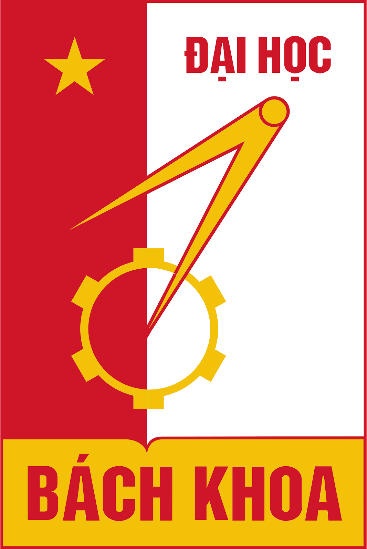
ĐẠI HỌC BÁCH KHOA HÀ NỘI

TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

──────── \* ───────



**BÁO CÁO NGHIÊN CỨU TỐT NGHIỆP 2**

**ĐỀ TÀI: Music Recommendation System**

**Giảng viên hướng dẫn : PGS Thân Quang Khoát**

**Sinh viên : Nguyễn Doãn Tuấn MSSV: 20205040**

Hà nội , tháng 1 năm 2024

# **LỜI NÓI ĐẦU**

# Hệ thống đề xuất âm nhạc (Music Recommendation System) là một bài toán quan trọng trong đời sống hàng ngày . Việc đề xuất âm nhạc sao chuẩn xác sẽ giúp tăng trải nghiệm nghe nhạc của người nghe. Hiện nay với sự phát triển vượt bậc của các mô hình học máy (Machine Learning) và học sâu (Deep learning) cùng với lượng dữ liệu lớn thì ta có thể đề xuất các thể loại bài hát phù hợp với người nghe .Để giải quyết bài toán này chúng tôi quyết định sử dụng phương pháp ***Content-Based Recommendation*** (Gợi ý dựa trên nội dung) trong hệ gợi ý để đề xuất âm nhạc dựa trên danh sách nhạc có sẵn.

# Hi vọng với những gì trình bày trong báo cáo sẽ giúp mọi người có thể có thêm nhiều góc nhìn về hệ gợi ý để giải quyết các bài toán trong thực tế.

# **Mục lục**

[**LỜI NÓI ĐẦU 2**](#_heading=h.gjdgxs)

[**Mục lục 3**](#_heading=h.30j0zll)

[**MỞ ĐẦU 4**](#_heading=h.3znysh7)

[1. Giới thiệu 4](#_heading=h.2et92p0)

[2. Mục tiêu đề tài 4](#_heading=h.tyjcwt)

[**Chương 1: NGHIÊN CỨU TỔNG QUAN. 5**](#_heading=h.3dy6vkm)

[1. Tổng quan về Hệ gợi ý (Recommendation system) 5](#_heading=h.1t3h5sf)

[2. Các phương pháp gợi ý](#_heading=h.4d34og8)  [6](#_heading=h.2s8eyo1)

[**Chương 2: CƠ SỞ LÝ THUYẾT 10**](#_heading=h.17dp8vu)

[1. Thuật toán phân cụm K-Means 10](#_heading=h.3rdcrjn)

[2. Hàm khoảng cách cosine (Cosine Similarity) 12](#_heading=h.26in1rg)

[**Chương 3: XÂY DỰNG HỆ THỐNG 13**](#_heading=h.z337ya)

[1. Các bước chính xây dựng hệ thống 13](#_heading=h.3j2qqm3)

[2. Đọc dữ liệu vào 13](#_heading=h.1y810tw)

[3. Trực quan hóa dữ liệu 15](#_heading=h.2xcytpi)

[4. Phân cụm với K-Means 17](#_heading=h.1ci93xb)

[5. Xây dựng hàm gợi ý 19](#_heading=h.3whwml4)

[1. Kết quả 22](#_heading=h.qsh70q)

[Tài liệu tham khảo 23](#_heading=h.49x2ik5)

# **MỞ ĐẦU**

## **Giới thiệu**

Cùng với sự phát triển mạnh mẽ của các loại hình truyền thông đa phương tiện thì âm nhạc là một trong những nội dung khá phổ biến và được xem như là một nhu cầu không thể thiếu trong cuộc sống, có thể chia sẻ bởi nhiều người từ nhiều quốc gia có ngôn ngữ và nền văn hóa khác nhau. Tuy nhiên, số lượng bài nhạc đang ngày càng tăng lên, đa dạng và phong phú cả về nội dung lẫn thể loại.

