TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN**

**NHẬP MÔN TÍNH TOÁN ĐA PHƯƠNG TIỆN**

**DỰ ĐOÁN CẢM XÚC CHO CÁC BÌNH LUẬN TRÊN YOUTUBE DỰA THEO MÔ HÌNH SENTIMENT ANALYSIS**

*Người hướng dẫn*: **GV. Vũ Đình Hồng**

**GV. Lê Anh Cường**

**GV. Phạm Văn Huy**

*Người thực hiện*: **NGUYỄN HỒ ĐĂNG KHOA – 51800563**

**NGUYỄN LÊ QUỐC BẢO – 51800523**

**NGUYỄN KHÁNH AN – 51800166**

Lớp **: 18050303**

Khoá  **: 22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2022**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN**

**NHẬP MÔN TÍNH TOÁN ĐA PHƯƠNG TIỆN**

**DỰ ĐOÁN CẢM XÚC CHO CÁC BÌNH LUẬN TRÊN YOUTUBE DỰA THEO MÔ HÌNH SENTIMENT ANALYSIS**

*Người hướng dẫn*: **GV. Vũ Đình Hồng**

**GV. Lê Anh Cường**

**GV. Phạm Văn Huy**

*Người thực hiện*: **NGUYỄN HỒ ĐĂNG KHOA – 51800563**

**NGUYỄN LÊ QUỐC BẢO – 51800523**

**NGUYỄN KHÁNH AN – 51800166**

Lớp **: 18050303**

Khoá  **: 22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2022**

LỜI CẢM ƠN

Sau thời gian học tập online vừa qua là những tháng ngày mà chúng em tiếp thu rất nhiều kiến thức của các thầy đã truyền đạt cho chúng em. Những tháng học tập môn “Nhập môn tính toán đa phương tiện” tại Trường đại Tôn Đức Thắng là quãng thời gian quý báu để chúng em học tập và biết thêm nhiều kiến thức chuyên môn có thể áp dụng thực tế cũng như đồ án của chúng em.

Để hoàn thành bài đồ án này trước tiên chúng em xin gửi đến các quý thầy, cô giảng viên trường Đại học Tôn Đức Thắng lời cảm ơn chân thành và sâu sắc nhất. Đặc biệt, em xin gửi đến thầy Vũ Đình Hồng, thầy Lê Anh Cường, thầy Phạm Văn Huy – người đã tận tình hướng dẫn, giúp đỡ em hoàn thành đồ án cuối kì này lời cảm ơn sâu sắc nhất.

Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế của một sinh viên, bài báo cáo này không thể tránh được những thiếu sót. Chúng em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các quý thầy cô để em có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Chúng em xin chân thành cảm ơn!

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của chúng tôi và được sự hướng dẫn của GV Vũ Đình Hồng, GV Lê Anh Cường, GV Phạm Văn Huy. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*\*

*Nguyễn Hồ Đăng Khoa*

*Nguyễn Lê Quốc Bảo*

*Nguyễn Khánh An*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Ngày nay, công nghệ thông tin là một trong những lĩnh vực rất cần thiết đối với cuộc sống của chúng ta. Nó đã trở thành một nhân tố không thể thiếu trong nền kinh tế của nhiều nước trên thế giới, đặc biệt là những nước đang phát triển, trong đó có Việt Nam. Chính vì vậy, công nghệ thông tin là một trong những mũi nhọn của nước ta trong những năm gần đây.

Qua đó, trí tuệ nhân tạo cũng được phát triển song song theo. Trong đó, xử lý ngôn ngữ tự nhiên là một chủ đề đang được chú trọng đến vì tính ứng dụng cao của nó. Phân tích cảm xúc là việc khai thác văn bản theo ngữ cảnh nhằm xác định và trích xuất thông tin chủ quan trong văn bản nguồn, chẳng hạn như nhận ra các cảm xúc (thích thú, sợ hãi, buồn bã, ngạc nhiên…). Hệ thống sẽ tự đánh giá các bình luận và phân tích dữ liệu, sau đó dán nhãn cho các bình luận đó.

# Qua tìm hiểu và những kiến thức đã được dạy về xử lý ngôn tự nhiên, chúng em chọn đề tài: “Dự đoán cảm xúc cho các bình luận trên Youtube dựa theo mô hình Sentiment Analysis”, với model là Multi-layer Perceptron (mạng MLP).

Trang web được xây dựng bằng Django framework, viết bằng ngôn ngữ lập trình Python.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc92455260)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc92455261)

[TÓM TẮT iv](#_Toc92455262)

[MỤC LỤC 1](#_Toc92455263)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 3](#_Toc92455264)

[CHƯƠNG 1 – MỞ ĐẦU 4](#_Toc92455265)

[1.1 Lý do chọn đề tài 4](#_Toc92455266)

[1.2 Cấu trúc báo cáo 4](#_Toc92455267)

[CHƯƠNG 2 – MÔ TẢ BÀI TOÁN 5](#_Toc92455268)

[2.1 Sentiment analysis 5](#_Toc92455269)

[2.1.1 Sentiment analysis là gì? 5](#_Toc92455270)

[2.1.2 Phân tích cảm xúc bằng văn bản 5](#_Toc92455271)

[2.2 Mô tả bài toán 6](#_Toc92455272)

[CHƯƠNG 3 – MÔ TẢ PHƯƠNG PHÁP, THUẬT TOÁN 7](#_Toc92455273)

[3.1 Phương pháp 7](#_Toc92455274)

[3.1.1 Giới thiệu về Django framework 7](#_Toc92455275)

[3.1.1.1 Django framework 7](#_Toc92455276)

[3.1.1.2 Những lợi thế của Django 7](#_Toc92455277)

[3.1.1.3 Cách đặt thư viện Django 8](#_Toc92455278)

[3.1.2 Giới thiệu về Natural Language Processing (NLP) 8](#_Toc92455279)

[3.1.2.1 Khái niệm 8](#_Toc92455280)

[3.1.2.2 Các loại xử lý của NLP 9](#_Toc92455281)

[3.1.2.3 Một số ứng dụng của NLP 9](#_Toc92455282)

