**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG**

**THÀNH PHỐ HỒ CHÍ MINH**

A circle with a book and a person with a leaf

Description automatically generated

**BÁO CÁO CUỐI KÌ**

**CÔNG NGHỆ DỮ LIỆU LỚN**

**ĐỀ TÀI: PHÂN TÍCH SỰ QUAN TÂM CỦA NGƯỜI DÙNG FACEBOOK ĐỐI VỚI MỘT SẢN PHẨM**

|  |  |  |
| --- | --- | --- |
| **Thành viên** | **:** | **Lý Gia Hưng Hưng -1050080053** |
|  |  | **Hoàng Phúc Lộc – 1050080058** |
|  |  | **Nguyễn Khắc Thiên – 1050080075** |
| **Lớp** | **:** | **K10 – ĐH – Công nghệ thông tin 2** |
| **Môn học** | **:** | **Công nghệ dữ liệu lớn** |

Tp. Hồ Chí Minh, ngày tháng năm

**PHÂN TÍCH SỰ QUAN TÂM CỦA NGƯỜI DÙNG FACEBOOK ĐỐI VỚI MỘT SẢN PHẨM**

**THÔNG TIN NHÓM 11**

Link youtube:

+Phân tích reactions : <https://youtu.be/0O_1evgVUHk?si=5r9qu1cBTMxZAySa>

+Phân tích comments : <https://youtu.be/32hDdZeS7Go?si=riRc0UdPiWW_1pBK>

|  |  |
| --- | --- |
| **Họ và tên thành viên** | **Đóng góp và tự đánh giá** |
| 1.    Lý Gia Hưng – 1050080053  A portrait of a person  Description automatically generated | -  Đóng góp: Cài đặt mô hình, thực nghiệm  - Tự đánh giá quá trình: 8  - Tự đánh giá cuối kì: 8 |
| 2.    Hoàng Phúc Lộc - 1050080058  A portrait of a young person  Description automatically generated | -  Đóng góp: Cài đặt mô hình, thực nghiệm  - Tự đánh giá quá trình: 8  - Tự đánh giá cuối kì:8 |
| 3.    Nguyễn Khắc Thiên -1050080075  A person in a white shirt  Description automatically generated | -  Đóng góp: Cài đặt mô hình, thực nghiệm  - Tự đánh giá quá trình: 8  - Tự đánh giá cuối kì: 8 |

PHỤ LỤC

[TÓM TẮT BÁO CÁO 5](#_Toc150310629)

[1. Lý do chọn đề tài 5](#_Toc150310630)

[2. Tầm quan trọng của đề tài 5](#_Toc150310631)

[3. Nội dung báo cáo 7](#_Toc150310632)

[BÁO CÁO CHI TIẾT 8](#_Toc150310633)

[1. Lý thuyết 8](#_Toc150310634)

[1.1. Big Data là gì ? 8](#_Toc150310635)

[1.2. Bài toán 9](#_Toc150310636)

[2. Công nghệ sử dụng 10](#_Toc150310637)

[2.1. Môi trường thực nghiệm (IDE): Google Colab 10](#_Toc150310638)

[2.2. Ngôn ngữ sử dụng: Python 10](#_Toc150310639)

[2.3. Cloud : Googe Driver 11](#_Toc150310640)

[2.4. Facebook Graph API 11](#_Toc150310641)

[2.5. facebook\_scaper 11](#_Toc150310642)

[2.6. Matplotlib 11](#_Toc150310643)

[2.7. Mô hình Naive Bayes 11](#_Toc150310644)

[3. Mô hình 13](#_Toc150310645)

[4. Cài đặt thực nghiệm 14](#_Toc150310646)

[4.1. Lấy Token có quyền Facebook Graph API 14](#_Toc150310647)

[4.2. Lấy dữ liệu comments của bài viết bằng API Graph Facebook 17](#_Toc150310648)

[4.3. Phân tích nội dung comment bằng toán TF-IDF và mô hình Naive Bayes 20](#_Toc150310649)

[4.4. Lấy thông tin của người comment bài viết 26](#_Toc150310650)

[4.5. Tiền xử lý dữ liệu thông tin của người comments 31](#_Toc150310651)

