



DỰ ÁN AI-DS

# **HỆ THỐNG GỢI Ý BÀI HÁT DỰA TRÊN THÓI QUEN CỦA NGƯỜI DÙNG BẰNG PHƯƠNG PHÁP K-MEANS**

Present by Group 3 - AI1707

# Nhóm 3



Lê Đăng Hùng

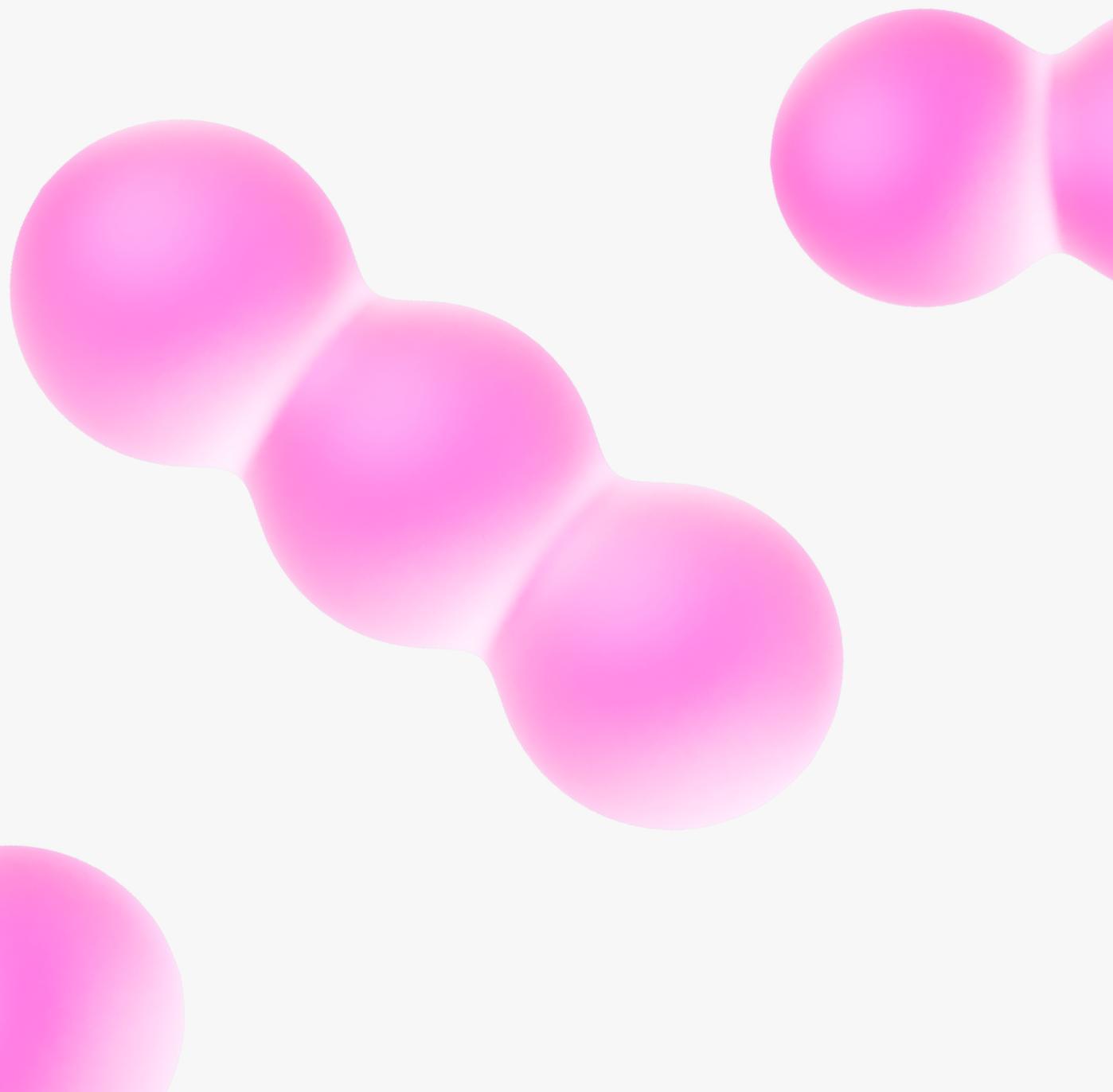


Lê Hồng Quang



Trần Tuấn Anh

# Outline



- Giới thiệu