Vì vậy, vấn đề đặt ra là khi một người sử dụng muốn tìm nghe những bài nhạc mà mình yêu thích, người sử dụng sẽ cần đến công cụ tìm kiếm Google và/hoặc vào một website về âm nhạc để tìm nghe. Mặc dù vậy, ở đó có nhiều bản nhạc mà người sử dụng sẽ không thể nghe thử hết để tìm ra những bài mà họ thích (điều này tốn thời gian mà lại không hiệu quả). Do đó, nhu cầu cần có một hệ thống gợi ý có khả năng dự đoán mức độ ưa thích của người sử dụng với từng bản nhạc và gợi ý cho họ các bản nhạc mới mà hệ thống cho là phù hợp.

## **Mục tiêu đề tài**

Trong bài viết này, chúng tôi giới thiệu một hệ thống gợi ý bài hát dựa vào danh sách nhạc đã có sẵn. Từ đó gợi ý bài hát cho người dùng, giúp tăng trải nghiệm khi nghe nhạc.

**CHƯƠNG 1: Nghiên cứu tổng quan**

1. **Tổng quan về Hệ gợi ý (Recommendation system)**

A person with glasses and papers around his head

Description automatically generated Hệ thống gợi ý (Recommendation system) là một công cụ lọc thông tin hữu ích được thiết kế để dự đoán và đề xuất các mục tiêu (items) hoặc hành động có thể được người dùng quan tâm, dựa trên thông tin về sở thích, lịch sử tương tác, và các thuộc tính cá nhân khác của họ. Mục tiêu của hệ thống gợi ý là cung cấp trải nghiệm cá nhân hóa và giúp người dùng khám phá, tìm kiếm, hoặc tiêu thụ nội dung một cách hiệu quả. Hệ thống gợi ý ngày càng đóng một vai trò quan trọng và thiết yếu hơn trong các hệ thống truy cập thông tin khác nhau để thúc đẩy hoạt động kinh doanh và tạo điều kiện thuận lợi cho quá trình ra quyết định .

Hệ thống gợi ý được áp dụng rộng rãi trong nhiều lĩnh vực, bao gồm:

* Thương mại điện tử: Hỗ trợ người dùng tìm kiếm và mua sắm trực tuyến theo sở thích cá nhân.
* Phim và âm nhạc: Gợi ý phim, bài hát và nghệ sĩ tương tự dựa trên lịch sử xem và nghe của người dùng.
* Mạng xã hội: Đề xuất kết bạn, nhóm, hoặc nội dung liên quan dựa trên sự tương tác và quan tâm của người dùng.
* Tin tức và blog: Gợi ý các bài viết, bài báo hoặc blog phù hợp với quan tâm và sở thích của người đọc.
* Du lịch và nhà hàng: Đề xuất địa điểm du lịch, khách sạn hoặc nhà hàng dựa trên lịch sử du lịch và ưu tiên của người dùng.

Đối với thương mại điện tử, người mua và người bán không cần gặp gỡ trực tiếp, mà họ giao dịch với nhau trên trang web. Vì vậy, rất cần có một trợ lí bán hàng tự động. Việc tìm ra một “chuyên gia” tư vấn thông minh, thân thiện, am hiểu khách hàng trên các website vô cùng quan trọng và cũng chính là một thách thức.

Vấn đề này có thể được giải quyết bằng việc tích hợp các kĩ thuật gợi ý (Recommender Systems) trong các website bán hàng. Hệ gợi ý như 1 chuyên gia tư vấn, dự đoán thông minh sở thích của khách hàng và cung cấp những thông tin mà họ thực sự quan tâm.

Các kĩ thuật hệ gợi ý đã và đang được nghiên cứu, ứng dụng một cách mạnh mẽ và mang lại lợi ích cho cả người cung cấp dịch vụ và người sử dụng dịch vụ.

