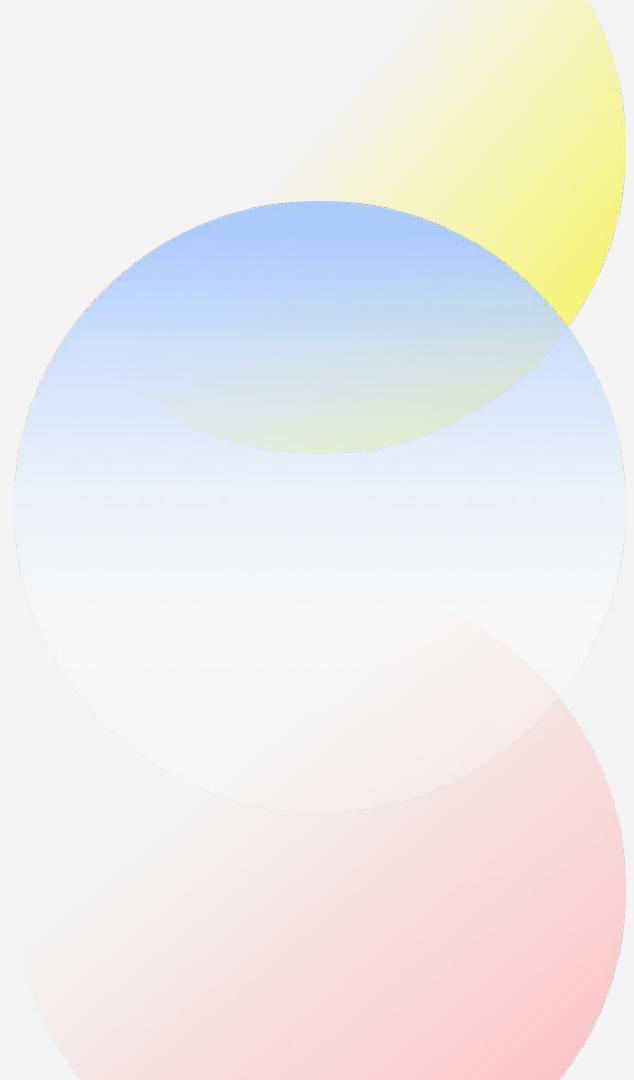


# Discovering sub-genres in music



# Table of Contents

01

Defining the  
Problem

02

Dataset &  
Preprocessing

03

EDA & Feature  
Selection

04

Clustering

05

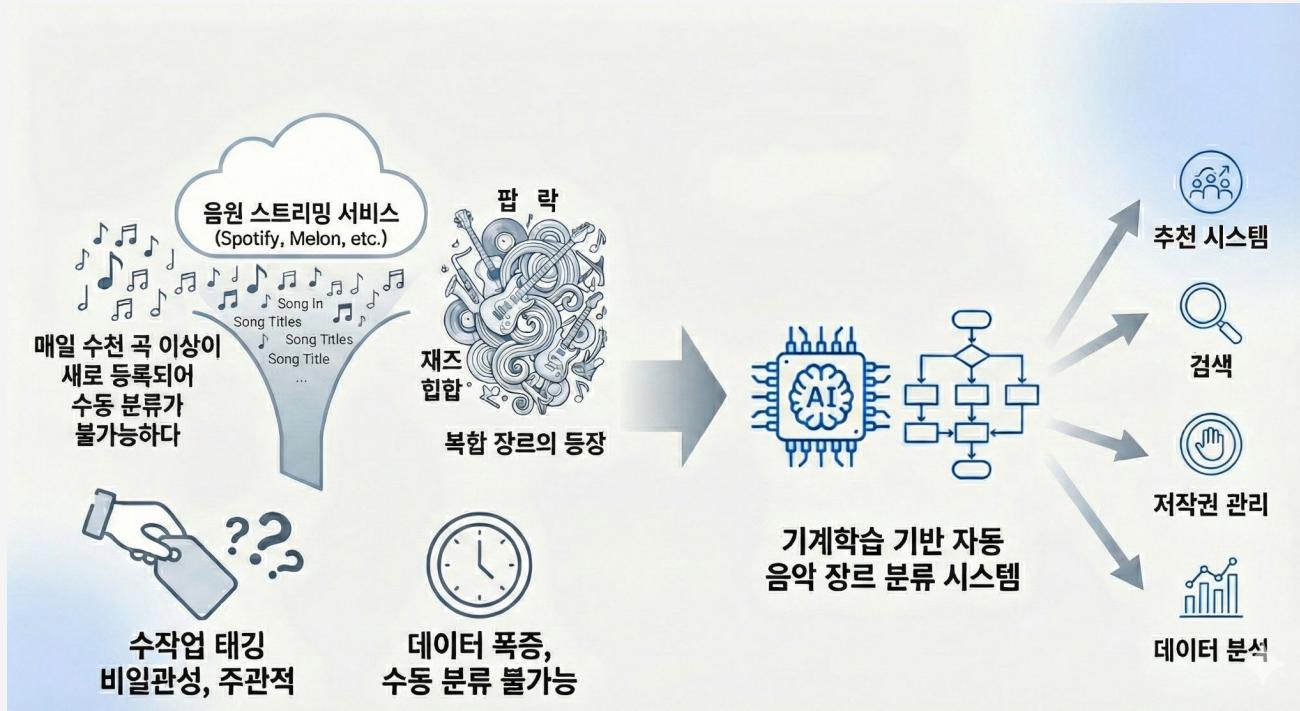
Classifying

06

Evaluation &  
Limitation

# 01 Defining the problem

목표: clustering을 이용해서 음악적 패턴 기반 세부 장르 분류 후 supervised learning으로 정확도 확인



with...