[4.6. Phân tích thông tin người dùng comments – Biểu diễn biểu đồ 34](#_Toc150310652)

[4.6.1. Phân tích giới tính người comment 34](#_Toc150310653)

[4.6.2. Phân tích tuổi người comment 36](#_Toc150310654)

[4.6.3. Phân tích nơi ở hiện tại (top 10 trong danh sách comment) 37](#_Toc150310655)

[4.7. Lấy dữ liệu reactions bằng facebook\_scaper 38](#_Toc150310656)

[4.8. Phân tích reactions 40](#_Toc150310657)

[5. Kết quả và đánh giá 42](#_Toc150310658)

[5.1. Kết quả 42](#_Toc150310659)

[5.2. Đánh giá 45](#_Toc150310660)

[6. Tài liệu tham khảo 45](#_Toc150310661)

# TÓM TẮT BÁO CÁO

## 1. Lý do chọn đề tài

Hiểu sâu về người dùng: Phân tích dữ liệu từ Facebook giúp doanh nghiệp hiểu rõ hơn về đối tượng tiềm năng, sở thích, và nhu cầu của họ. Điều này có thể giúp cải thiện chiến dịch tiếp thị và phát triển sản phẩm.

Đánh giá sự quan tâm: Dữ liệu từ Facebook có thể giúp đánh giá mức độ quan tâm của người dùng đối với sản phẩm, dự đoán sự thành công hoặc thất bại của sản phẩm trước khi ra mắt.

Tối ưu hóa chiến dịch tiếp thị: Dựa trên thông tin từ Big Data, doanh nghiệp có thể tối ưu hóa chiến dịch tiếp thị, chọn đúng phương tiện truyền thông xã hội và thời điểm phù hợp để tạo sự quan tâm đối với sản phẩm.

Đối phó với phản hồi của người dùng: Big Data cũng giúp doanh nghiệp phát hiện sớm phản hồi và phàn nàn của người dùng, giúp họ có cơ hội giải quyết vấn đề và cải thiện sản phẩm.

Tóm lại, đề tài này đem lại giá trị to lớn cho việc phát triển sản phẩm và chiến dịch tiếp thị thông qua sử dụng Big Data để nắm bắt thông tin về sự quan tâm của người dùng trên mạng xã hội.

## 2. Tầm quan trọng của đề tài

Tầm quan trọng của đề tài "Phân tích sự quan tâm của người dùng Facebook đối với một sản phẩm" trong lĩnh vực Big Data là không thể phủ nhận. Dưới đây là một số điểm quan trọng về tầm quan trọng của đề tài này:

* Hiểu rõ người dùng và thị trường: Việc nắm bắt sự quan tâm của người dùng Facebook đối với một sản phẩm giúp doanh nghiệp hiểu rõ hơn về nhu cầu và mong muốn của khách hàng tiềm năng. Điều này giúp cải thiện chiến lược tiếp thị và phát triển sản phẩm để đáp ứng tốt hơn sự quan tâm của họ.
* Tối ưu hóa chiến dịch tiếp thị: Bằng cách phân tích dữ liệu về sự quan tâm của người dùng, doanh nghiệp có thể tạo ra các chiến dịch tiếp thị tập trung và hiệu quả hơn. Điều này giúp tiết kiệm tài nguyên và tăng cường khả năng thu hút sự quan tâm của khách hàng.
* Phát triển sản phẩm và dịch vụ: Hiểu rõ sự quan tâm của người dùng cũng giúp doanh nghiệp phát triển sản phẩm và dịch vụ phù hợp hơn với thị trường. Dựa trên thông tin này, bạn có thể điều chỉnh hoặc mở rộng danh mục sản phẩm để đáp ứng nhu cầu người dùng.
* Dự đoán xu hướng thị trường: Sự quan tâm của người dùng trên mạng xã hội có thể phản ánh những xu hướng tiềm năng trong tương lai. Phân tích dữ liệu sự quan tâm này có thể giúp doanh nghiệp dự đoán và chuẩn bị cho những thay đổi trong thị trường.
* Tối ưu hóa kinh doanh dựa trên dữ liệu: Big Data cho phép bạn thu thập và phân tích lượng lớn dữ liệu từ người dùng Facebook. Điều này giúp bạn tối ưu hóa quy trình kinh doanh, từ quản lý cổng thông tin đến dự đoán sự quan tâm của khách hàng.
* Cạnh tranh hiệu quả: Bằng cách hiểu rõ sự quan tâm của khách hàng và thị trường, doanh nghiệp có thể cạnh tranh hiệu quả hơn. Họ có thể thích nghi nhanh chóng với thay đổi và cung cấp giá trị tốt hơn cho khách hàng.