Hầu hết các trang thương mại lớn trên thế giới, như: Amazon.com, Alibaba.com, youtube.com, facebook.com, ebay.com, MovieFinder.com… đều sử dụng các kĩ thuật gợi ý trong website của mình để nâng cao trải nghiệm cho khách hàng, nâng cao chất lượng dịch vụ và thu lại lợi nhuận tốt hơn.

Ở Việt Nam, những trang thương mại điện tử thành công, như: lazada.vn, Vatgia.com, Shopee.com, www.fptshop.com.vn, Thegioididong.com, Tiki.vn… đều có sử dụng hệ gợi ý trong hệ thống thông tin.

1. **Các phương pháp gợi ý.**
   1. ***Gợi ý dựa trên Lọc cộng tác (Collaborative filtering)***

Một cách tiếp cận để thiết kế các hệ thống recommender được sử dụng rộng rãi là**lọc cộng tác**. Các phương pháp lọc cộng tác dựa trên việc thu thập và phân tích một lượng lớn thông tin về hành vi, hoạt động hoặc sở thích của người dùng và dự đoán những gì người dùng sẽ thích dựa trên sự tương đồng của họ với người dùng khác.

Một lợi thế quan trọng của phương pháp lọc cộng tác là nó không dựa vào nội dung phân tích máy và do đó nó có khả năng đề xuất chính xác các mục phức tạp như phim mà không yêu cầu “*hiểu biết*” về mục đó. Nhiều thuật toán đã được sử dụng để đo lường sự giống nhau của người dùng hoặc sự tương đồng về mặt hàng trong các hệ thống giới thiệu. Ví dụ, cách tiếp cận hàng xóm gần nhất (k-nearest neighbor k-NN) và Pearson Correlation được Allen triển khai lần đầu tiên.

Lọc cộng tác dựa trên giả định rằng những người đã đồng ý trong quá khứ sẽ đồng ý trong tương lai và rằng họ sẽ thích các loại mặt hàng tương tự như họ thích trong quá khứ.

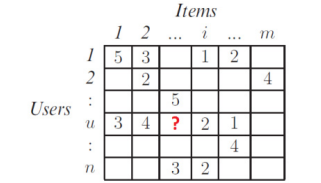
1. **Lọc cộng tác dựa trên người dùng (User-Based Collaborative Filtering):**

Phương pháp này đề xuất các mục cho một người dùng dựa trên sự tương đồng giữa người dùng đó và các người dùng khác. Cụ thể, nó so sánh sự tương đồng giữa người dùng dựa trên các đánh giá và sở thích của họ.

Hệ thống tìm kiếm những người dùng có sở thích tương tự với người dùng hiện tại và đề xuất những mục tiêu được đánh giá cao bởi những người dùng đó.

1. **Lọc cộng tác dựa trên sản phẩm (Item-Based Collaborative Filtering):**

Phương pháp này đề xuất các items cho một người dùng dựa trên sự tương đồng giữa các item. Nó tìm các item tương tự với các item được xem xét và đề xuất các item tương tự đó.

 Hệ thống xác định tương đồng giữa các mục tiêu và đề xuất các mục tiêu tương tự với những mục tiêu đã được người dùng đánh giá cao.

**Hình 1: Ma trận biểu diễn xếp hạng của người dùng trên mục tin (user-item-rating matrix)**

1. **Ưu điểm và nhược điểm của lọc cộng tác**

* *Ưu điểm:*
* Không yêu cầu thông tin mô tả về mục tiêu hoặc người dùng
* Có tính cá nhân hóa, tạo ra gợi ý cá nhân dựa trên hành vi cả người dùng
* Có thể hoạt động hiệu quả trong môi trường có nhiều người dùng và mục tiêu, đặc biệt khi dữ liệu có sẵn.
* *Nhược điểm:*
* Gặp vấn đề cold start: hệ thống gặp khó khăn trong việc đưa ra gợi ý cho người dùng mới hoặc các mục mới được thêm vào hệ thống
* Khả năng mở rộng : Trong nhiều môi trường mà các hệ thống này đưa ra các khuyến nghị, có hàng triệu người dùng và sản phẩm. Do đó, một lượng lớn công suất tính toán thường là cần thiết để tính toán các gợi ý.
* Sự thua thớt : Số lượng các mặt hàng được bán trên các trang web thương mại điện tử lớn là cực kỳ lớn. Những người dùng tích cực nhất sẽ chỉ đánh giá một tập con nhỏ của cơ sở dữ liệu tổng thể. Do đó, ngay cả những mặt hàng phổ biến nhất cũng có rất ít xếp hạng.
  1. ***Gợi ý dựa trên nội dung (Content-Based Recommendation)***