[3.1.3 Mạng Multi-layer Perceptron (MLP) 10](#_Toc92455283)

[3.1.3.1 Multilayer Perceptron Classifier – MLP là gì? 10](#_Toc92455284)

[3.1.3.2 Phân biệt Perceptron nhiều lớp (MLP) và hồi quy logistics (Logistic Regression) 11](#_Toc92455285)

[CHƯƠNG 4 - DEMO 13](#_Toc92455286)

[TÀI LIỆU THAM KHẢO 20](#_Toc92455287)

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

[Hình 4.1 Anaconda Prompt (anaconda3) 13](#_Toc92455288)

[Hình 4.2 Trỏ đến thư mục 51800166\_51800563\_51800523 14](#_Toc92455289)

[Hình 4.3 Mở jupyter lab 14](#_Toc92455290)

[Hình 4.4 Project được mở trên Jupyter 15](#_Toc92455291)

[Hình 4.5 Gõ lệnh để chạy server 15](#_Toc92455292)

[Hình 4.6 Server đã sẵn sàng 16](#_Toc92455293)

[Hình 4.7 Giao diện trang web 16](#_Toc92455294)

[Hình 4.8 Nhập link youtube vào ô input 17](#_Toc92455295)

[Hình 4.9 Sau khi đã Crawl Comment 18](#_Toc92455296)

[Hình 4.10 Tổng số Comments đã được phân tích tương ứng với từng nhãn 18](#_Toc92455297)

[Hình 4.11 Comments đã được phân tích và gán nhãn 19](#_Toc92455298)

CHƯƠNG 1 – MỞ ĐẦU

1.1 Lý do chọn đề tài

Cố gắng hiện thực hóa ý tưởng tạo ra một trang web có thể thu thập một số lượng nhận xét (comments) từ một đường link youtube bất kỳ được lựa chọn và dự đoán một cách khách quan cảm xúc (Emotion) của từng nhận xét đươc thu thập, đồng thời thống kê số lượng cảm xúc đã đươc dự đoán trước đó.

Chính vì lẽ đó nên nhóm chúng em chọn đề tài **“Dự đoán cảm xúc cho các bình luận trên Youtube dựa theo mô hình Sentiment Analysis”**, với model là Multi-layer Perceptron (mạng MLP).

1.2 Cấu trúc báo cáo

Cấu trúc báo cáo được chia thành các chương như sau :

* Chương 1: Mở đầu
* Chương 2: Mô tả bài toán
* Chương 3: Mô tả phương pháp, thuật toán
* Chương 4: Demo

Trong quá trình làm bài, chúng em đã cố gắng trình bày một cách đầy đủ, dễ hiểu nhất, đồng thời chúng em đã nỗ lực trong quá trình tìm hiểu tài liệu và viết bài báo cáo này, nhưng không thể tránh khỏi những sai lầm, thiếu sót. Chúng em mong sẽ nhận được những ý kiến đóng góp từ phía thầy cô.

CHƯƠNG 2 – MÔ TẢ BÀI TOÁN

2.1 Sentiment analysis

2.1.1 Sentiment analysis là gì?

Sentiment analysis (Phân tích cảm xúc) là công nghệ được sử dụng để đo lường xúc cảm trong thông điệp truyền tải dựa vào những đặc điểm được lập trình sẵn dựa trên thang điểm mặc định trong hệ thống, có sự tác động của ngữ cảnh, không gian, thời gian,..

2.1.2 Phân tích cảm xúc bằng văn bản

Là chức năng khai thác và phân tích bằng văn bản dựa vào ngữ cảnh nhằm xác định và trích xuất thông tin khách hàng trong hệ CSDL, giúp doanh nghiệp đo lường và đánh giá các phản hồi của người dùng từ các phản hồi bằng ký tự. Nó được sử dụng rộng rãi trong các chiến dịch đo lường sức khỏe thương hiệu dựa trên những biến động của các thông tin từ luồng truyền thông xã hội (Như các ứng dụng Facebook, Linkedin, Google, Báo,…).

Phân tích cảm xúc bằng văn bản giúp doanh nghiệp biết được: Các khía cạnh, vấn đề của sản phẩm/dịch vụ của thương hiệu mà khách hàng quan tâm; ý định, cảm nghĩ của người dùng về vấn đề đó.

Các hình thức phân tích cảm xúc bằng văn bản:

* **Phân tích tình cảm:** Dựa trên các lập trình có khả năng nhận biết đặc điểm càm xúc như: Icon, từ cảm thán,…
* **Phân tích chủ đích:** Phân tích mục đích của đoạn văn bản để xem xét và xác định mục tiêu của khách hàng như: ý thức, tin tức, khiếu nại, đề xuất,..
* **Phân tích theo ngữ cảnh:** Dựa trên định vị trên môi trường mà khách hàng đang tương tác như: cổng thanh toán, forum thảo luận, đánh giá sản phẩm.

2.2 Mô tả bài toán

Thu thâp một số lượng comment từ đường link youtube mà người dùng nhập vào trang web. Dự đoán trang thái cảm xúc của từng comment đươc thu thập và hiện thi lên trang web.

Các trang thái cảm xúc bao gồm:

* Anger (Giận dữ)
* Disgust (Kinh tởm)
* Enjoyment (Thích thú)
* Fear (Sợ hãi)
* Sadness (Buồn bã)
* Surprise (Bất ngờ)
* Other (Khác)

CHƯƠNG 3 – MÔ TẢ PHƯƠNG PHÁP, THUẬT TOÁN

3.1 Phương pháp

3.1.1 Giới thiệu về Django framework

3.1.1.1 Django framework

Django là một framework bậc cao của Python có thể thúc đẩy việc phát triển phần mềm thần tốc và clean, thiết kế thực dụng. Được xây dựng bởi nhiều lập trình viên kinh nghiệm, Django tập trung lớn những vấn đề phát triển Web, bạn có thể phát triển trang web của bạn mà không cần xây dựng từ những căn bản. Đặc biệt nó free và open source.