Clustering

Classifying

# 02 Dataset & Preprocessing

사용 데이터셋: **GTZAN music genre dataset**

- 10개 장르(Classical, Pop, Rock, Metal, Jazz 등) × 각 100곡
- 각 곡 길이 약 **30초**, 총 **1,000곡**

세그먼트 단위로 확장

- 30초 오디오를 **3초, 30초** 단위로 잘라 세그먼트 생성
- 약 **10,000개** 세그먼트를 학습에 활용

CSV 기반 전처리 데이터 사용

- 원본 오디오가 아니라 미리 추출된 **feature CSV** 사용
- 각 장르별 80개 곡은 **train**, 20개 곡은 **test**로 사용

# 02 Dataset & Preprocessing

## 장르 다양성

- Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock

서로 다른 성격의 10개 장르 포함

## 현실적인 난이도

- 장르 간 경계가 명확하지 않고, **복합 장르·유사 장르**가 많음
- t-SNE 시각화에서도 일부 장르는 서로 섞여 있는 **분포**를 보임

## 연구에 많이 쓰이는 벤치마크

- 음악 장르 분류 연구에서 가장 많이 사용되는 공개 데이터셋 중 하나
- baseline·여러 모델을 비교하기 용이

## Sub-genre / Clustering 실험에 적합

- 기존 라벨의 한계를 보완하기 위해, **feature 기반 세부 장르(sub-genre)**를 만들고 실험하기 좋음

ANDRADA · UPDATED 6 YEARS AGO  
910 · Code · Download · ...  
**GTZAN Dataset - Music Genre Classification**  
Audio Files | Mel Spectrograms | CSV with extracted features  
[GTZAN Dataset - Music Genre Classification](#)

# 02 Dataset & Preprocessing

## 데이터 로드

- Csv 파일 불러오기
- 각 row = 3초 세그먼트 1개, label = 원래 GTZAN 장르

## Feature 선택

- Energy, Spectral, Rhythm, MFCC 1~13의 mean/var (총 26개), Chroma/Harmony 을 중심으로 진행