Gợi ý dựa trên nội dung là một phương pháp gợi ý dựa trên đặc điểm và thuộc tính của các mục tiêu (sản phẩm, nội dung, v.v.) và sở thích của người dùng.

Hệ thống xây dựng mô hình hồ sơ người dùng và mô hình hồ sơ mục tiêu dựa trên các thuộc tính như từ khóa, thể loại, mô tả, v.v.

Khi người dùng yêu cầu gợi ý, hệ thống so sánh hồ sơ người dùng với hồ sơ mục tiêu và đề xuất các mục tiêu có thuộc tính tương tự.

Ví dụ: một user xem rất nhiều các bộ phim về cảnh sát hình sự, vậy thì gơi ý một bộ phim trong cơ sở dữ liệu có chung đặc tính hình sự tới user này, ví dụ phim Người phán xử.

* *Ưu điểm:*
* Tăng tính cá nhân hóa và khả năng cung cấp nội dung phù hợp với người dùng.
* Độc lập với thông tin từ người dùng khác.
* Giải quyết được vấn đề cold start
* Hữu ích khi dữ liệu về hành vi người dùng hạn chế.
* *Nhược điểm:*
* Phụ thuộc vào việc xác định các thuộc tính và đặc điểm quan trọng để tạo hồ sơ mục tiêu và người dùng.
* Không thể đánh giá sở thích phức tạp của người dùng.
* Nếu người dùng thường xem những mục tiêu tương tự và không được đề xuất nôi dung mới thì họ sẽ chỉ tiếp cận với các mục tiêu giống nhau
  1. ***Hệ thống gợi ý lai (Hybrid recommender systems)***

Nghiên cứu gần đây đã chứng minh rằng một phương pháp lai, kết hợp lọc cộng tác và lọc dựa trên nội dung có thể hiệu quả hơn trong một số trường hợp. Các phương pháp lai có thể được thực hiện theo nhiều cách:

* Bằng cách đưa ra các dự đoán dựa trên nội dung và dựa trên lọc cộng tác riêng biệt và sau đó kết hợp chúng.
* Bằng cách thêm các khả năng dựa trên nội dung vào phương pháp cộng tác (và ngược lại).
* Bằng cách thống nhất các phương pháp tiếp cận thành một mô hình.

Một số nghiên cứu thực nghiệm so sánh hiệu suất của phương pháp lai với các phương pháp cộng tác thuần túy và chứng minh rằng các phương pháp lai có thể cung cấp các khuyến nghị chính xác hơn các phương pháp thuần túy. Những phương pháp này cũng có thể được sử dụng để khắc phục một số vấn đề thường gặp trong hệ thống gợi ý như Cold Start và vấn đề thưa thớt.

Một loạt các kỹ thuật đã được đề xuất làm cơ sở cho các hệ thống gợi ý: các kỹ thuật hợp tác (collaborative), dựa trên nội dung (content-based), dựa trên kiến ​​thức (knowledge-based) và nhân khẩu học (demographic techniques). Mỗi kỹ thuật này đều có những thiếu sót, như vấn đề Cold Start cho các hệ thống cộng tác và dựa trên nội dung (phải làm gì với người dùng mới với ít xếp hạng) và tắc nghẽn kỹ thuật tri thức (knowledge engineering bottleneck) trong các phương pháp dựa trên tri thức . Một hệ thống gợi ý lai là một hệ thống trong đó kết hợp nhiều kỹ thuật với nhau để đạt được một số sức mạnh tổng hợp giữa chúng.

