**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**Logo, icon

Description automatically generated**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN HỌC: MẠNG XÃ HỘI - IS353.N11.HTCL**

**ĐỀ TÀI:**

**KHẢO SÁT MẠNG CẠNH TRANH CÁC THỂ LOẠI NHẠC GIỮA NHỮNG CA SĨ, NGHỆ SĨ THÔNG QUA BÀI HÁT**

**Giảng viên hướng dẫn:**

**NGUYỄN THỊ KIM PHỤNG**

**Sinh viên thực hiện:**

**NGUYỄN TRÍ MINH - 19521847**

**Thành phố Hồ Chí Minh, 10/2022**

MỤC LỤC

[MỤC LỤC 3](#_Toc117713021)

[NHẬN XÉT CỦA GIẢNG VIÊN 4](#_Toc117713022)

[QUẢN LÝ TÀI LIỆU 5](#_Toc117713023)

[phân chia công việc 6](#_Toc117713024)

[PHẦN A. XỬ LÝ BẰNG phương pháp LẬP TRÌNH 7](#_Toc117713025)

[CHƯƠNG 1. mở đầu 7](#_Toc117713026)

[1.1. Giới thiệu 7](#_Toc117713027)

[1.2. Tiếp cận bài toán 7](#_Toc117713028)

[1.3. Dữ liệu 7](#_Toc117713029)

[PHẦN b. XỬ LÝ BẰNG phương pháp THỦ CÔNG 9](#_Toc117713030)

NHẬN XÉT CỦA GIẢNG VIÊN

*……., ngày……... tháng…… năm 2022*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên)*

QUẢN LÝ TÀI LIỆU

|  |  |  |  |
| --- | --- | --- | --- |
| **Thời gian** | **Phiên bản** | **Người sửa đổi** | **Chi tiết** |
| 26/10/2022 | v1.0 | Nguyễn Trí Minh | Khởi tạo tài liệu |
| 27/10/2022 | v1.1 | Nguyễn Trí Minh | Thêm nội dung và chỉnh sửa |

phân chia công việc

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Nội dung công việc** | **Ước tính % đóng góp** |
| 1 | Nguyễn Trí Minh | 19521847 | Thực hiện đồ án | 100 |

PHẦN A. XỬ LÝ BẰNG phương pháp LẬP TRÌNH

Giới thiệu về phương pháp lập trình:

* Ngôn ngữ sử dụng: Python.
* IDE: Visual Studio Code.
* Extension: Jupiter và Jupiter Keymap, cho phép thực thi Jupiter notebook trực tiếp trên Visual Studio Code.

# mở đầu

## Giới thiệu

Ngày nay, nhu cầu giải trí đa phương tiện là nhu cầu không thể thiếu đối với mỗi chúng ta. Từ lâu, việc nghe nhạc hay thưởng thức âm nhạc, vốn là một thành phần quan trọng trong nhiều thể loại giải trí đa phương tiện, luôn là một phương thức hiệu quả để giúp chúng ta giải trí sau những giờ làm việc hay học tập căng thẳng. Thị trường âm nhạc luôn có nhiều biến động bởi nhiều bài hát với giai điệu hấp dẫn tương ứng với nhiều thể loại nhạc khác nhau (như Pop, R&B, Rap/Hiphop, Rock,…) được sáng tác và biểu diễn bởi các ca sĩ, nhà sản xuất âm nhạc, nghệ sĩ nổi tiếng trên khắp thế giới. Việc ngày càng xuất hiện nhiều nền tảng (platform) nghe nhạc điện tử như Spotify, Apple Music, Deezer,… giúp cho người dùng dễ dàng tiếp cận và nghe nhạc một cách dễ dàng hơn để ủng hộ cho những nghệ sĩ, ca sĩ ấy. Từ đó, việc cạnh tranh về những thể loại âm nhạc của thị trường âm nhạc trên những nền tảng nói trên của các ca sĩ ngày càng sôi nổi hơn bao giờ hết. Chính vì thế, môn học Mạng xã hội trong phạm vi của môn học có thể giúp cho chúng ta phân tích **mạng cạnh tranh trên thị trường âm nhạc** của các thể loại nhạc giữa những ca sĩ thông qua bài hát của họ bằng các độ đo (dựa trên đường đi, phổ, khoảng cách,...) và khai thác dữ liệu trên mạng này (bằng phương pháp gom cụm) một cách có hiệu quả.

## Tiếp cận bài toán

* Input: Tập dữ liệu ban đầu trên nguồn dữ liệu Kaggle đã qua tiền xử lý dữ liệu.
* Output: Đưa ra độ đo, đưa ra cộng đồng, phục vụ cho việc phân tích mạng xã hội - “top các bài hát trên nền tảng Spotify từ năm 2010 đến 2019”.