- Chuẩn bị dữ liệu
- Phương pháp
- Kết quả
- Kết luận và thảo luận



GROUP 3 -AI1707

# 1. Giới thiệu

# 1. Giới thiệu



## Hệ thống gợi ý dựa trên nội dung (content-based)

- Cần nhiều thời gian hơn xử lý
- Gợi ý những bài hát không cùng sở thích

## Hệ thống khuyến nghị hợp tác (collaborative)

- Không thể đề xuất tất cả các bài hát

K-Means Clustering

# Làm thế nào ?



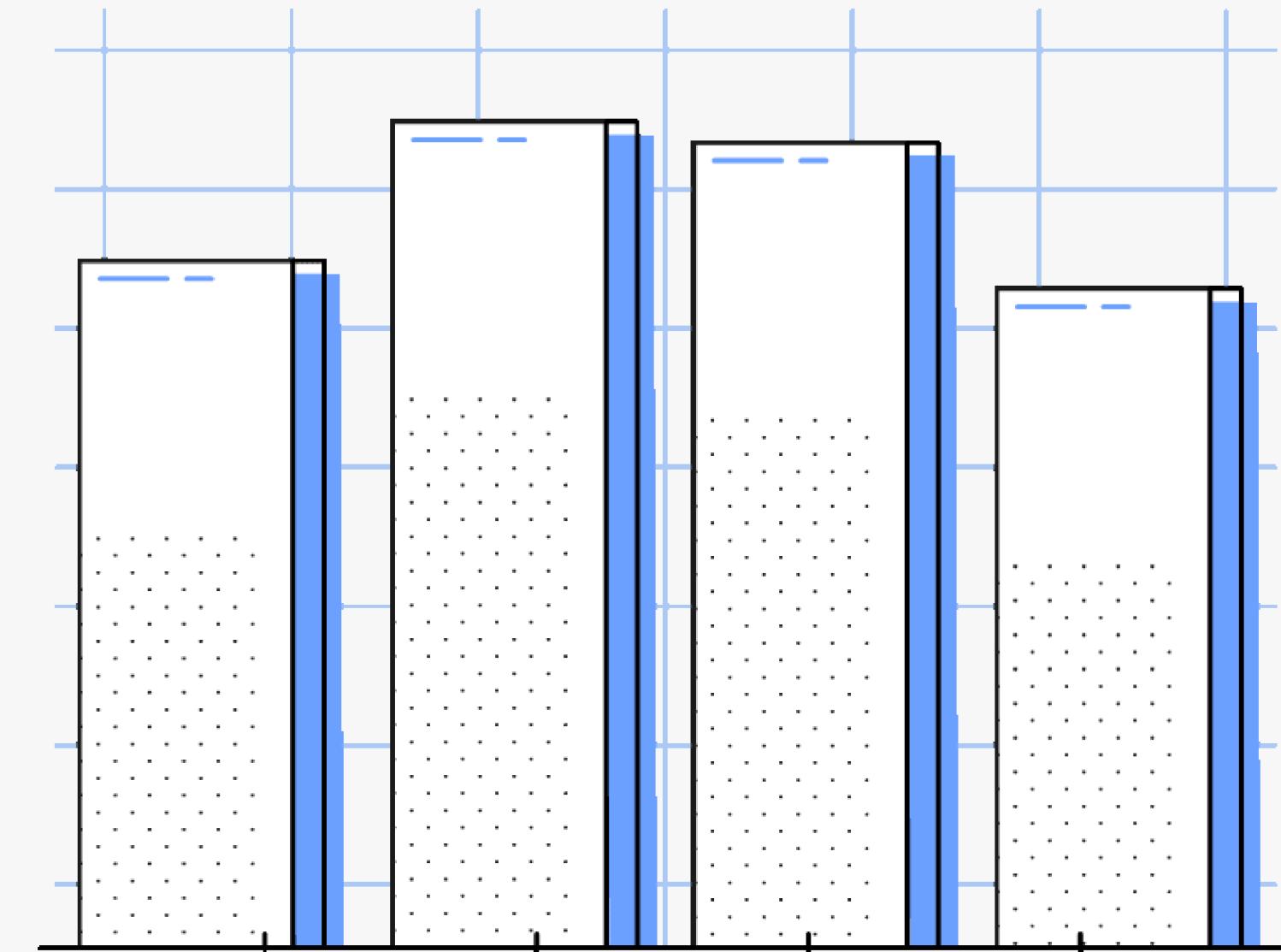
Phân tích thuộc  
tính

Phân tích sở thích  
của người dùng