Tóm lại, đề tài "Phân tích sự quan tâm của người dùng Facebook đối với một sản phẩm" trong lĩnh vực Big Data đóng vai trò quan trọng trong việc giúp doanh nghiệp hiểu rõ khách hàng, tối ưu hóa chiến dịch tiếp thị, phát triển sản phẩm và dịch vụ, dự đoán xu hướng thị trường và cạnh tranh hiệu quả. Điều này có thể giúp tăng trưởng kinh doanh và đảm bảo sự cạnh tranh trong môi trường kinh doanh ngày càng cạnh tranh.

## 3. Nội dung báo cáo

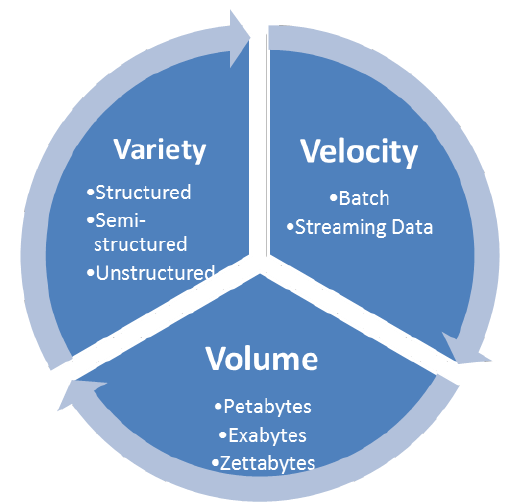
* Lý thuyết
* Công nghệ sử dụng
* Mô hình
* Cài đặt thực nghiệm
* Kết quả và đánh giá

# BÁO CÁO CHI TIẾT

## 1. Lý thuyết

### 1.1. Big Data là gì ?

Big Data là một thuật ngữ thường được sử dụng để mô tả lượng dữ liệu lớn và phức tạp, vượt xa khả năng của công cụ và phương pháp truyền thống để quản lý, xử lý và phân tích. Dữ liệu này có thể bao gồm thông tin từ nhiều nguồn khác nhau như máy tính cá nhân, thiết bị di động, cảm biến, mạng xã hội, dữ liệu giao dịch, hình ảnh, video và nhiều nguồn dữ liệu khác. Dữ liệu Big Data thường được đánh giá dựa trên ba yếu tố chính, được gọi là "3V":



- Volume (Lượng dữ liệu): Big Data thường chứa lượng dữ liệu lớn. Điều này có thể bao gồm terabytes, petabytes hoặc thậm chí exabytes dữ liệu.

- Velocity (Tốc độ): Dữ liệu này thường được tạo ra và cập nhật với tốc độ nhanh, đặc biệt là trong các ngữ cảnh như Internet of Things (IoT) hoặc trên mạng xã hội.

- Variety (Đa dạng): Big Data đến từ nhiều nguồn, và nó có thể tồn tại ở nhiều định dạng khác nhau, bao gồm văn bản, hình ảnh, video, âm thanh và dữ liệu không cấu trúc.

Big Data có nhiều ứng dụng quan trọng, bao gồm:

- Phân tích dự đoán: Sử dụng Big Data để dự đoán xu hướng tương lai, ví dụ như dự đoán sự quan tâm của người dùng đối với một sản phẩm.

- Quản lý dữ liệu khách hàng: Doanh nghiệp sử dụng Big Data để hiểu rõ hơn về khách hàng của họ và cải thiện dịch vụ hoặc sản phẩm.

- Nghiên cứu khoa học: Big Data hỗ trợ nghiên cứu trong nhiều lĩnh vực, từ y học đến khoa học xã hội.