## Dữ liệu

* + 1. **Giới thiệu nguồn dữ liệu**
* Nguồn dataset: *Top Spotify songs from 2010-2019 - BY YEAR*, <https://www.kaggle.com/datasets/leonardopena/top-spotify-songs-from-20102019-by-year>.
* Dữ liệu “Top Spotify songs from 2010-2019 - BY YEAR” được cung cấp trên nền tảng Kaggle chứa dữ liệu các bài hát trên nền tảng Spotify từ năm 2010 đến 2019. Dữ liệu gồm 603 dòng với 15 thuộc tính bao gồm:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| 1 | ID | int | ID của bài hát |
| 2 | title | char | Tên bài hát |
| 3 | artist | char | Tên ca sĩ |
| 4 | top genre | char | Thể loại của bài hát |
| 5 | year | int | Năm phát hành của bài hát trên Billboard |
| 6 | bpm | int | Tempo của bài hát |
| 7 | nrgy | int | Chỉ số năng lượng của bài hát |
| 8 | dnce | int | Chỉ số mang lại cảm hứng “nhảy” khi nghe bài hát |
| 9 | dB | int | Độ lớn âm thanh của bài hát |
| 10 | live | int | Chỉ số cho biết khả năng có thể được thu trực tiếp của bài hát |
| 11 | val | int | Chỉ số cho biết khả năng mang lại sự tích cực từ bài hát |
| 12 | dur | int | Độ dài của bài hát |
| 13 | acous | int | Chỉ số acoustic của bài hát |
| 14 | spch | int | Chỉ số cho biết độ lớn theo số lượng từ được hát từ bài hát |
| 15 | pop | int | Chỉ số độ phổ biến của bài hát |

* + 1. **Xử lý và phân tích dữ liệu**
* Làm sạch dữ liệu: Nhận thấy không có thuộc tính bị thiếu dữ liệu, tiến hành loại bỏ dữ liệu trùng:

A screenshot of a computer

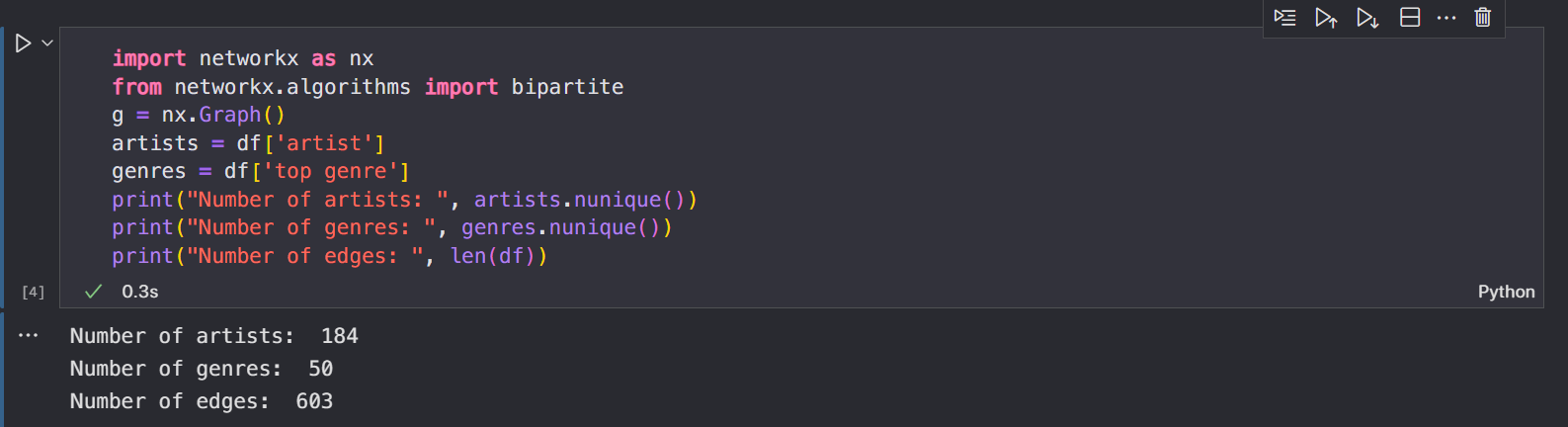
Description automatically generated with medium confidence

A screen shot of a computer

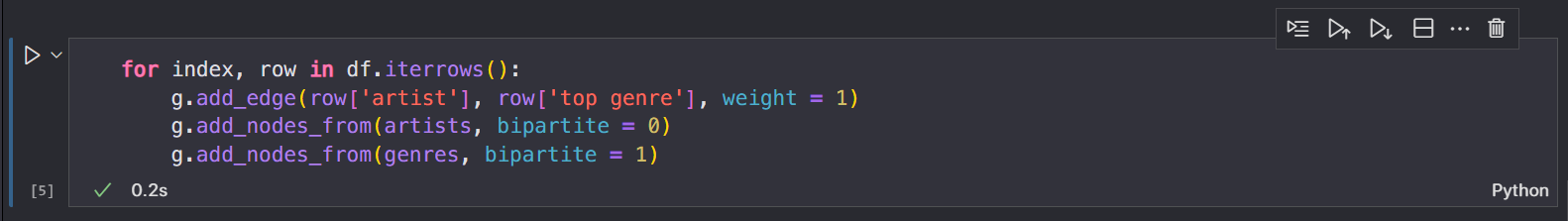
Description automatically generated with low confidence

Dữ liệu được đọc từ file “top\_song.csv” đưa vào dataframe. Sau khi xoá bỏ dữ liệu trùng thì vẫn được dataset có 603 dòng và 15 cột, chứng tỏ dữ liệu đã được tiền xử lý tốt, không có giá trị trùng lập và giá trị rỗng.

* Chuyển đổi dataframe thành đồ thị:
* Đồ thị 2 phía:
* Node: Ca sĩ và thể loại bài hát, edge: mối quan hệ giữa việc ca sĩ phát hành 1 bài nhạc thuộc thể loại nào đó.
* In ra số ca sĩ, số thể loại và số cạnh:



* Đưa dữ liệu vào đồ thị 2 phía:



* Hiển thị đồ thị 2 phía:
* Đồ thị 1 phía:

PHẦN b. XỬ LÝ BẰNG phương pháp THỦ CÔNG