**CHƯƠNG 2: Cơ sở lý thuyết**

1. **Thuật toán phân cụm K-Means**

K-means là một thuật toán phân cụm đơn giản thuộc loại học không giám sát(tức là dữ liệu không có nhãn) và được sử dụng để giải quyết bài toán phân cụm. Ý tưởng của thuật toán phân cụm k-means là phân chia 1 bộ dữ liệu thành các cụm khác nhau. Trong đó số lượng cụm được cho trước là k. Công việc phân cụm được xác lập dựa trên nguyên lý: Các điểm dữ liệu trong cùng 1 cụm thì phải có cùng 1 số tính chất nhất định. Tức là giữa các điểm trong cùng 1 cụm phải có sự liên quan lẫn nhau. Đối với máy tính thì các điểm trong 1 cụm đó sẽ là các điểm dữ liệu gần nhau.

A diagram of different colored circles

Description automatically generated with medium confidence

**Hình 2: Bài toán với 3 clusters**

***Ý tưởng của thuật toán k-means***

1. Khởi tạo K điểm dữ liệu trong bộ dữ liệu và tạm thời coi nó là tâm của các cụm dữ liệu của chúng ta.
2. Với mỗi điểm dữ liệu trong bộ dữ liệu, tâm cụm của nó sẽ được xác định là 1 trong K tâm cụm gần nó nhất.
3. Sau khi tất cả các điểm dữ liệu đã có tâm, tính toán lại vị trí của tâm cụm để đảm bảo tâm của cụm nằm ở chính giữa cụm.
4. Bước 2 và bước 3 sẽ được lặp đi lặp lại cho tới khi vị trí của tâm cụm không thay đổi hoặc tâm của tất cả các điểm dữ liệu không thay đổi.

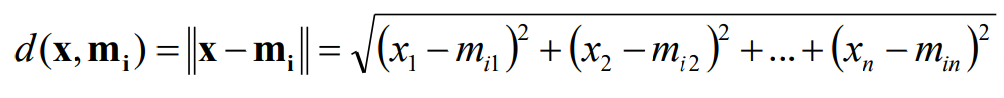
**Xác định điểm trung tâm: Điểm trung bình (Mean centroid)**

**A math symbols with numbers and symbols

Description automatically generated with medium confidence**

* (vectơ) **mi** là điểm trung tâm (centroid) của cụm **Ci**
* |Ci | kích thước của cụm Ci (tổng số quan sát trong Ci )

**Hàm khoảng cách: Euclidean distance**



* (vectơ) **mi** là điểm trung tâm (centroid) của cụm **Ci**
* d(**x,mi** ) là khoảng cách giữa **x** và điểm trung tâm **mi**

**Ưu điểm và nhược điểm của K-Means**

* ***Ưu điểm:***
* Đơn giản: dễ cài đặt, rất dễ hiểu
* Rất linh động: cho phép dùng nhiều độ đo khoảng cách khác nhau → phù hợp với các loại dữ liệu khác nhau.
* Hiệu quả (khi dùng độ đo Euclide)

+ Độ phức tạp tính toán tại mỗi bước ~ O(r.k)

▪ r: Tổng số các quan sát (kích thước của tập dữ liệu)

▪ k: Tổng số cụm thu được

◼ Thuật toán có độ phức tạp trung bình là đa thức.

* K-means là giải thuật phân cụm được dùng phổ biến nhất
* ***Nhược điểm:***
* Số cụm k phải được xác định trước

◼ Thường ta không biết chính xác !

* Giải thuật K-means nhạy cảm (gặp lỗi) với các quan sát ngoại lai (outliers)

+ Các quan sát ngoại lai là các quan sát (rất) khác biệt với tất các quan sát khác

+ Các quan sát ngoại lai có thể do lỗi trong quá trình thu thập/lưu dữ liệu

+ Các quan sát ngoại lai có các giá trị thuộc tính (rất) khác biệt với các giá trị thuộc tính của các quan sát khá

1. **Hàm khoảng cách cosine (Cosine Similarity)**

Độ đo tương đồng Cosine Similarity (tương đồng cosin) là một trong những phép đo tương đồng phổ biến được sử dụng trong lọc cộng tác và xử lý văn bản để đo lường sự tương tự giữa các vectơ. Độ đo này đo lường cosin của góc giữa hai vectơ trong không gian nhiều chiều và cho biết mức độ tương tự giữa chúng. Độ đo Cosine Similarity thường sử dụng trong việc so sánh tương tự giữa các người dùng hoặc các mục dựa trên đánh giá của họ.