## 정규화 (Scaling)

- StandardScaler로 평균 0, 분산 1로 스케일링
- Train 데이터로 fit, Test에는 transform만 적용

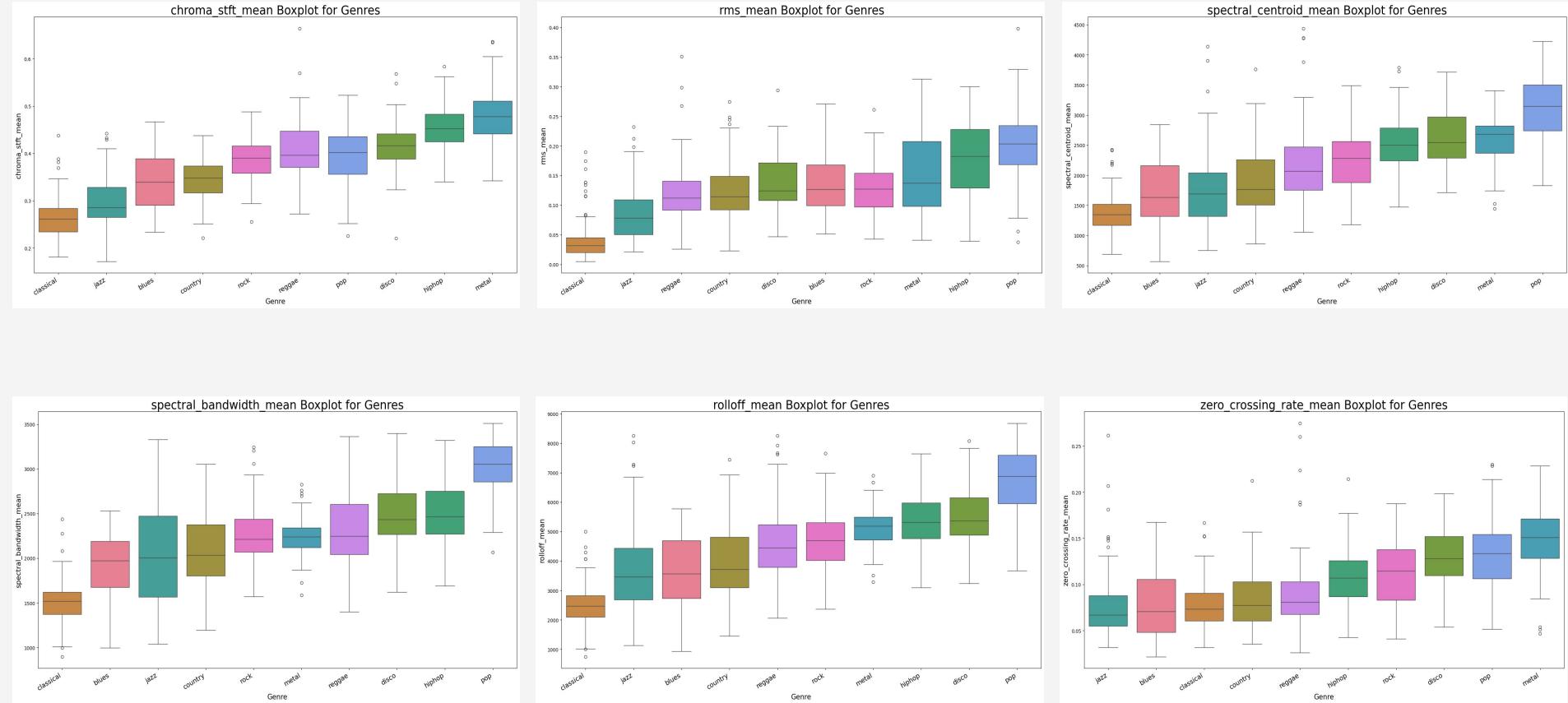
# 03 EDA & Feature Selection

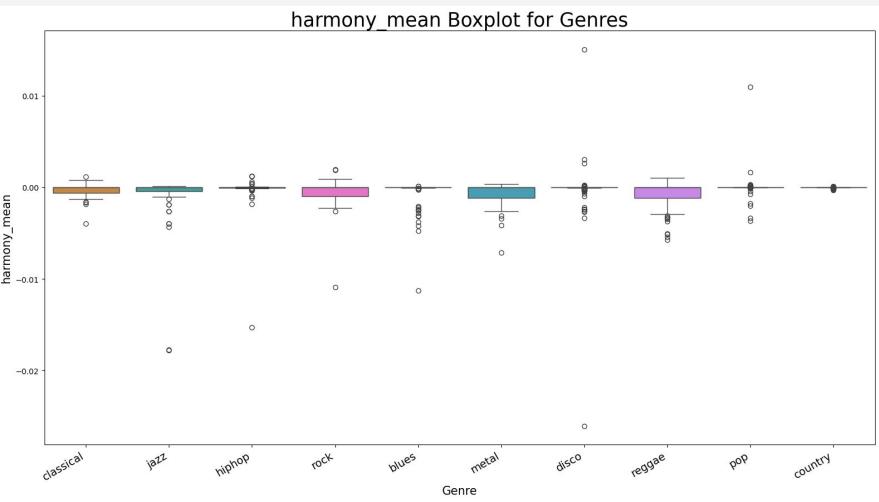
## Feature 선택

- **Energy:** rms\_mean, rms\_var
- **Spectral:** spectral\_centroid\_\*, spectral\_bandwidth\_\*, rolloff\_\*
- **Rhythm:** zero\_crossing\_rate\_\*, tempo
- **MFCC 1~13의 mean/var** (총 26개)
- **Chroma/Harmony:** chroma\_stft\_\*, harmony\_\* (분석용, 중요도 낮음)

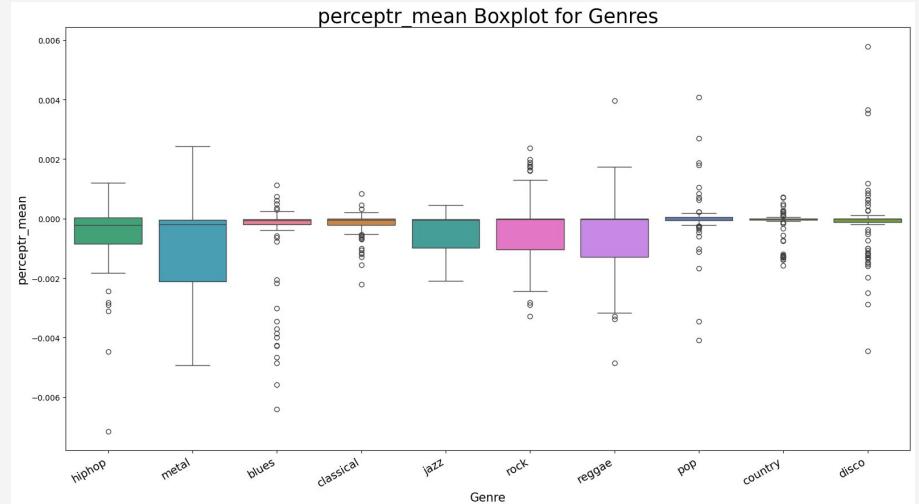
## 사용한 분석 방법(EDA)