- An toàn mạng: Big Data có vai trò quan trọng trong phát hiện và ngăn chặn các mối đe dọa an ninh trực tuyến.

- Dự đoán và tối ưu hóa vận hành: Các ngành công nghiệp như logistics và sản xuất sử dụng Big Data để tối ưu hóa quy trình vận hành và dự đoán cần tài nguyên gì và khi nào.

Big Data đã trở thành một phần quan trọng của nhiều lĩnh vực và mang lại nhiều cơ hội và thách thức cho doanh nghiệp và tổ chức trên khắp thế giới.

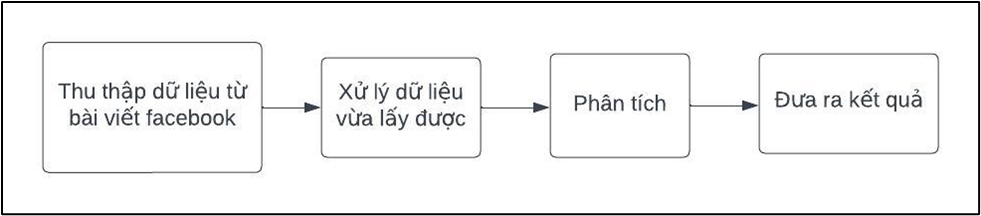
### 1.2. Bài toán

Bài toán: Các doanh nghiệp đăng bài về sản phẩm của họ lên faebook dưới dạng một bài post. Để đánh giá sản phẩm đó, chúng ta không thể nào ngồi đếm từng lượt tương tác (reactions) của người dùng facebook cũng như không thể ngồi đọc từng comment họ nói về sản phẩm được.

Bài toán đặt ra cho chủ đề này là phải thu thập dữ liệu từ một bài viết

* Thu thập về lượt reaction: thích, thương thương, yêu thích, bất ngờ, tức giận, buồn bã.
* Thông tin của comment
* Thông tin của người dùng tương tác với bài viết: tuổi, giới tính, nơi sống, quê quán,..

Sau khi có được dữ liệu, ta phải xử lý dữ liệu rồi phân tích sự quan tâm từ số liệu mới thu được rồi cuối cùng là đưa ra kết quả phân tích.



## 2. Công nghệ sử dụng

### 2.1. Môi trường thực nghiệm (IDE): Google Colab

Google Colab là một môi trường phát triển tích hợp trực tuyến (Integrated Development Environment - IDE) dựa trên nền tảng Jupyter Notebook. Môi trường này đã trở thành công cụ quan trọng trong lĩnh vực học máy và phân tích dữ liệu. Google Colab cho phép các nhà nghiên cứu và nhà phát triển làm việc trên cùng một tài liệu, chia sẻ dự án, và sử dụng tài nguyên tính toán mạnh mẽ, bao gồm cả tài nguyên GPU và TPU, mà không cần phải cài đặt bất kỳ phần mềm nào trên máy tính cá nhân.

### 2.2. Ngôn ngữ sử dụng: Python

Python là một trong những ngôn ngữ lập trình phổ biến nhất trên thế giới, và nó đã thống trị lĩnh vực học máy và phân tích dữ liệu. Sự phổ biến của Python đến từ cú pháp đơn giản, cộng đồng lớn mạnh, và các thư viện phong phú như NumPy, pandas, scikit-learn, và TensorFlow, giúp người dùng thực hiện các tác vụ phức tạp một cách dễ dàng.

### 2.3. Cloud : Googe Driver

Google Drive là một dịch vụ lưu trữ đám mây được tích hợp chặt chẽ với Google Colab. Nó cho phép bạn lưu trữ và chia sẻ dự án và dữ liệu dễ dàng, giúp bạn truy cập dữ liệu từ mọi nơi với kết nối internet. Sự kết hợp giữa Google Colab và Google Drive làm cho việc làm việc trên dự án cùng với đồng nghiệp trở nên dễ dàng hơn bao giờ hết.