Công thức tính Cosine Similarity giữa hai vectơ A và B được xác định bằng cách tính tích vô hướng của chúng và chia cho tích của độ dài (norm) của hai vectơ:

**A black text on a white background

Description automatically generated**

Trong đó:

* A và B là hai vectơ cần so sánh.
* A · B là tích vô hướng (dot product) giữa hai vectơ A và B.
* |A| và |B| là độ dài (norm) của hai vectơ A và B, tương ứng.

Giá trị Cosine Similarity nằm trong khoảng từ -1 đến 1:

* Nếu Cosine Similarity bằng 1, nghĩa là hai vectơ hoàn toàn tương tự và có cùng hướng.
* Nếu Cosine Similarity bằng 0, nghĩa là hai vectơ không có sự tương tự hoặc không tương quan.
* Nếu Cosine Similarity gần -1, nghĩa là hai vectơ tương tự nhưng có hướng trái ngược.

**CHƯƠNG 3: Xây dựng hệ thống**

* 1. **Các bước chính xây dựng hệ thống**

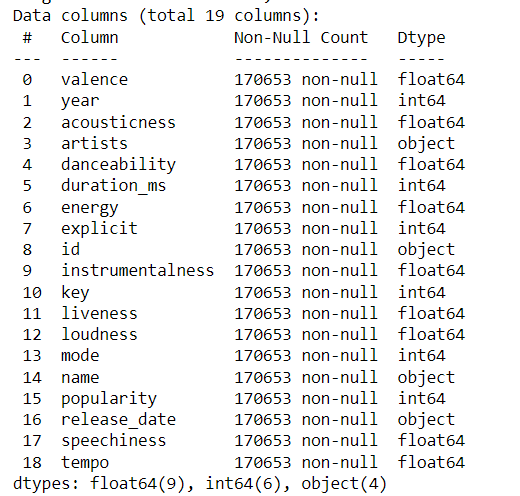
B1: Đọc dữ liệu vào

B2: Trực quan hóa dữ liệu và EDA

B3: Phân cụm với K-Means

B4: Đưa ra gợi ý

* 1. **Đọc dữ liệu vào**

Ta đã có sẵn 1 bộ dataset về bài hát và các thông tin của từng bài hát của từng bài hát từ Kaggle. Bộ dataset này có khoảng 29000 bản ghi bài hát, mỗi bản ghi có khoảng 20 thuộc tính. Ta chuẩn hóa các giá trị theo dải từ 0 đến 1.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A graph of a bar graph

Description automatically generated Sau khi đọc dữ liệu ta chuẩn hóa dữ liệu, ta sẽ kiểm tra tất cả các phân tích với mục tiêu là “popularity”.