3.1.1.2 Những lợi thế của Django

* **Hoàn thiện:** Django phát triển theo tư tưởng "Batteries included" (có thể hiểu ý nghĩa là tích hợp toàn bộ, chỉ cần gọi ra mà dùng). Nó cung cấp mọi thứ cho developer không cần phải nghĩ phải dùng cái ngoài. Chúng ta chỉ cần tập trung vào sản phẩm, tất cả đều hoạt động liền mạch với nhau.
* **Đa năng:** Django có thể được dùng để xây dựng hầu hết các loại website, từ hệ thống quản lý nội dung, cho đến các trang mạng xã hội hay web tin tức. Nó có thể làm việc với framework client-side, và chuyển nội dung hầu hết các loại format (HTML, RESS, JSON, XML, ...)
* **Bảo mật:** Django giúp các developer trang các lỗi bảo mật thông thường bằng cách cung cấp framework rằng có những kĩ thuật "phải làm như vậy" để bảo vệ website. Ví dụ: Django cung cấp bảo mật quản lý tên tài khoản và mật khẩu, tránh các lỗi cơ bản như để thông tin session lên cookie, mã hóa mật khẩu thay vì lưu thẳng.
* **Dễ Scale:** Django sử dụng kiến trúc shared-nothing dựa vào component (mỗi phần của kiến trúc sẽ độc lập với nhau, và có thể thay thế hoặc sửa đổi nếu cần thiết). Có sự chia tách rõ ràng giữa các phần nghĩa là nó có thể scale cho việc gia tăng traffic bằng cách thêm phần cứng ở mỗi cấp độ: caching, servers, database servers, hoặc application servers. Nhiều web về kinh doanh đã thành công khi Django được scale đáp ứng yêu cầu của họ.
* **Dễ Maintain:** code Django được viết theo nguyên tắc thiết kế và pattern có thể khuyến khích ý tưởng bảo trì và tái sử dụng code. Trên thực tế, nó sự theo khái niệm Don't Repeat Yourself làm cho không có sự lặp lại không cần thiết, giảm một lượng code.
* **Tính linh động:** Django được viết bằng Python, nó có thể chạy đa nền tảng. Nó có nghĩa rằng bạn không ràng buộc một platform server cụ thể. Django được hỗ trợ tốt ở nhiều nhà cung cấp hosting, họ sẽ cung cấp hạ tầng và tài liệu cụ thể cho hosting web Django.

3.1.1.3 Cách đặt thư viện Django

Ta sẽ cài đặt Django bằng trình lệnh pip của Python. Bây giờ bạn mở **Terminal/CMD** và gõ lệnh sau:

**pip install Django**

3.1.2 Giới thiệu về Natural Language Processing (NLP)

3.1.2.1 Khái niệm

Natural Language Processing (NLP) có thể hiểu là xử lý ngôn ngữ tự nhiên, là một dạng trí tuệ nhân tạo (Artificial intelligence – AI). Nó tập trung vào những tương tác giữa ngôn ngữ tự nhiên của con người và máy tính. Trong trí tuệ nhân tạo thì xử lý ngôn ngữ tự nhiên là một trong những phần khó nhất. Vì nó liên quan đến việc phải hiểu ý nghĩa ngôn ngữ của con người – công cụ hoàn nhất của tư duy và giao tiếp. NLP có khả năng tự động hóa hỗ trợ, nâng cao trải nghiệm của người dùng và phân tích, phản hồi.

Xử lý ngôn ngữ tự nhiên có thể bao gồm xử lý tiếng nói (speech processing) và xử lý văn bản (text processing). Xử lý tiếng nói tập trung nghiên cứu, phát triển các thuật toán, chương trình máy tính xử lý ngôn ngữ của con người ở dạng tiếng nói (dữ liệu âm thanh). Xử lý văn bản tập trung vào phân tích dữ liệu văn bản. Các ứng dụng quan trọng của xử lý văn bản bao gồm tìm kiếm và truy xuất thông tin, dịch máy, tóm tắt văn bản tự động, hay kiểm lỗi chính tả tự động.

Ví dụ về xử lý ngôn ngữ tự nhiên: nhận dạng chữ viết, tiếng nói, dịch tự động, tìm kiếm thông tin, tóm tắt văn bản, khai phá dữ liệu, …

3.1.2.2 Các loại xử lý của NLP

**Phân tích hình thái:** Trong bước này các kí tự không chữ (dấu câu) sẽ được tách ra khỏi các từ. Trong nhiều ngôn ngữ, các từ được phân tác với nhau bằng dấu cách. Tuy nhiên, trong tiếng Việt dấu cách được dùng để phân tách các tiếng (âm tiết) chứ không phải từ.

**Phân tích cú pháp:** Dãy các từ sẽ được biến đổi thành các cấu trúc thể hiện sự liên kết giữa các từ này. Sẽ có những dãy từ bị loại do vi phạm các luật văn phạm.

**Phân tích ngữ nghĩa:** Thêm ngữ nghĩa vào các cấu trúc được tạo ra bởi bộ phân tích cú pháp.

**Tích hợp văn bản:** Ngữ nghĩa của một câu riêng biệt có thể phụ thuộc vào những câu đứng trước, đồng thời nó cũng có thể ảnh hưởng đến các câu phía sau.

**Phân tích thực nghĩa:** Cấu trúc thể hiện được phát ngôn sẽ được thông dịch lại để xác định nó thật sự có nghĩa là gì.

3.1.2.3 Một số ứng dụng của NLP

**Dịch máy (Machine Translation – MT):** Là việc sử dụng máy tính để tự động hóa một phần hoặc toàn bộ quá trình dịch từ ngôn ngữ này sang ngôn ngữ khác. Các phương pháp dịch máy phổ biến bao gồm dịch máy dựa trên ví dụ (example-based machine translation – EBMT), dịch máy dựa trên luật (rule-based machine translation – RBMT), dịch máy thống kê (statistical machine translation – SMT), và dịch máy sử dụng mạng nơ-ron (neural machine translation).