- **Boxplot:** 주요 feature 6개에 대한 장르별 분포 비교
- **t-SNE:** 고차원 feature를 2D로 축소하여 장르 간 유사도·군집 구조 확인

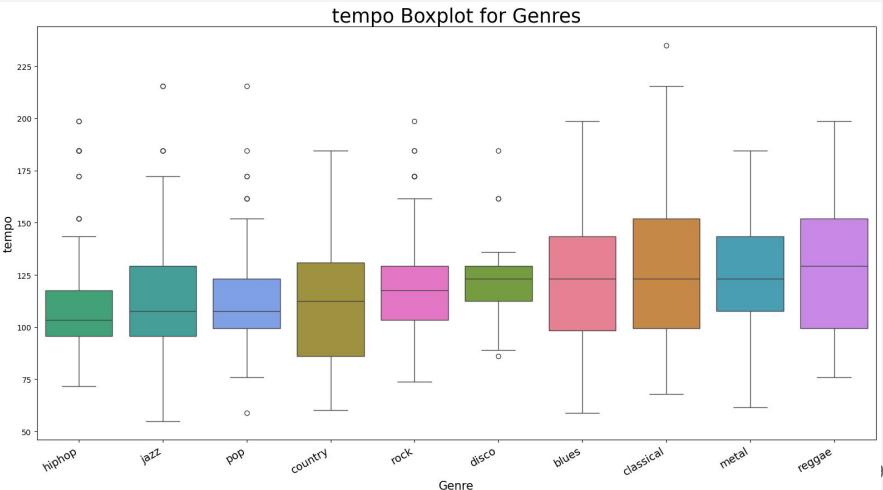




perceptr\_mean Boxplot for Genres

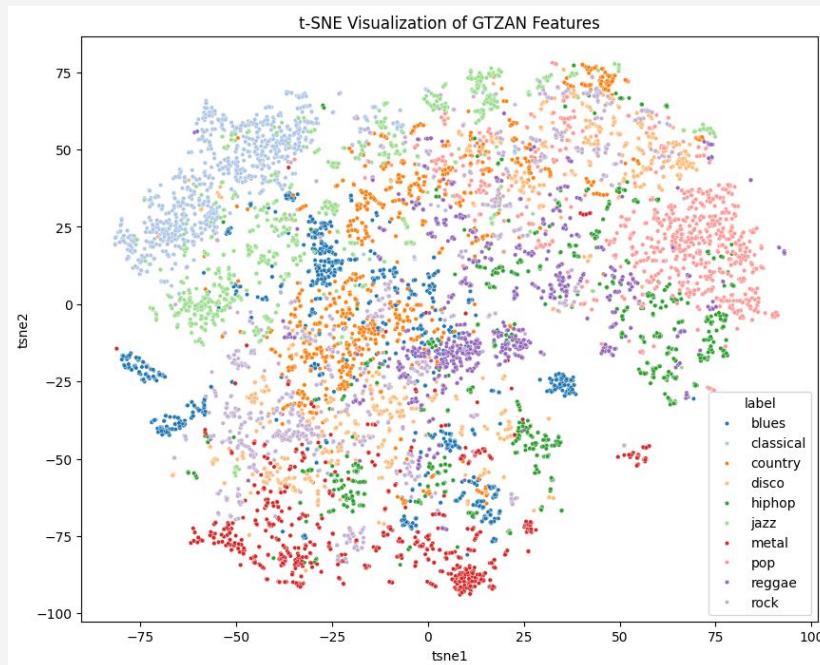


tempo Boxplot for Genres



# 03 EDA & Feature Selection

## T-SNE Visualization

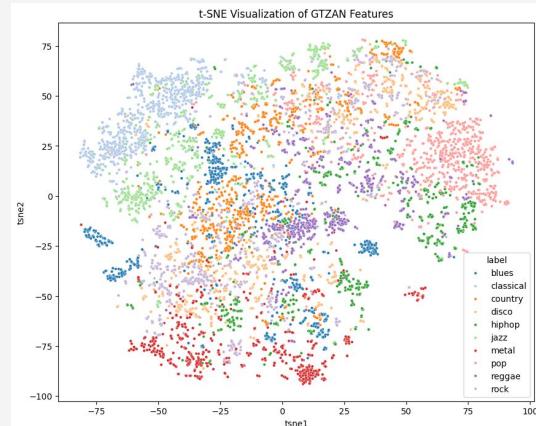


# 04 Clustering

Clustering 사용 이유

→ GTZAN에서 지정된 label 결과 정확도가 낮음

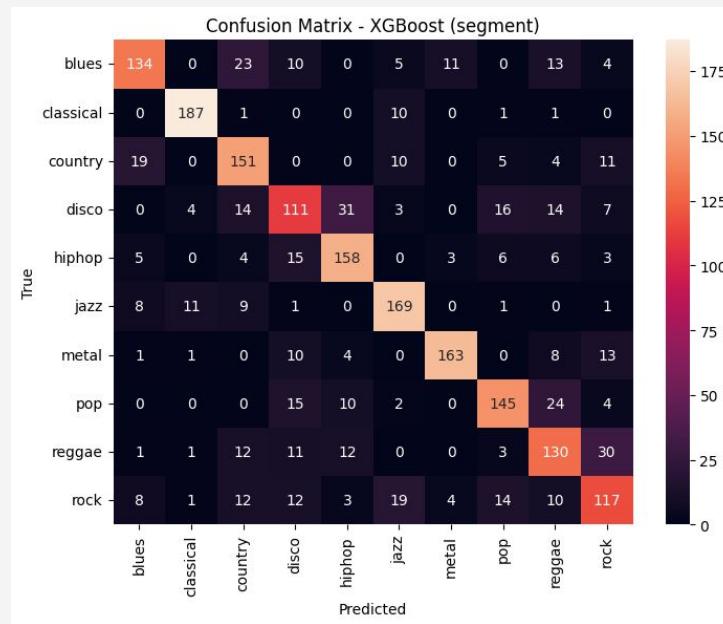
Model		Accuracy	Macro_F1	Weighted_F1
0	KNN	0.6545	0.653938	0.653938
1	Decision Tree	0.4950	0.495362	0.495362
2	Random Forest	0.6685	0.662763	0.662763
3	AdaBoost	0.4965	0.485082	0.485082
4	XGBoost	0.7325	0.731784	0.731784



# 04 Clustering

Clustering 사용 이유

→ GTZAN에서 지정된 label 결과 정확도가 낮음



# 04 Clustering

## Clustering 사용 이유

- GTZAN에서 지정된 label 결과 정확도가 낮음
- t-sne 결과에서 장르 경계가 명확하지 않음
- 음악 장르는 중에서도 비슷하거나 겹치는 요소가 많음
- hip-hop 요소가 있는 pop,
- Clustering을 사용하여 음악 features 기반 더 세분화된 sub-genre 식별
- Clustering 적용 후 supervised learning

# 04 Clustering

비슷한 방법론: clustering & supervised learning

Article Full-text available

K-Means-Based Pseudo-Labeling Technique in Supervised Learning Models for Regional Classification Based on Types of Non-Communicable Diseases