### 2.4. Facebook Graph API

Facebook Graph API là một tập hợp các giao diện lập trình ứng dụng (API) do Facebook cung cấp, cho phép các nhà phát triển truy cập và tương tác với dữ liệu trên mạng xã hội Facebook. Các ứng dụng sử dụng Facebook Graph API có thể trích xuất thông tin về người dùng, bài viết, trang, nhóm, và nhiều loại dữ liệu khác từ Facebook. Điều này làm cho nó trở thành một công cụ quan trọng trong việc nghiên cứu và phân tích dữ liệu xã hội.

### 2.5. facebook\_scaper

facebook\_scraper là một thư viện Python giúp thu thập dữ liệu từ Facebook thông qua web scraping. Thư viện này cho phép bạn trích xuất thông tin từ bài viết, bình luận, và nhiều thành phần khác trên trang Facebook một cách tự động. Nó cung cấp một phương tiện thuận tiện để thu thập dữ liệu từ Facebook cho các nghiên cứu và dự án liên quan đến mạng xã hội.

### 2.6. Matplotlib

Matplotlib là một thư viện Python mạnh mẽ dùng để vẽ biểu đồ và trực quan hóa dữ liệu. Nó cung cấp các công cụ để tạo các biểu đồ đường, biểu đồ cột, biểu đồ phân tán, biểu đồ hình tròn, và nhiều loại biểu đồ khác. Matplotlib giúp người dùng hiển thị kết quả phân tích dữ liệu một cách trực quan và dễ hiểu.

### 2.7. Mô hình Naive Bayes

Mô hình Naive Bayes là một mô hình học máy dựa trên nguyên tắc xác suất Bayes và giả định "ngây thơ" (naive) rằng tất cả các đặc trưng đầu vào độc lập với nhau. Mô hình này thường được sử dụng trong các tác vụ phân loại và phân tích dữ liệu. Một trong những ứng dụng phổ biến nhất của Naive Bayes là phân loại văn bản, ví dụ như phân loại email thành email thường và email rác (spam).

+ Ưu điểm của Mô hình Naive Bayes:

Dễ triển khai và hiệu quả: Mô hình Naive Bayes dễ triển khai và hoạt động tốt trên các tập dữ liệu lớn. Nó cần ít dữ liệu huấn luyện so với một số mô hình phức tạp khác.

Xử lý dữ liệu văn bản tốt: Naive Bayes thường hiệu quả trong việc phân loại văn bản, ví dụ: phân tích cảm xúc từ các bài đăng trên mạng xã hội, phân loại tin tức, hoặc phát hiện spam email.

Khả năng xử lý đặc trưng lớn: Mô hình Naive Bayes có khả năng xử lý một lượng lớn các đặc trưng (features) mà không gây hiện tượng quá khớp (overfitting).

Không đòi hỏi nhiều tài nguyên tính toán: Mô hình Naive Bayes cần ít tài nguyên tính toán so với một số mô hình phức tạp khác như các mạng nơ-ron sâu (deep neural networks).

+ Nhược điểm của Mô hình Naive Bayes:

Giả định ngây thơ không thực tế: Giả định rằng tất cả các đặc trưng đầu vào độc lập với nhau có thể không phù hợp với thực tế trong một số tình huống. Tuy nhiên, mô hình này vẫn hoạt động tốt trong nhiều trường hợp thực tế.

Hiệu suất có thể bị hạn chế: Mô hình Naive Bayes có thể có hiệu suất thấp hơn so với một số mô hình phức tạp khác, đặc biệt trong trường hợp dữ liệu phức tạp và không tuân theo giả định ngây thơ.

Khả năng xử lý các tương tác phức tạp: Mô hình này không thể mô hình hóa các tương tác phức tạp giữa các đặc trưng. Điều này có nghĩa rằng nó có thể không phù hợp cho các tác vụ yêu cầu mô hình hóa các mối quan hệ phức tạp giữa các biến đầu vào.