**Phân loại văn bản:** Là quá trình gán các thẻ hoặc danh mục cho văn bản theo nội dung của nó; Đó là một vấn đề cơ bản trong NLP và có thể được thực hiện thủ công (tẻ nhạt, tốn thời gian và dễ bị lỗi của con người) hoặc bằng cách tận dụng các kỹ thuật ML.

**Phân tích cảm xúc:** Là việc khai thác văn bản theo ngữ cảnh nhằm xác định và trích xuất thông tin chủ quan trong văn bản nguồn, chẳng hạn như nhận ra các cực (tích cực, tiêu cực, trung tính), xác định cảm xúc, v.v. Một ví dụ điển hình là trong ngành thương mại điện tử, nơi khai thác và phân tích các bài đánh giá để có được những hiểu biết sâu sắc về sự hài lòng và trải nghiệm của khách hàng, việc xác định các khu vực tiềm năng để cải thiện là rất quan trọng.

**Chatbot:** Là việc chương trình máy tính có khả năng trò chuyện (chat), hỏi đáp với con người qua hình thức hội thoại dưới dạng văn bản (text). Chatbot thường được sử dụng trong ứng dụng hỗ trợ khách hàng, giúp người dùng tìm kiếm thông tin sản phẩm, hoặc giải đáp thắc mắc.

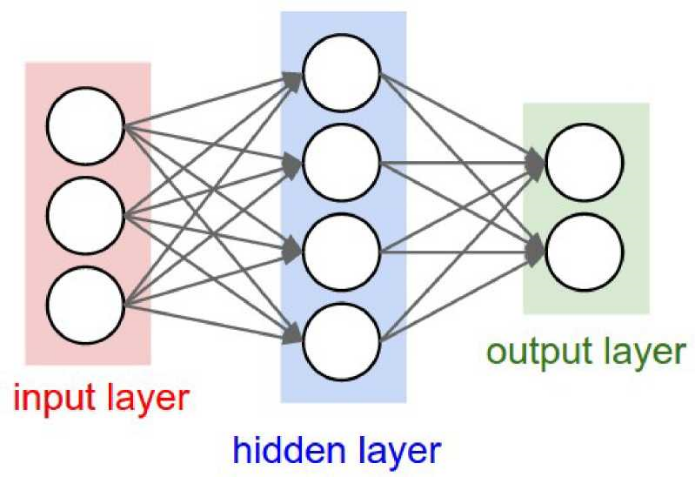
**Nhận dạng tiếng nói (Automatic Speech Recognition – ASR, hoặc Speech To Text – STT):** Chuyển đổi ngôn ngữ từ dạng tiếng nói sang dạng văn bản, thường được ứng dụng trong các chương trình điều khiển qua giọng nói.

3.1.3 Mạng Multi-layer Perceptron (MLP)

3.1.3.1 Multilayer Perceptron Classifier – MLP là gì?

Multilayer Perceptron (Perceptron nhiều lớp) là một thuật toán học máy có giám sát (Machine Learning) thuộc lớp Mạng nơ-ron nhân tạo, là tập hợp của các perceptron chia làm nhiều nhóm, mỗi nhóm tương ứng với một layer. Thuật toán về cơ bản được đào tạo trên dữ liệu để học một hàm phi tuyến tính nhằm mục đích phân loại hay hồi quy, với một tập hợp các tính năng và một biến mục tiêu (ví dụ: nhãn). Bài viết này tập trung vào trường hợp phân loại.

Một Mạng nơ-ron nhân tạo (Artificial Neural Network- ANN) với 3 lớp: Input layer (lớp đầu vào), Output layer (lớp đầu ra) và Hidden layer (lớp ẩn).



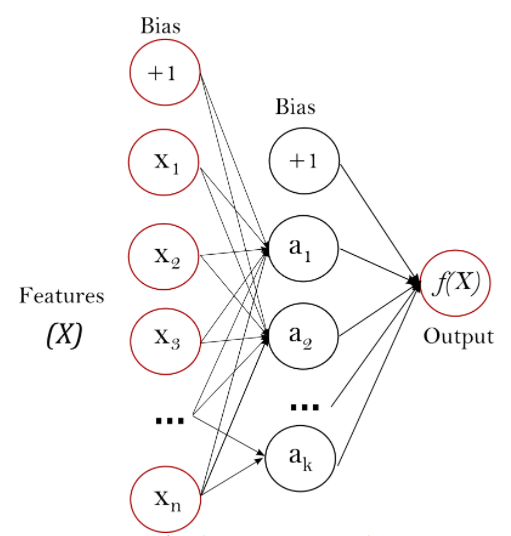
Minh họa Perceptron nhiều lớp

3.1.3.2 Phân biệt Perceptron nhiều lớp (MLP) và hồi quy logistics (Logistic Regression)

Hồi quy logistic chỉ có hai lớp, bao gồm lớp đầu vào (input) và đầu ra (output), còn với MLP, ngoài lớp Input và Output ra thì các lớp nơ-ron ở giữa được gọi chung là lớp Hidden (lớp ẩn). Do tính chất phi tuyến tính này, MLP có thể học các hàm phi tuyến tính phức tạp và phân biệt dữ liệu không thể phân tách tuyến tính.

Tuy nhiên, MLP cũng có 1 số nhược điểm, chẳng hạn như hàm bị phân tách cho lớp ẩn (hidden) dẫn đến vấn đề tối ưu hóa không lồi và tồn tại cực tiểu cục bộ. Vì vậy việc khởi tạo khối lượng đầu vào khác nhau có thể dẫn đến đầu ra với các kết quả khác nhau.

Ngoài ra, MLP sử dụng một số siêu tham số, bao gồm số lượng tế bào nơ-ron ẩn hay thông tin các lớp cần được điều chỉnh, phân tích các siêu tham số này dễ dẫn đến tốn thời gian và nguồn lực. Đặc biệt, MLP có thể nhạy cảm với việc mở rộng tính năng.



Một MLP với một lớp ẩn và với đầu ra vô hướng.

