지만, 그 옷장을 완성한 건 팀워크였다.

B : 기존에 개발한 쇼핑몰을 기준으로 실제로 서비스 가능한 기능을 AI로 구현하는 과정이 좋았다.

C : 멀게만 느껴졌던 AI를 실습을 통해 공부를 할 수 있어서 좋은 경험이 되었다.

D : 어려웠던 AI 공부를 같이 하다 보니까 좀 더 쉽게 할 수 있어서 많은 도움이 되었다.

E : 잊고 있었던 머신러닝 지식을 다시 되돌아볼 수 있어서 좋았고,  
AI를 활용해야 한다면 딥하게 배우는 시간이 더 길었으면 하는 생각이 있습니다.

## 프로젝트 회고 - 추후 개발사항

- 현재 모델은 사용자 특성(체형, 나이 등)에 기반
  - > 구매 이력, 스타일 선호도 등 다양한 데이터 추가 수집
  - > 더 정교하고 개인화된 추천 시스템 구축
- **Fast API** 연동 서버 도입 후 기존 서비스 추가
  - 사용자 페이지:
    - **상품 추천** 기능
    - 바디 수치 입력 후 **바디 타입 판단** 기능
  - 신뢰도 낮은 리뷰를 **필터링**하거나, **리뷰 품질을 관리**할 수 있는 기능
- **RAG** 활용 **LLM 챗봇** 서비스 기능 추가

## 참고 문헌

1) [계층적 군집분석을 활용한 아시아 컨테이너 항만 클러스터링 측정] (2021)

([http://kportea.or.kr/filedown/Treatise/2022/3.%20%5B%ED%95%AD%EB%A7%8C%EA%B2%BD%EC%A0%9C%2037%EA%B6%8C%201%ED%98%B8%5D%20%EA%B3%84%EC%B8%B5%EC%A0%81%20%EA%B5%B0%EC%A7%91%EB%B6%84%EC%84%9D\(%EC%B5%9C%EB%8B%A8.%20%EC%B5%9C%EC%9E%A5,%20%ED%8F%89%EA%B7%A0,%20%EC%A4%91%EC%95%99%EC%97%B0%EA%B2%B0\)%EB%B0%A9%EB%B2%95%EC%97%90%20%EC%9D%98%ED%95%9C%20%EC%95%84%EC%8B%9C%EC%95%84%20%EC%BB%A8%ED%85%8C%EC%9D%B4%EB%84%88%20%ED%95%AD%EB%A7%8C%EC%9D%98%20%ED%81%B4%EB%9F%AC%EC%8A%A4%ED%84%B0%EB%A7%81%20%EC%B8%A1%EC%A0%95%20%EB%B0%8F%20%EC%8B%A4\\_.pdf](http://kportea.or.kr/filedown/Treatise/2022/3.%20%5B%ED%95%AD%EB%A7%8C%EA%B2%BD%EC%A0%9C%2037%EA%B6%8C%201%ED%98%B8%5D%20%EA%B3%84%EC%B8%B5%EC%A0%81%20%EA%B5%B0%EC%A7%91%EB%B6%84%EC%84%9D(%EC%B5%9C%EB%8B%A8.%20%EC%B5%9C%EC%9E%A5,%20%ED%8F%89%EA%B7%A0,%20%EC%A4%91%EC%95%99%EC%97%B0%EA%B2%B0)%EB%B0%A9%EB%B2%95%EC%97%90%20%EC%9D%98%ED%95%9C%20%EC%95%84%EC%8B%9C%EC%95%84%20%EC%BB%A8%ED%85%8C%EC%9D%B4%EB%84%88%20%ED%95%AD%EB%A7%8C%EC%9D%98%20%ED%81%B4%EB%9F%AC%EC%8A%A4%ED%84%B0%EB%A7%81%20%EC%B8%A1%EC%A0%95%20%EB%B0%8F%20%EC%8B%A4_.pdf))

2) [Body shape classification using measurements] (Springer, 2018)

([https://link.springer.com/chapter/10.1007/978-3-319-77700-9\\_8?utm\\_source=chatgpt.com](https://link.springer.com/chapter/10.1007/978-3-319-77700-9_8?utm_source=chatgpt.com))

[Human body shape prediction and classification for mass customization] (Elsevier, 2019)

([https://biomedicaloptics.spiedigitallibrary.org/conference-proceedings-of-spie/4309/0000/3D-measurement-of-human-body-for-apparel-mass-customization/10.1117/12.410883.short?utm\\_source=chatgpt.com](https://biomedicaloptics.spiedigitallibrary.org/conference-proceedings-of-spie/4309/0000/3D-measurement-of-human-body-for-apparel-mass-customization/10.1117/12.410883.short?utm_source=chatgpt.com))

# Q&A