## 3. Mô hình

A diagram of a company

Description automatically generated

A diagram of a company

Description automatically generated with medium confidence

## 4. Cài đặt thực nghiệm

### 4.1. Lấy Token có quyền Facebook Graph API

- Tạo một tài khoản <https://developers.facebook.com/>

- Lấy full quyền acess token thong qua ứng dụng. Ở đây chúng ta sẽ dùng thông qua Instagram

A screenshot of a computer

Description automatically generated

- Sao chép đường dẫn sau dán vào trình duyệt: view-source:https://www.facebook.com/dialog/oauth?client\_id=124024574287414&redirect\_uri=https://www.instagram.com/accounts/signup/&&scope=email&response\_type=token



- Sau đó ta được đoạn token có kí tự đầu là “EAAB”

- Copy accessToken như trên sau đó vào link dưới để check quyền:  
<https://developers.facebook.com/tools/debug/accesstoken/>

A screenshot of a computer

Description automatically generated  
  
- Lấy cookies từ accessToken trên

https://graph.facebook.com/me?access\_token=YOUR\_ACCESS\_TOKEN\_HERE

Điền access\_token =

Nhấn F12 và load lại trang.

- Bấm vào tab network -> status: 200 -> Headers -> Request Headers -> CopyCookie  
A screenshot of a computer

Description automatically generated

- Sau bước trên thì ta đã có 2 loại mã là token và cookie.

### 4.2. Lấy dữ liệu comments của bài viết bằng API Graph Facebook

- Thay “PAGE” bằng id page facebook muốn lấy thông tin

- Thay ACCESS\_TOCKEN và COOKIE bằng chuỗi mới lấy được ở phần 4.1

A screenshot of a computer

Description automatically generatedA screen shot of a computer

Description automatically generated

- Kết quả chúng ta có file “XiaomiVietnam.csv”:

A screenshot of a computer

Description automatically generated

- Mở đọc file “XiaomiVietnam.csv”:

A screenshot of a computer

Description automatically generated

- Từ việc đọc file, ta biết cần phải lấy thông tin của trường “message” ra và tổng hợp lại thành một file

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

>> Kết quả chúng ta đã lấy được nội dung comment của người dùng về sản phẩm

### 4.3. Phân tích nội dung comment bằng toán TF-IDF và mô hình Naive Bayes

- Cài đặt pyvi

A screenshot of a computer

Description automatically generated

- Tiền xử lý, làm sạch dữ liệu

A screenshot of a computer code

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated A screenshot of a computer program

Description automatically generated

A screen shot of a computer code

Description automatically generated

Chúng ta có kết quả sau khi gán nhãn:

A close-up of a text

Description automatically generated

- Cài đặt mô hình NB

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated  
A screenshot of a computer program

Description automatically generated

>> Kết quả :

A screenshot of a computer

Description automatically generatedA close up of numbers

Description automatically generated

### 4.4. Lấy thông tin của người comment bài viết

- Tạo một phiên làm việc mới với google colab

from google.colab import drive

drive.mount("/content/drive")

import json

import csv

# Đường dẫn đến tệp JSON trên Google Drive

file\_json\_data = '/content/drive/MyDrive/project\_facebook\_bigdata/all\_data\_file/comments.json'

# Đọc tệp JSON

with open(file\_json\_data, 'r') as file:

   data = json.load(file)

data

A computer screen shot of a computer code

Description automatically generated

Chúng ta sẽ thấy trong dữ liệu comments bài viết sẽ có một document dạng 'from': {'name': 'Nguyễn Hồng Tâm', 'id': '100010678567302'},

Dựa vào id ta sẽ dùng API graph Facebook để lấy thông tin user

# Đọc tệp JSON

with open(file\_json\_data, 'r') as file:

    data = json.load(file)

# Tạo danh sách rỗng để lưu trữ thông tin name và id

names\_and\_ids = []

# Lặp qua các đối tượng trong tệp JSON và trích xuất name và id

for item in data:

    if 'from' in item and 'name' in item['from'] and 'id' in item['from']:

        name = item['from']['name']

        user\_id = item['from']['id']

        names\_and\_ids.append({'name': name, 'id': user\_id})