November 2025 · Jurnal Online Informatika · 10(2):465-473

DOI: [10.15575/join.v10i2.1609](https://doi.org/10.15575/join.v10i2.1609)

License: CC BY-ND 4.0

 Herison Surbakti ·  Tb Ai Munandar

Automated night-time fog detection and masking using machine learning from near real-time satellite observations

Narendra Nelli <sup>a</sup>, Diana Francis <sup>a</sup>  , Charfeddine Cherif <sup>a</sup>, Ricardo Fonseca <sup>a</sup>, Hosni Ghedira <sup>b</sup>

[1] Surbakti, Herison & Munandar, Tb. (2025). K-Means-Based Pseudo-Labeling Technique in Supervised Learning Models for Regional Classification Based on Types of Non-Communicable Diseases. *Jurnal Online Informatika*. 10. 465-473. [10.15575/join.v10i2.1609](https://doi.org/10.15575/join.v10i2.1609). Narendra Nelli, Diana Francis, Charfeddine Cherif, Ricardo Fonseca, Hosni Ghedira,

[2] Automated night-time fog detection and masking using machine learning from near real-time satellite observations, *Science of Remote Sensing*, Volume 12, 2025, 100297, ISSN 2666-0172, <https://doi.org/10.1016/j.srs.2025.100297> (<https://www.sciencedirect.com/science/article/pii/S2666017225001038>)

# 04 Clustering

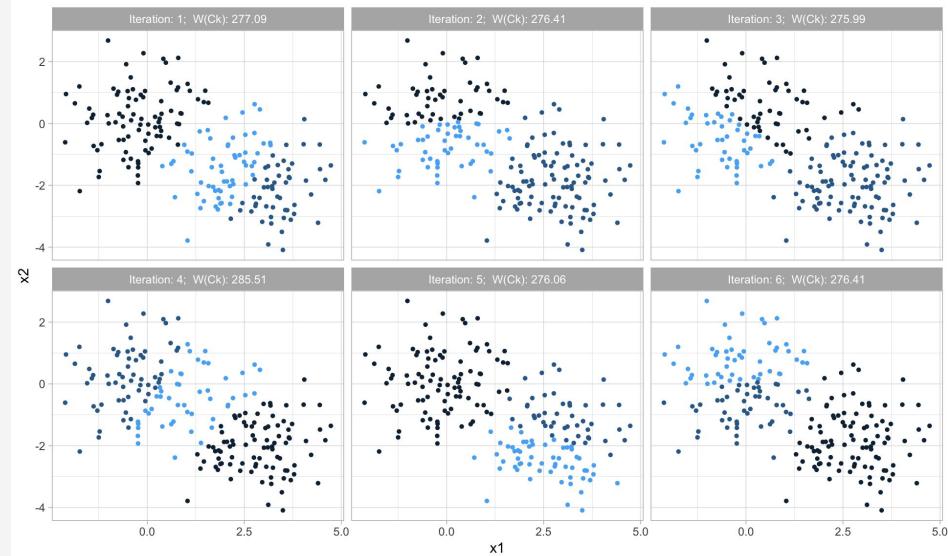
## K-means:

→ 같은 특징을 가진 데이터끼리 모음

(예: 밝은 소리, 어두운 소리 등을 묶음)

→ 군집 중심(centroid)을 기준으로 feature space를 분할: MFCC, Spectral Centroid / Rolloff, RMS (loudness), Spectral Bandwidth 등의 feature를 사용해 음악의 음색, 밝기, 에너지 등 핵심적 acoustic feature로 군집 형성

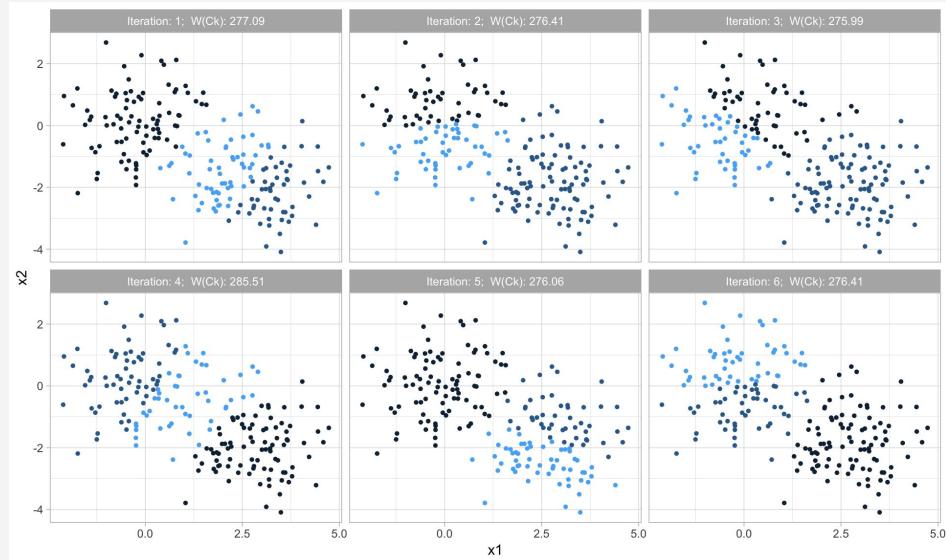
→ 사운드가 비슷하게 들리는 음악을 알아서 한 그룹으로 묶음



# 04 Clustering

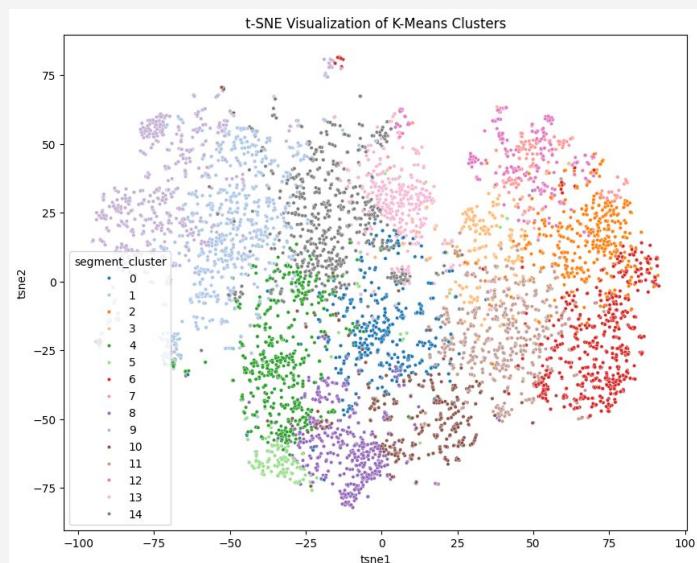
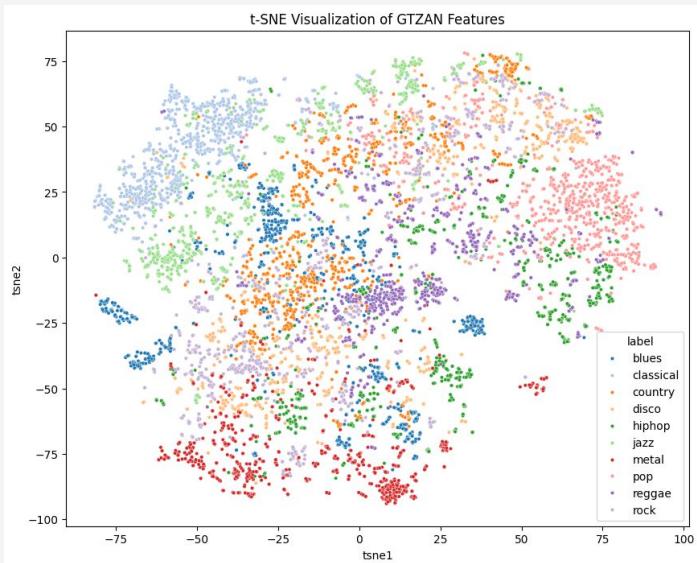
K-means:

- classifier(XGBoost, KNN 등)의 타깃 라벨로 사용 가능
- cluster 별 feature 중요도 분석도 쉬움
- 각 cluster 가 어떤 음향적 특성을 가지는지도 분석 가능



# 04 Clustering

K-means result:



# 04 Clustering

K-means label names:

```
--- Subcluster: classical_jazz_0 ---
Total segments: 937
Genre distribution:
  classical  :  696  ( 74.3%)
  jazz       :  148  ( 15.8%)
  country    :   38  (  4.1%)
  blues      :   33  (  3.5%)
  rock       :    5  (  0.5%)
  reggae     :    5  (  0.5%)
  disco      :    5  (  0.5%)
  pop        :    4  (  0.4%)
  hiphop     :    2  (  0.2%)
  metal      :    1  (  0.1%)
```



“Orchestral jazz”

# 04 Clustering

K-means label names:

```
--- Subcluster: pop_hiphop_reggae_6 ---
Total segments: 578
Genre distribution:
pop      : 313  ( 54.2%)
hiphop   : 156  ( 27.0%)
reggae   :  88  ( 15.2%)
disco    :  14  (  2.4%)
jazz     :   4  (  0.7%)
country  :   2  (  0.3%)
rock    :   1  (  0.2%)
```



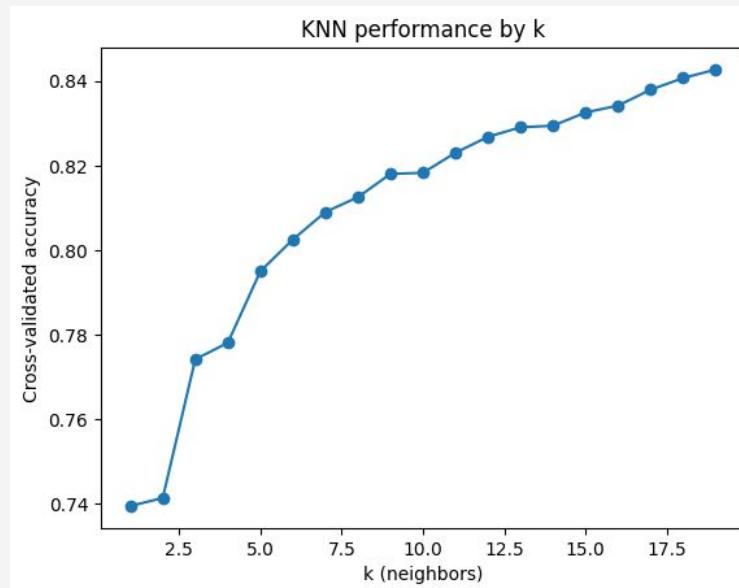
“Dance pop”

# 05 Classifying

- KNN → 거리 기반
- Decision Tree → 규칙 기반
- Random Forest → 여러 트리의 앙상블
- Ada Boost → 약한 모델을 반복적으로 보완
- XGboost → 복잡한 비선형 경계까지 학습 가능

# 05 Classifying

- KNN parameter



# 06 Evaluation & Limitation

Baseline (Accuracy)  
GTZAN label 사용

모델	곡 단위 정확도
KNN	0.598
Decision Tree	0.430
Random Forest	0.445
AdaBoost	0.400

GTZAN label 사용 + Random-split 사용  
+ fine-tuning

	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.6545	0.653938	0.653938
1	Decision Tree	0.5135	0.514497	0.514497
2	Random Forest	0.6685	0.662763	0.662763
3	AdaBoost	0.4965	0.485082	0.485082
4	XGBoost	0.7325	0.731784	0.731784