Chọn bản nhạc phù  
hợp



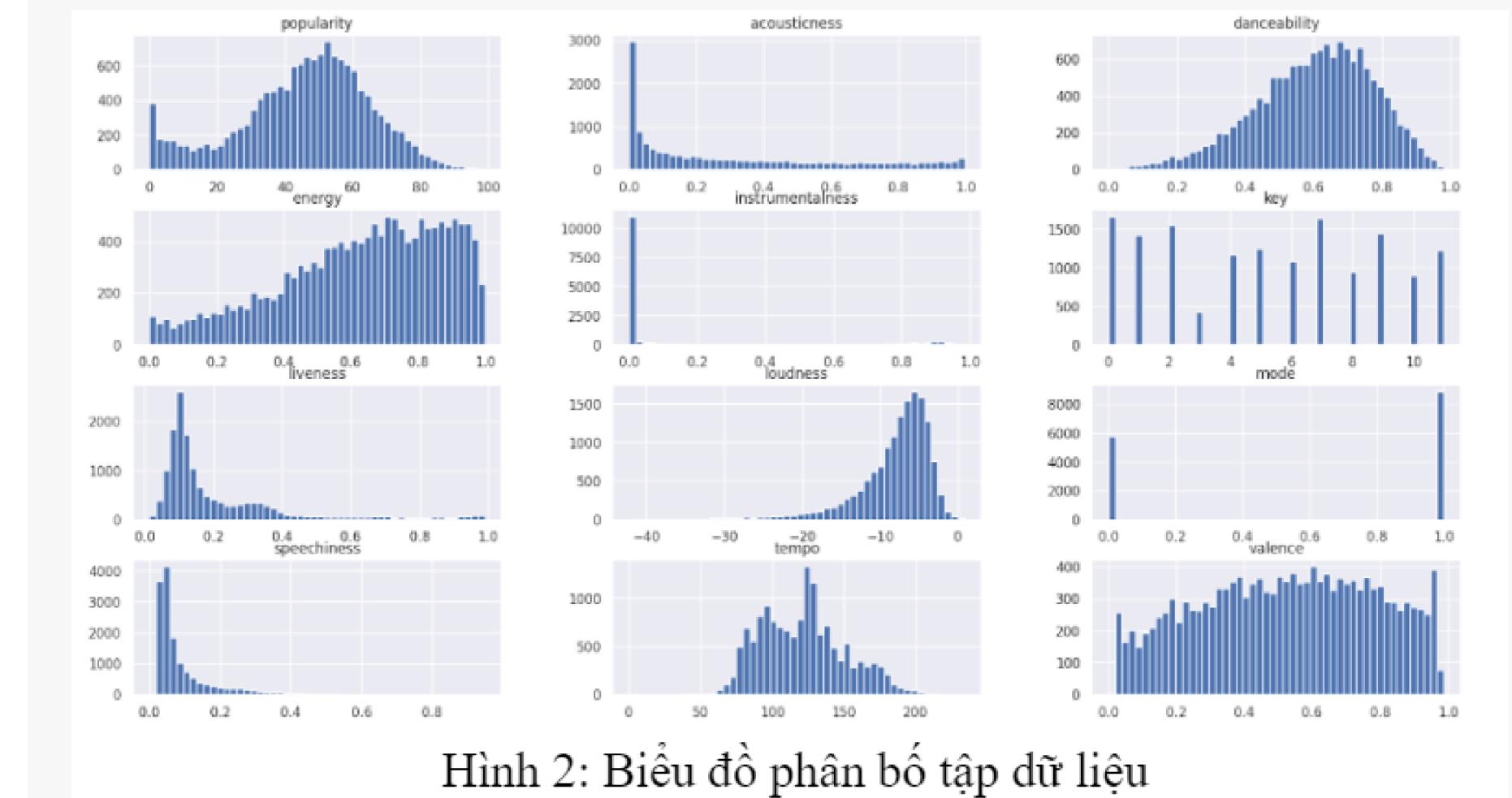
## 2. Chuẩn bị dữ liệu



## 2. Chuẩn bị dữ liệu

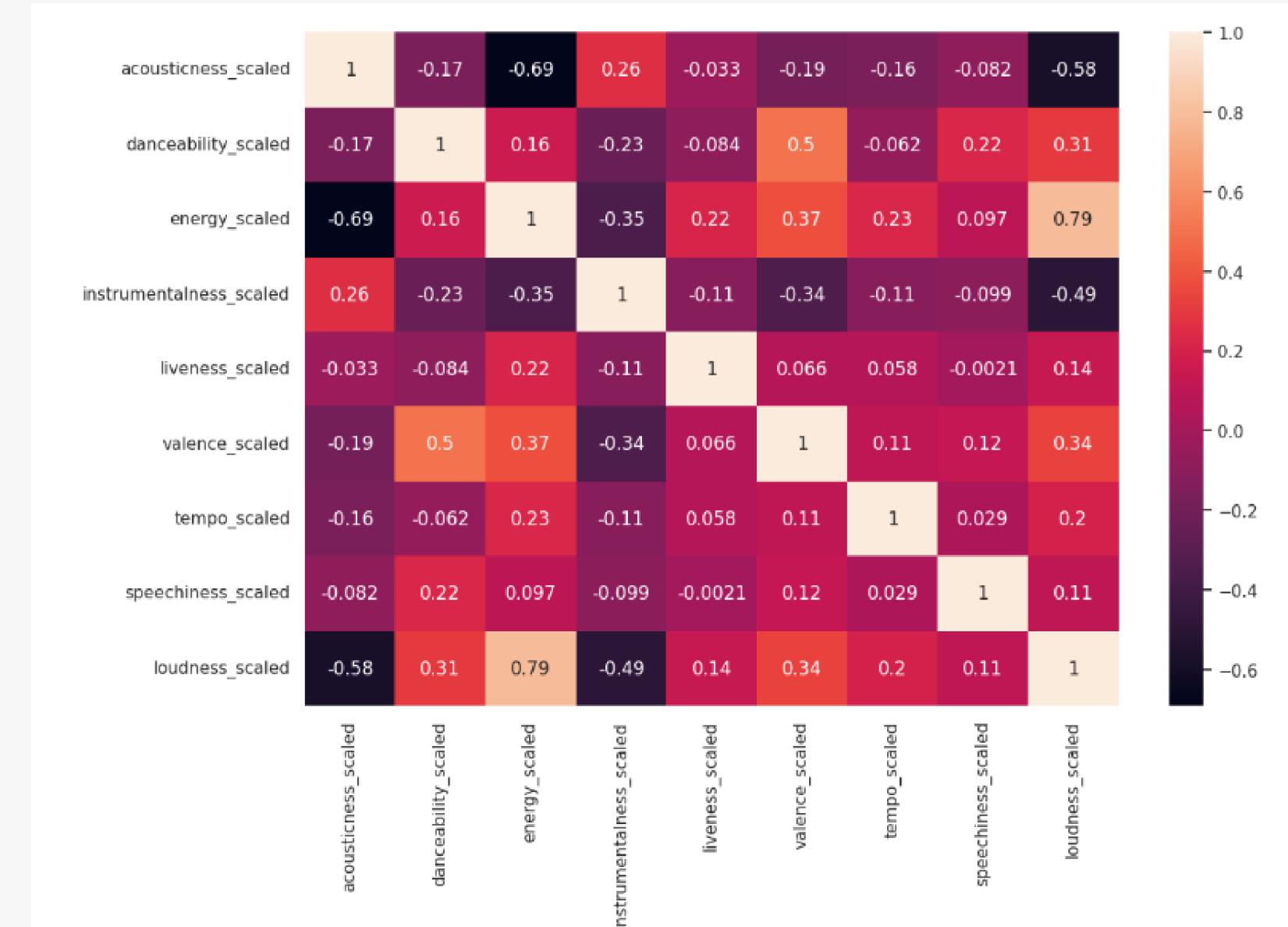
- Spotify API
- Thư viện bài hát: 14675 hàng và 19 cột
- Lịch sử nghe của người dùng: 85 bài hát