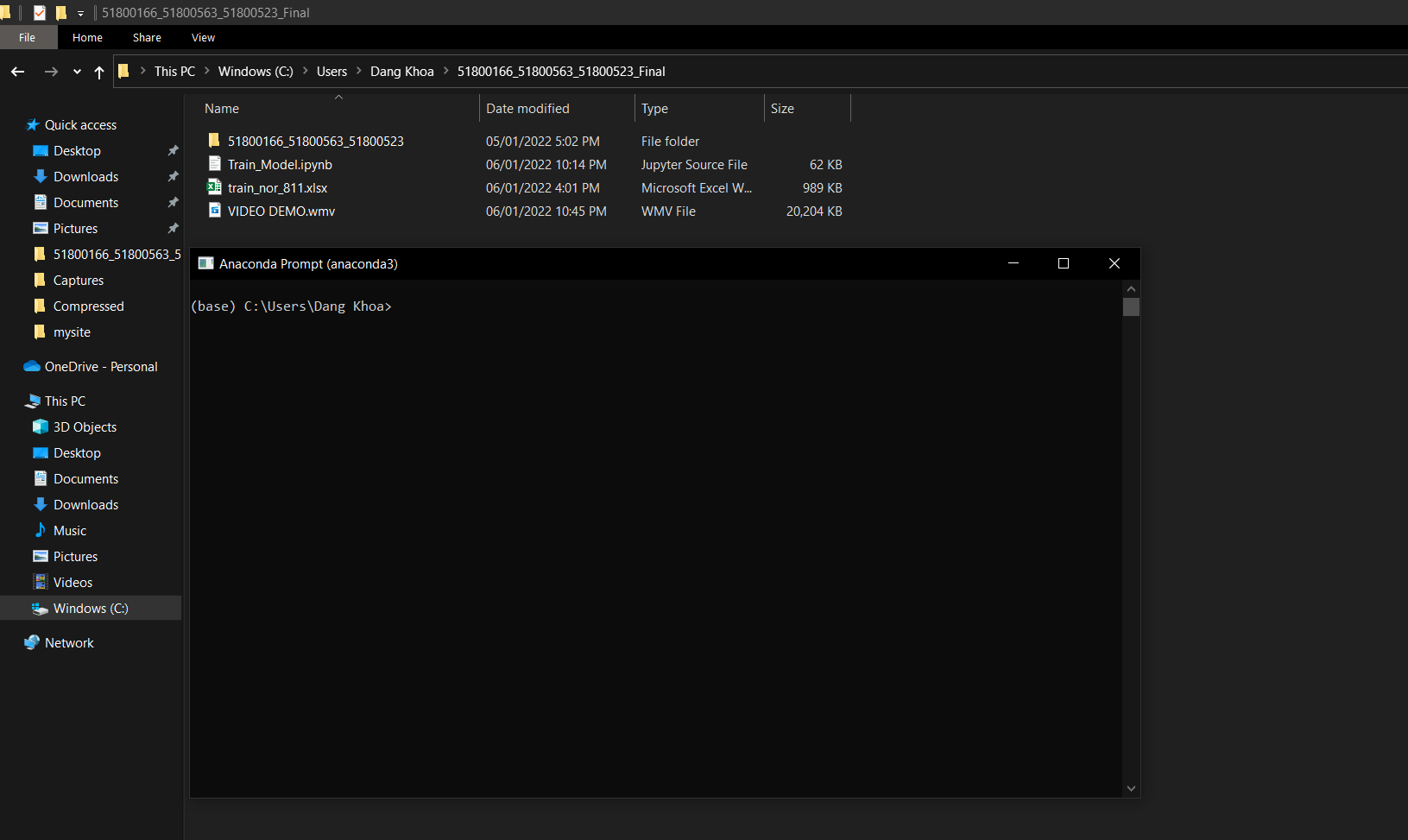
3.1.3.3 Đánh giá mô hình Perceptron nhiều lớp (MLP)

Perceptron nhiều lớp (MLP) là một mô hình mạng nơ-ron rất mạnh mẽ cho phép học các hàm phi tuyến tính đối với dữ liệu phức tạp.

Phương pháp này đặc biệt phù hợp với các tài liệu chữ, số viết tay, hỗ trợ phân loại và trích xuất nhanh chóng.