* 1. **Trực quan hóa dữ liệu**

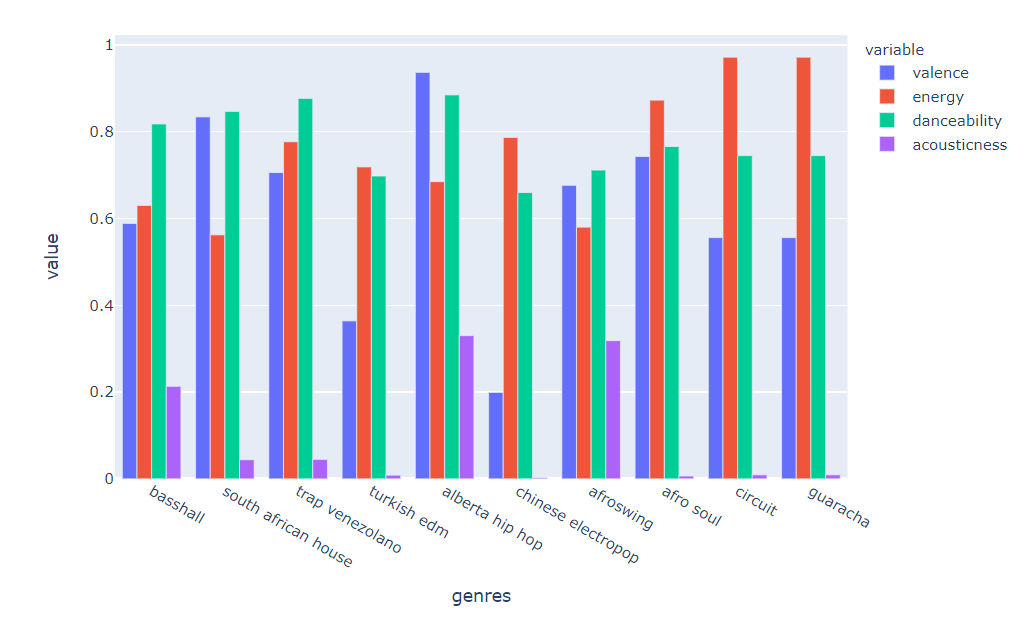
**A graph of different colored bars

Description automatically generated**Sử dụng dữ liệu được nhóm theo năm, chúng ta có thể hiểu âm thanh tổng thể của âm nhạc đã thay đổi như thế nào từ năm 1921 đến năm 2020.

**A graph showing different colored lines

Description automatically generated**

Tập dữ liệu này chứa các tính năng âm thanh cho các bài hát khác nhau cùng với các tính năng âm thanh cho các thể loại khác nhau. Chúng ta có thể sử dụng thông tin này để so sánh các thể loại khác nhau và hiểu sự khác biệt độc đáo về âm thanh của chúng.

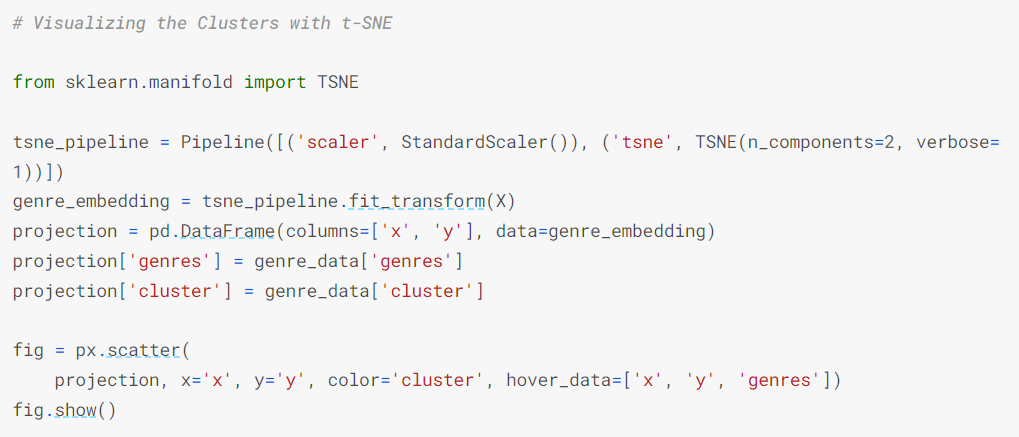
****

* 1. **Phân cụm với K- Means**

1. ***Phân cụm thể loại với K-Means***

**A screen shot of a computer code

Description automatically generated** Ở đây, thuật toán phân cụm K-mean đơn giản được sử dụng để chia các thể loại trong tập dữ liệu này thành mười cụm dựa trên các đặc điểm âm thanh số của từng thể loại.

 Tiếp theo thực hiện việc giảm chiều dữ liệu và hiển thị dữ liệu trong không gian 2 chiều sử dụng phương pháp t-SNE (t-Distributed Stochastic Neighbor Embedding).