# 2. Chuẩn bị dữ liệu



Hình 1: Biểu đồ mối tương quan

# Chuẩn hóa dữ liệu



Hình 4: Biểu đồ mối tương quan dữ liệu được chuẩn hóa



GROUP 3 - AI1707

# 3.Phương pháp



**Huấn luyện  
mô hình**

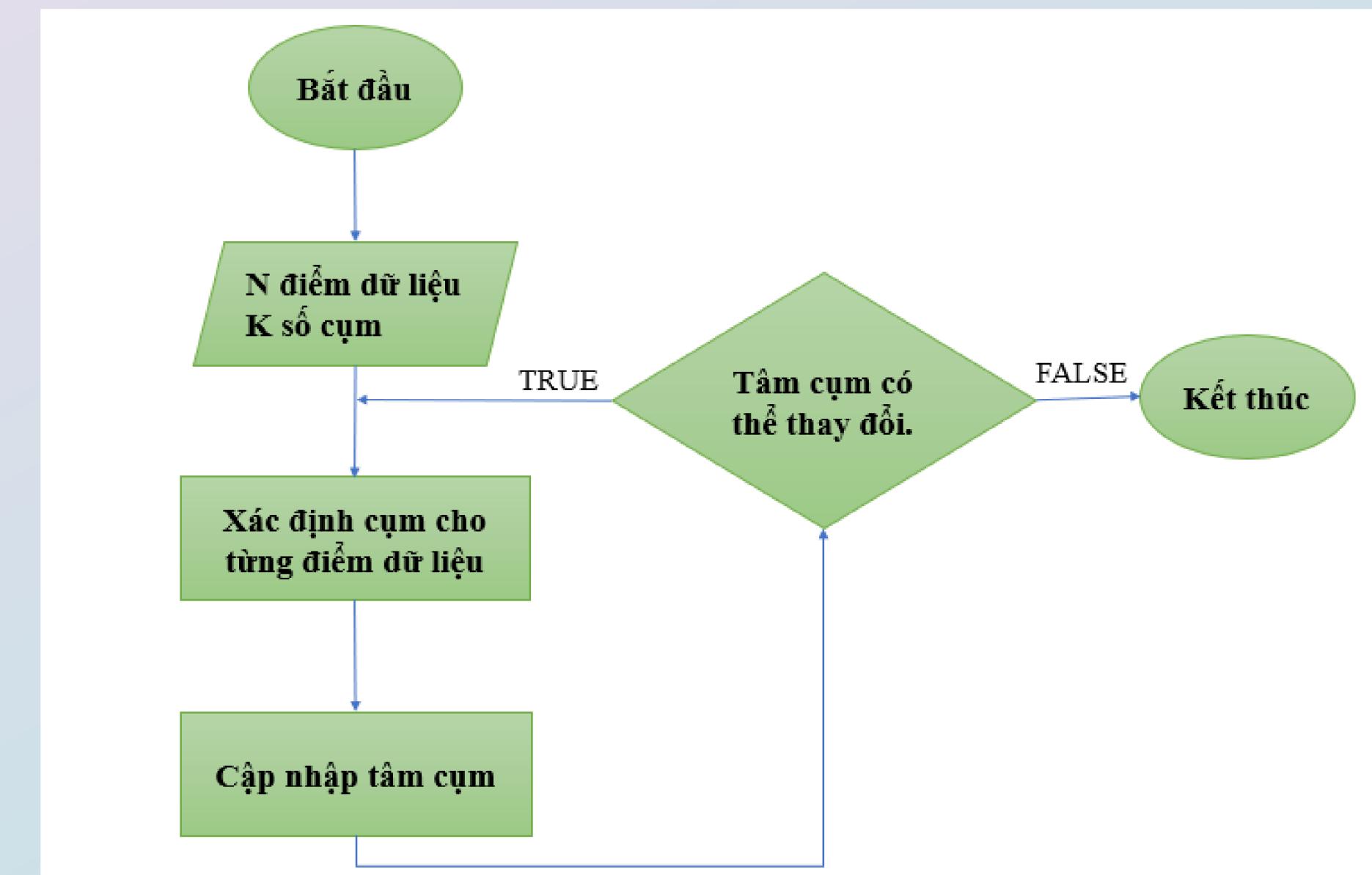
Thuật toán K-Means

**Trực quan hóa**

Phép phân tích thành  
phần chính

### 3.1. Huấn luyện mô hình bằng thuật toán K-means

## K-Means là gì ?



# Tìm số cụm tối ưu



Phương pháp Elbow



Giá trị Silhouette

### 3.1.1. Giá trị Silhouette

- Công thức tính:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