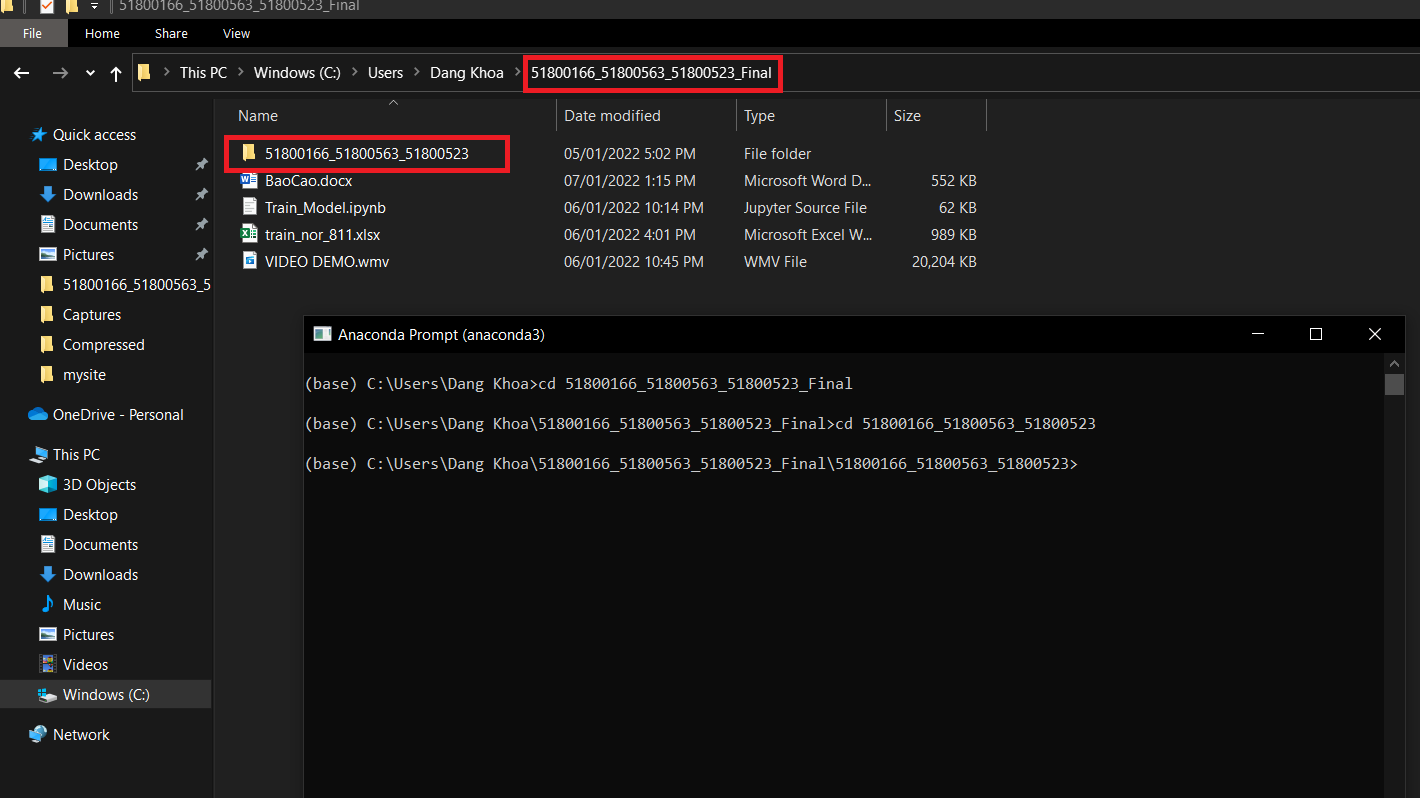
CHƯƠNG 4 - DEMO

Để có thể chạy project, đầu tiên ta mở Anaconda Prompt (anaconda3)



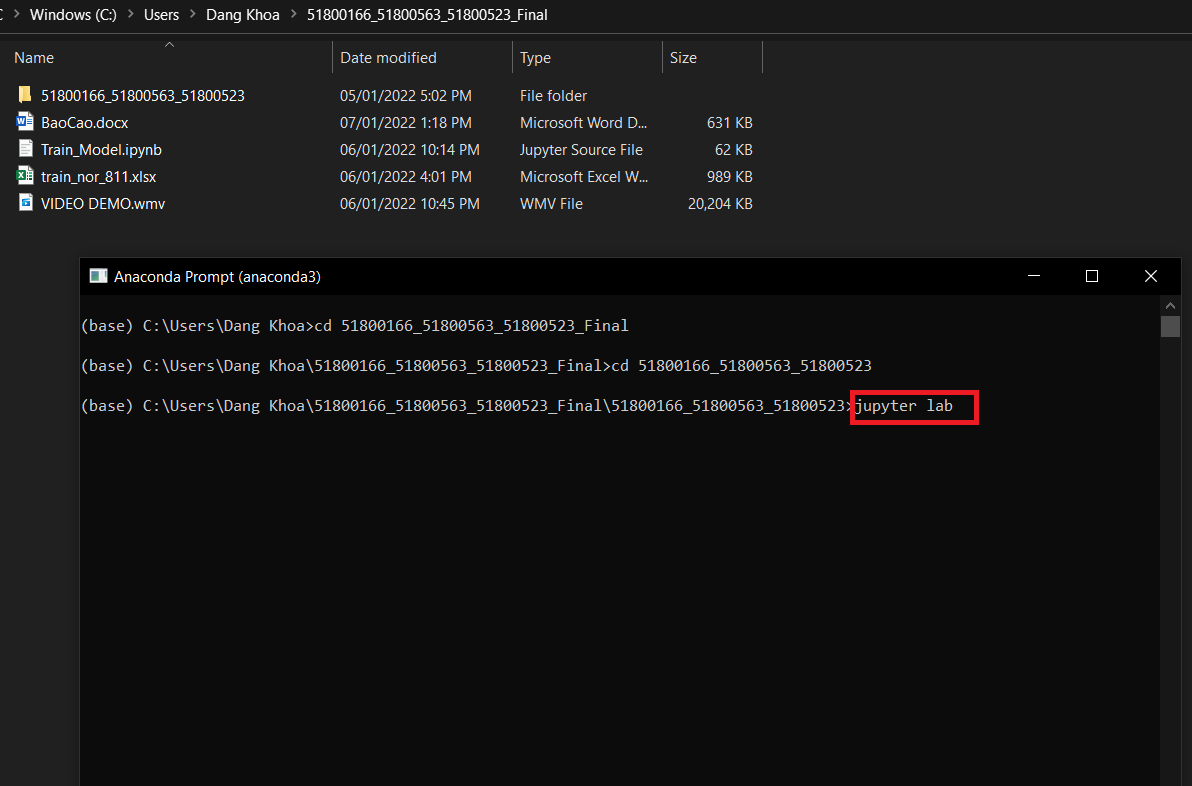
Hình 4.1 Anaconda Prompt (anaconda3)

Trong CMD của Anaconda Prompt (anaconda3) thực hiện trỏ thư mục đến thư mục 51800166\_51800563\_51800523 chính là thư mục project WEB