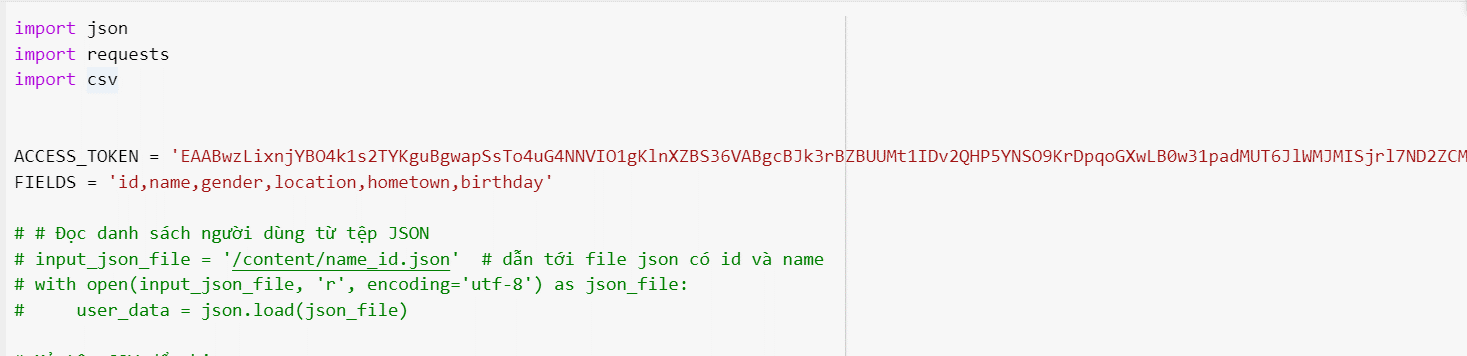
print(len(names\_and\_ids))

names\_and\_ids

A screenshot of a computer

Description automatically generated

- Dùng vòng lặp và sử dụng API Graph facebook để truy cập đến từng Id của người dùng để lấy thống tin



A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

Chúng ta đã lấy được dữ liệu của những user comment bài viết gồm có các trường như gender, location,hometwon,birthday,..

A screenshot of a computer

Description automatically generated

### 4.5. Tiền xử lý dữ liệu thông tin của người comments

- Dữ liệu tuổi

A screenshot of a computer program

Description automatically generated

A computer screen shot of text

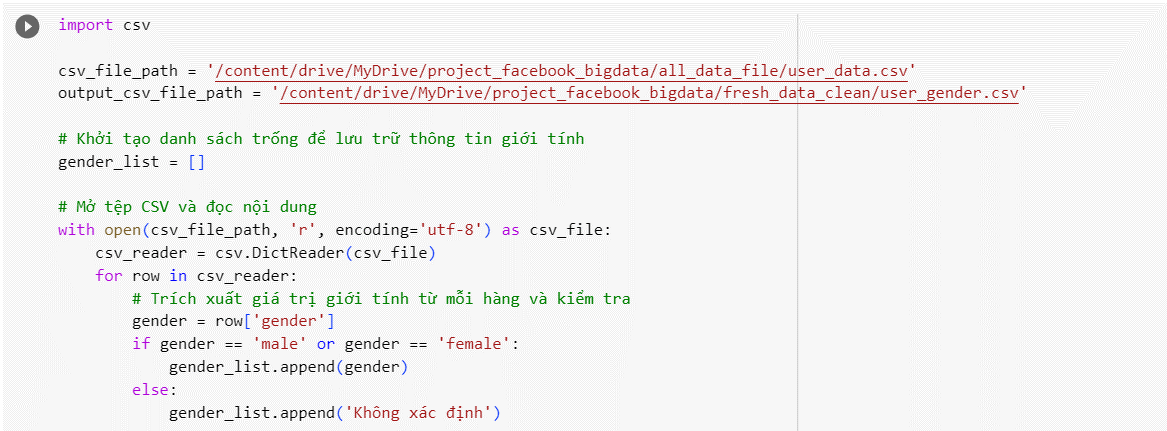
Description automatically generated

>> Kết quả :

A screenshot of a computer

Description automatically generated

- Dữ liệu giới tính



A screenshot of a computer

Description automatically generated

>> Kết quả:

A screenshot of a computer

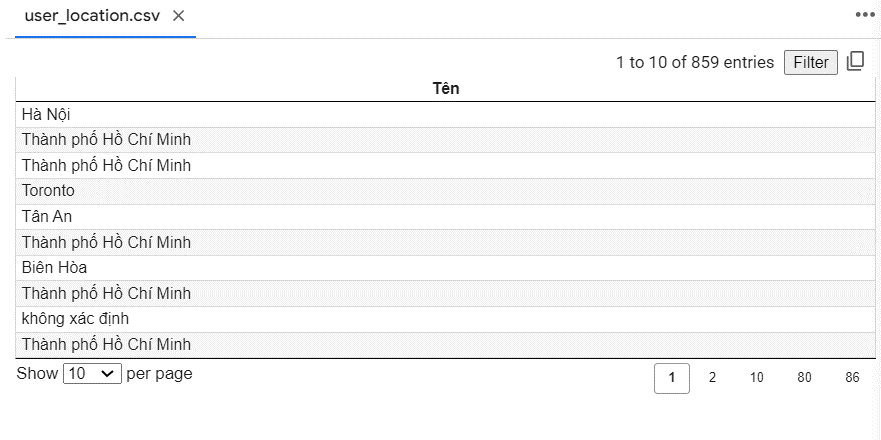
Description automatically generated

- Dữ liệu nơi ở hiện tại:

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

>>Kết quả:

### 4.6. Phân tích thông tin người dùng comments – Biểu diễn biểu đồ

- Import các mục cần thiết A screenshot of a computer

Description automatically generated

#### 4.6.1. Phân tích giới tính người comment

A computer screen shot of text

Description automatically generated

A screen shot of a computer code

Description automatically generated

>> Kết quả:

A blue and pink pie chart

Description automatically generated

#### 4.6.2. Phân tích tuổi người comment

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A white screen with black text

Description automatically generated

>> Kết quả:

A graph with a line

Description automatically generated

#### 4.6.3. Phân tích nơi ở hiện tại (top 10 trong danh sách comment)

A screenshot of a computer code

Description automatically generated

A computer screen shot of text

Description automatically generated

>>Kết quả:

A graph with blue bars

Description automatically generated

### 4.7. Lấy dữ liệu reactions bằng facebook\_scaper

- Tạo một cell làm việc mới tại google colab

A close-up of a white background

Description automatically generated

- Thay giá trị cho POST\_ID bằng id bài post muốn lấy dữ liệu và tương tự hay cookie bằng giá trị lấy ở mục 4.1

A screenshot of a computer

Description automatically generated

-A close-up of a computer screen

Description automatically generated

- Mục tiêu là chúng ta sẽ lấy giá trị reaction và lưu thành 1 file riêng để xử lý.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

### 4.8. Phân tích reactions

- Tại mục 4.7 ta đã lưu data reaction thành 1 file, giờ chúng ta sẽ đọc file đó nên và xử lý phân tích như bên dưới

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

>>Kết quả:

A pie chart with different colored circles

Description automatically generated

## 5. Kết quả và đánh giá

### 5.1. Kết quả

- Phân tích nội dung comments

A screenshot of a computer

Description automatically generated

- Biểu đồ biểu diễn thông tin của người comments bài viết

A blue and pink pie chart

Description automatically generated

A graph with a line

Description automatically generated

A graph with blue bars

Description automatically generated

- Phân tích reactions

A pie chart with different colored circles

Description automatically generated

### 5.2. Đánh giá

- Việc gán nhãn bước phân tích nội dung comments với mô hình Naive Bayes còn tốn thời gian nếu huấn luyện trên tệp tin lớn.

- Dữ liệu về nơi ở của người comment chưa thật sự được làm sạch khi vẫn trùng lặp nhau về địa điểm.

- Google Colab hạn chế kết nối : chỉ cho phép tối đa 3 phiên làm việc kết nối vào google driver cùng một lúc

- Chương trình phải thực hiện tuần tự, không linh động.

## 6. Tài liệu tham khảo

* Liu, B. (2021). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press
* Mitchell, R. (2018). Python Web Scraping. O'Reilly Media
* Holmes, D. E. (2016). Big Data: A Very Short Introduction. Oxford University Press
* McKinney, W. (2017). Python for Data Analysis. O'Reilly Media.
* Russell, M. A. (2015). Mining the Social Web, 3rd Edition. O'Reilly Media