### 3.1.2. Phương pháp Elbow

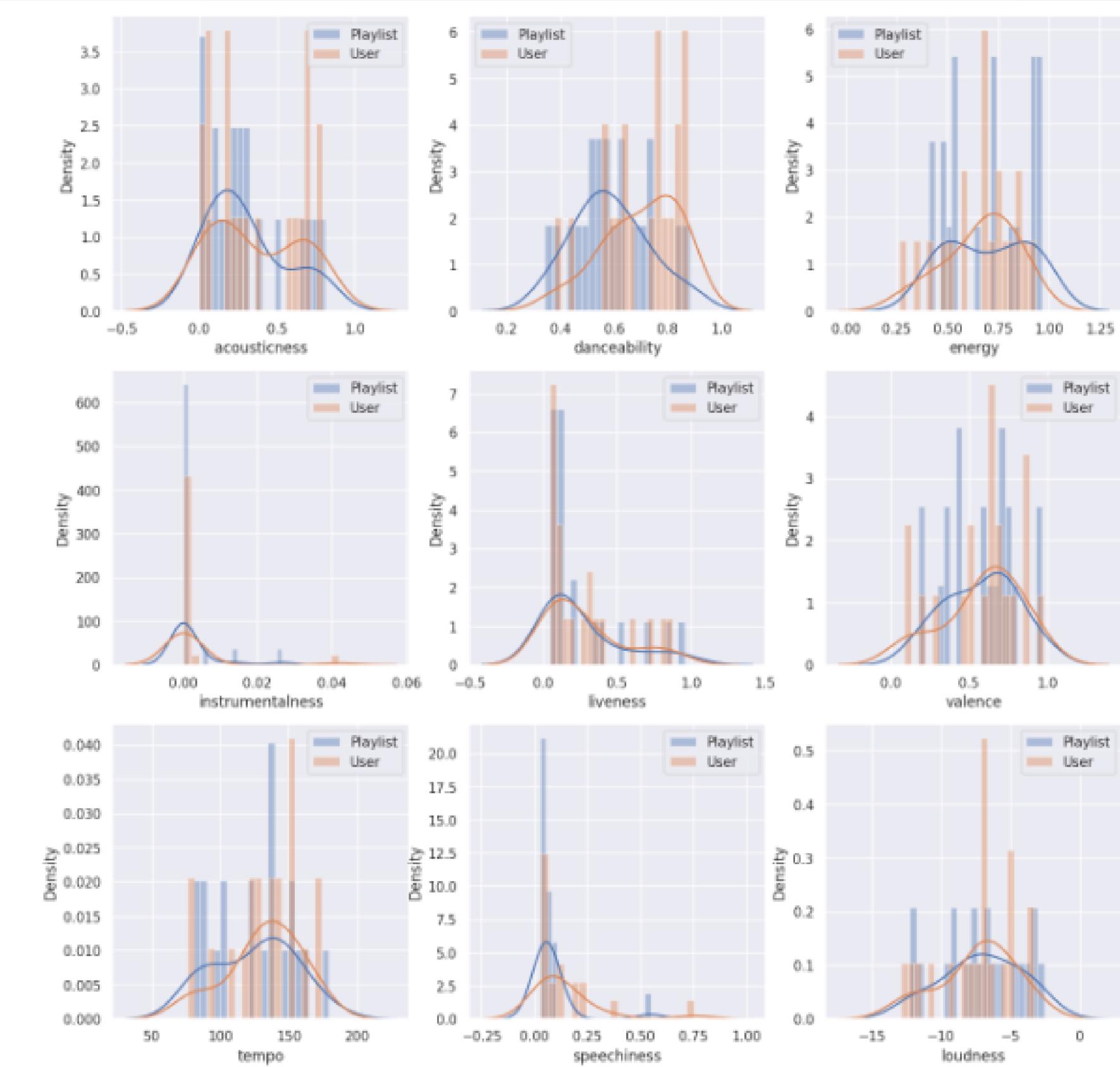
- Công thức tính SSD:

$$\text{SSD} = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

## 3.2. Trực quan hóa bằng PCA

- Đánh giá độ chuẩn xác của thuật toán đã phân cụm
- Tốc độ xử lý, đơn giản hóa dữ liệu và tiền xử lý dữ liệu tốt
- Mất thông tin về cụm ban đầu

# 4. Kết quả



Hình 7: Biểu đồ sự tương quan của danh sách bài hát gợi ý so với danh sách bài hát của người dùng

## **5. KẾT LUẬN VÀ THẢO LUẬN**