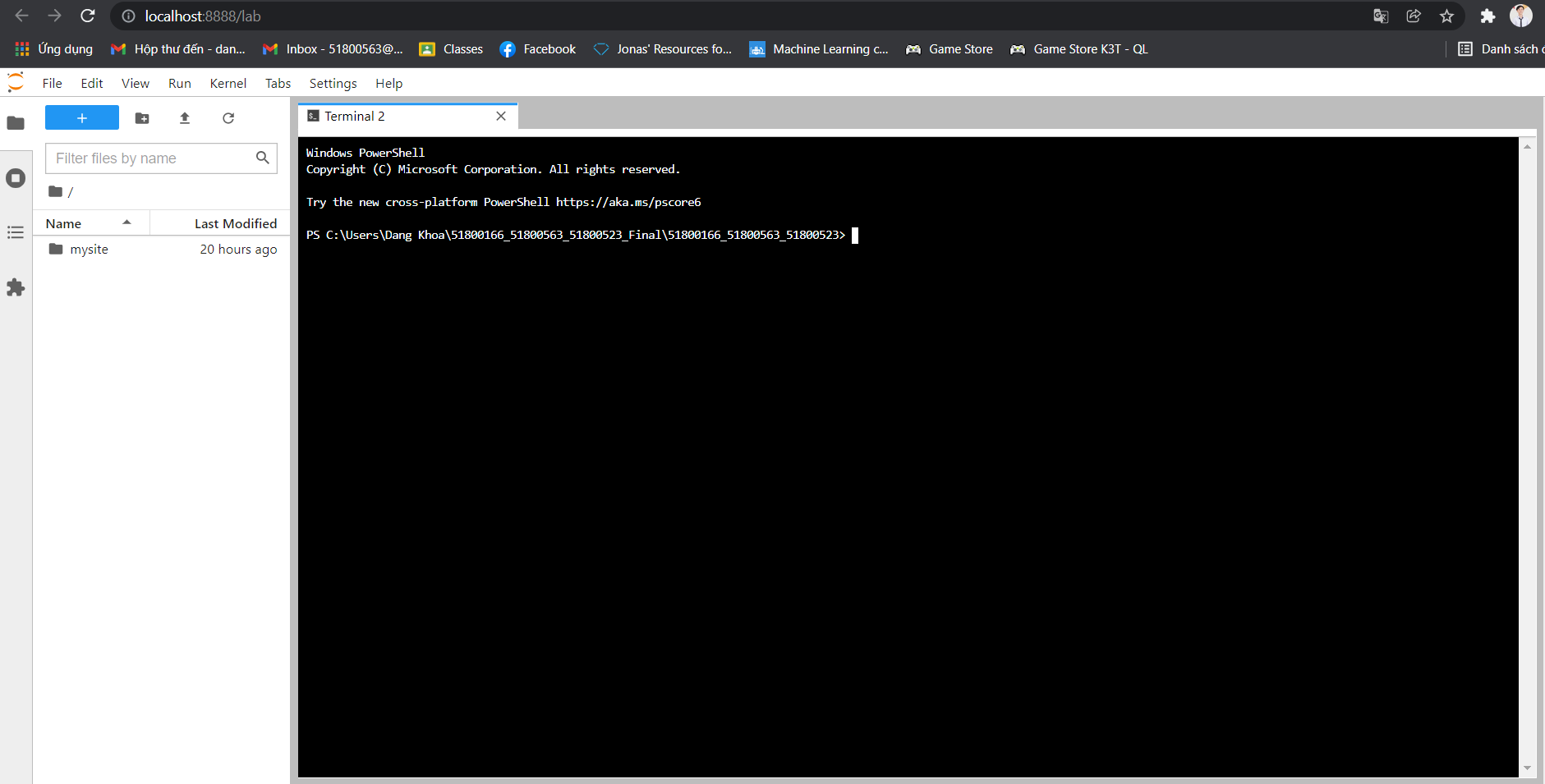


Hình 4.2 Trỏ đến thư mục 51800166\_51800563\_51800523

Tiếp tục trong CMD, gõ jupyter lab để chạy project

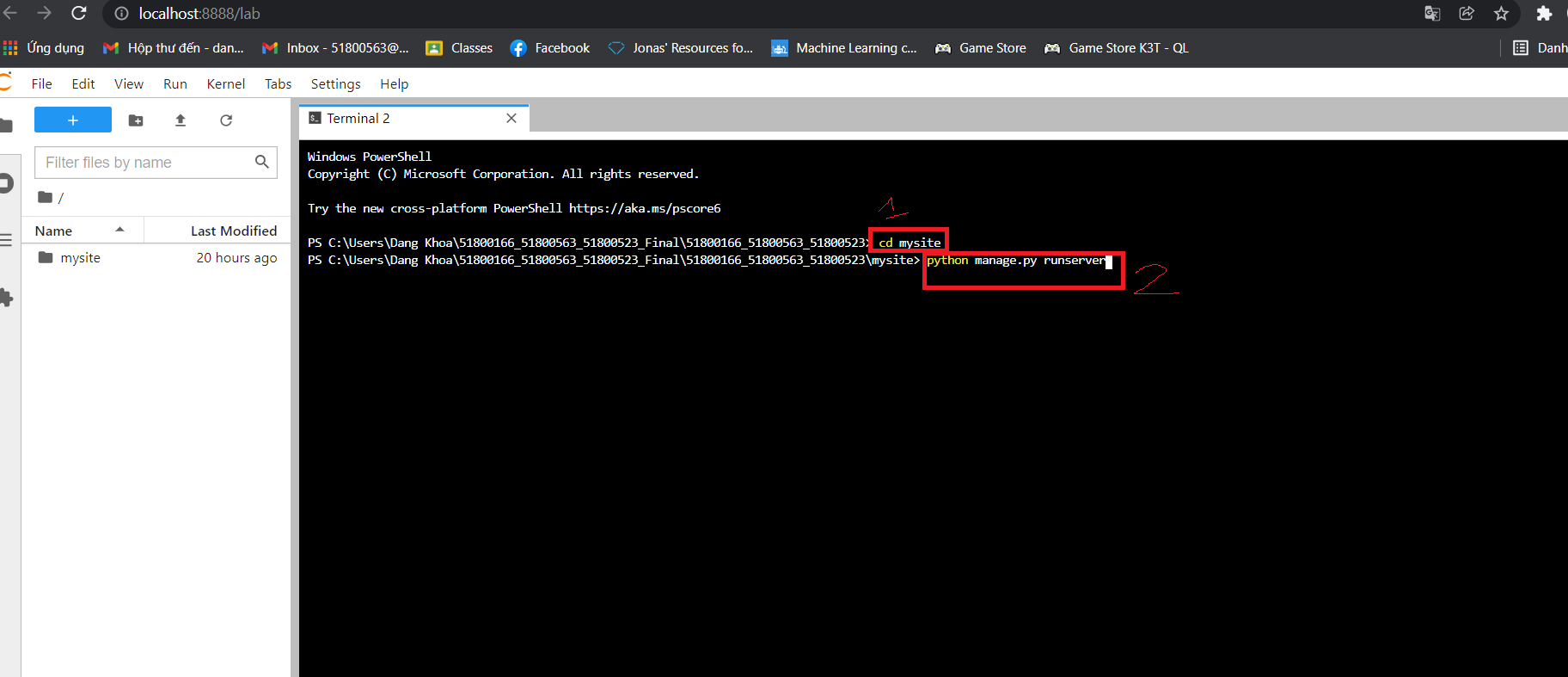


Hình 4.3 Mở jupyter lab