# 06 Evaluation & Limitation

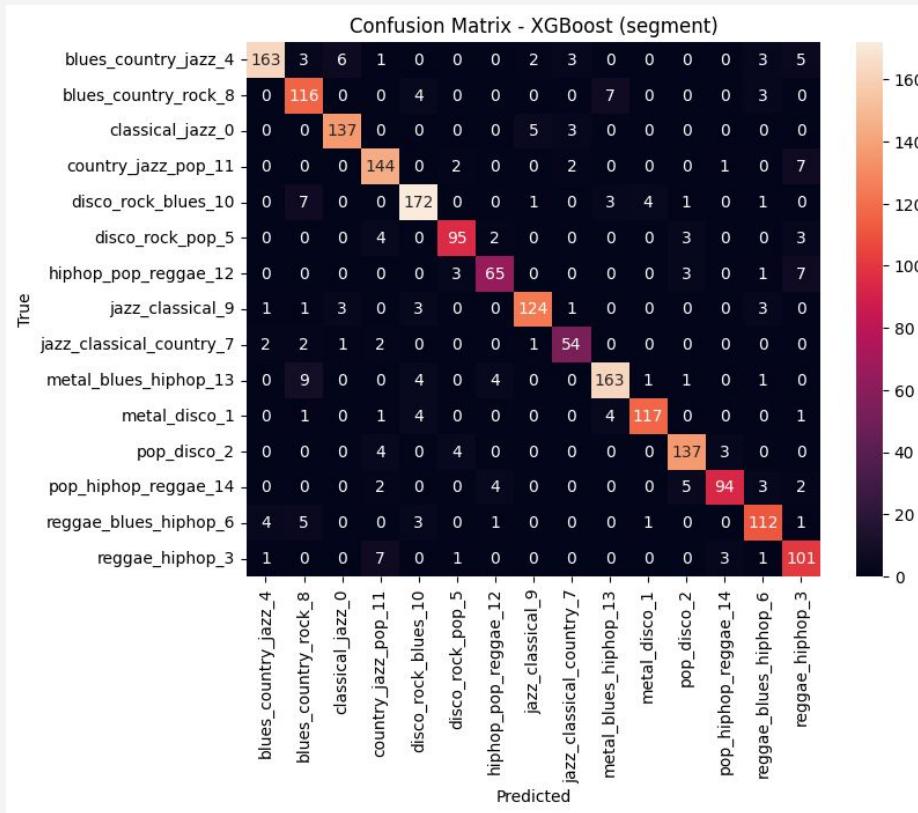
GTZAN label 사용 + Random-split  
사용 + fine-tuning

	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.6545	0.653938	0.653938
1	Decision Tree	0.5135	0.514497	0.514497
2	Random Forest	0.6685	0.662763	0.662763
3	AdaBoost	0.4965	0.485082	0.485082
4	XGBoost	0.7325	0.731784	0.731784

K-means cluster label 사용

	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.8530	0.843185	0.852399
1	Decision Tree	0.6965	0.684308	0.696345
2	Random Forest	0.8450	0.839672	0.845583
3	AdaBoost	0.6145	0.584505	0.590645
4	XGBoost	0.8970	0.892590	0.897424

# 06 Evaluation & Limitation



# 06 Evaluation & Limitation

리듬 기반 feature 추가  
X

\O	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.843333	0.819334	0.840384
1	Decision Tree	0.666667	0.632425	0.669686
2	Random Forest	0.848333	0.830849	0.845854
3	AdaBoost	0.633333	0.570372	0.605595
4	XGBoost	0.841667	0.816814	0.840503

리듬 기반 feature 추가  
O

\O X	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.843333	0.819334	0.840384
1	Decision Tree	0.663333	0.640817	0.664954
2	Random Forest	0.848333	0.830849	0.845854
3	AdaBoost	0.633333	0.570372	0.605595
4	XGBoost	0.841667	0.816814	0.840503

# 06 Evaluation & Limitation

Segment 길이 = 3

3	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.8530	0.843185	0.852399
1	Decision Tree	0.6900	0.677335	0.690478
2	Random Forest	0.8450	0.839672	0.845583
3	AdaBoost	0.6145	0.584505	0.590645
4	XGBoost	0.8880	0.881691	0.888296

Segment 길이 = 10

10	Model	Accuracy	Macro_F1	Weighted_F1
0	KNN	0.843333	0.819334	0.840384
1	Decision Tree	0.666667	0.632425	0.669686
2	Random Forest	0.848333	0.830849	0.845854
3	AdaBoost	0.633333	0.570372	0.605595
4	XGBoost	0.841667	0.816814	0.840503

# 06 Evaluation & Limitation

-데이터의 구조적 한계

Parameter Tuning 전

```
===== [Segment-level] KNN =====  
Accuracy      : 0.8755  
Macro F1       : 0.8594  
Weighted F1    : 0.8743
```



Parameter Tuning 후

```
==== Segment-level metrics (KNN) ====  
Accuracy : 0.8455  
Macro F1 : 0.8340046615290707  
Weighted F1 : 0.8447141717068677
```

# 06 Evaluation & Limitation

-데이터의 구조적 한계

## K-means limitation

원형 / 구형에서만 성능이 뛰어남.  
→ centroid 문제가 있어 음악 장르의  
복잡함을 나타내는 데 한계가 있음.



## Gaussian mixture model

타원형 cluster 가능/soft clustering  
→ 음악 장르의 혼합적, 복합적 특성  
표현 가능

# Future plans & Improvements

01

Cluster 개수 조정

silhouette + DB +  
stability(ARI) + 해석  
가능성으로 cluster 개수  
선택

02

GMM 도입

하드 라벨 + soft 확률로  
장르 혼합 정도까지 표현  
확률 벡터를 classifier 입력  
feature로 추가

03

Semi-Supervised  
Learning

원래 GTZAN 장르 +  
sub-genre pseudo-label을  
함께 활용

self-training, consistency  
등을 이용해 라벨 품질 개선



# Thank you