A colorful dots on a graph

Description automatically generatedVà đây là kết quả:

1. ***Phân cụm bài hát với K-Means***

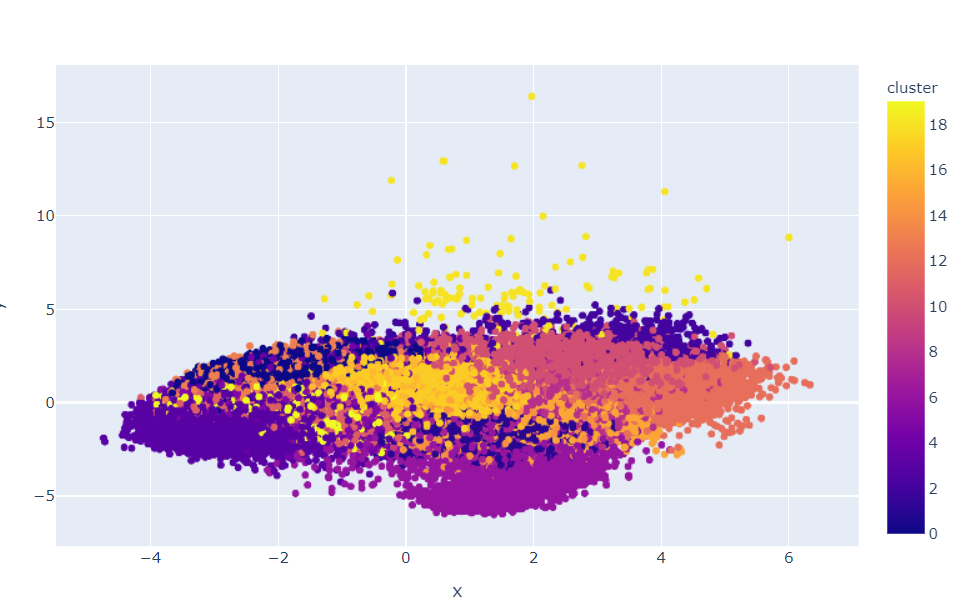
A screen shot of a computer code

Description automatically generated

Thực hiện việc sử dụng phân tích thành phần chính (PCA) để giảm chiều dữ liệu và sau đó tạo biểu đồ scatter plot để hiển thị sự phân bố của các bài hát trong không gian hai chiều

A screen shot of a computer code

Description automatically generated

Và đây là kết quả biểu đồ:

* 1. **Xây dựng hàm gợi ý**

Dựa trên phân tích và trực quan hóa, ta thấy các thể loại tương tự có xu hướng có các điểm dữ liệu nằm gần nhau trong khi các loại bài hát tương tự cũng được nhóm với nhau.

Các thể loại tương tự sẽ có âm thanh giống nhau và sẽ đến từ những khoảng thời gian giống nhau trong khi điều tương tự cũng có thể đối với các bài hát thuộc thể loại đó. Chúng ta có thể sử dụng ý tưởng này để xây dựng hệ thống gợi ý bằng cách lấy điểm dữ liệu của các bài hát mà người dùng đã nghe và đề xuất các bài hát tương ứng với các điểm dữ liệu gần đó.

A screenshot of a computer

Description automatically generated Spotify là một Python client dành cho API Web Spotify, giúp ta có thể dễ dàng tìm và truy vấn danh mục bài hát của Spotify. Sau khi cài đặt Spotipy, bạn sẽ cần tạo một ứng dụng trên trang của Nhà phát triển Spotify và lưu Client ID và secret key của mình.

Từ đó, ta xây dựng được hệ thống gợi ý sau:

A screenshot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated

Kết quả sau khi chạy hàm gợi ý ta được:

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Tài liệu tham khảo**

[1] Bài giảng môn Hệ gợi ý - Đại học Bách Khoa Hà Nội

[2] Scalable Machine Learning - Recommender Systems, Alex Smola

[3] Deep Learning based Recommender System: A Survey and New Perspectives, SHUAI ZHANG, LINA YAO, AIXIN SUN

[4] A Survey of Collaborative Filtering Techniques, Xiaoyuan Su, Taghi M. Khoshgoftaar

[5] Deep Learning for Sequential Recommendation: Algorithms, Influential Factors, and Evaluations, HUI FANG, DANNING ZHANG∗ , YIHENG SHU, GUIBING GUO

[6] Wikipedia – Hệ gợi ý