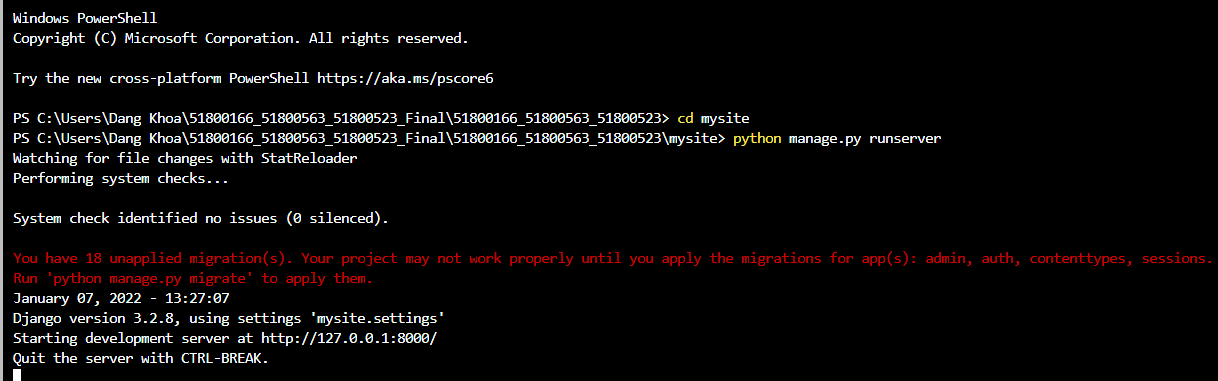


Hình 4.4 Project được mở trên Jupyter

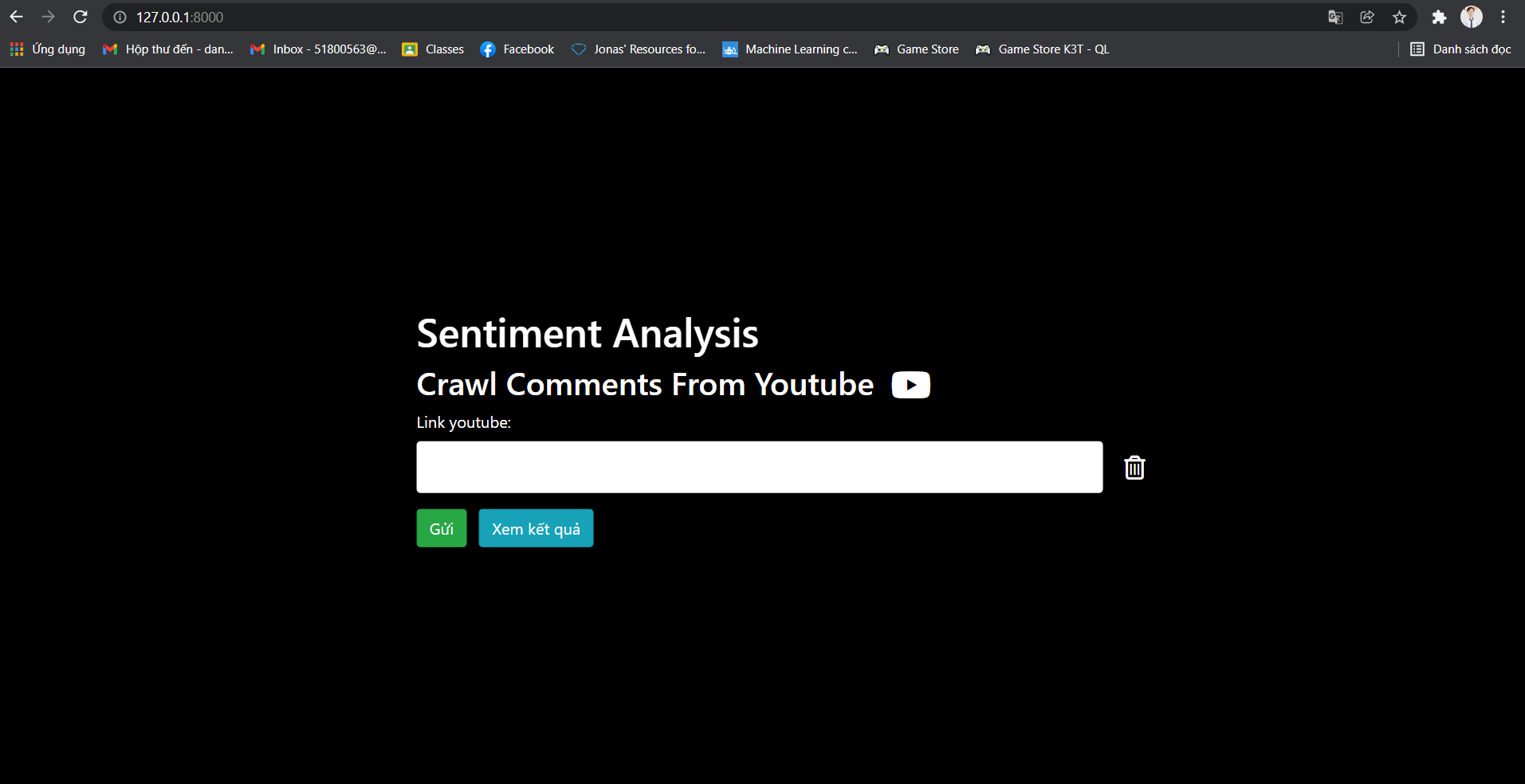
Chuyển đến thư mục mysite và gõ python manage.py runserver để chạy



Hình 4.5 Gõ lệnh để chạy server

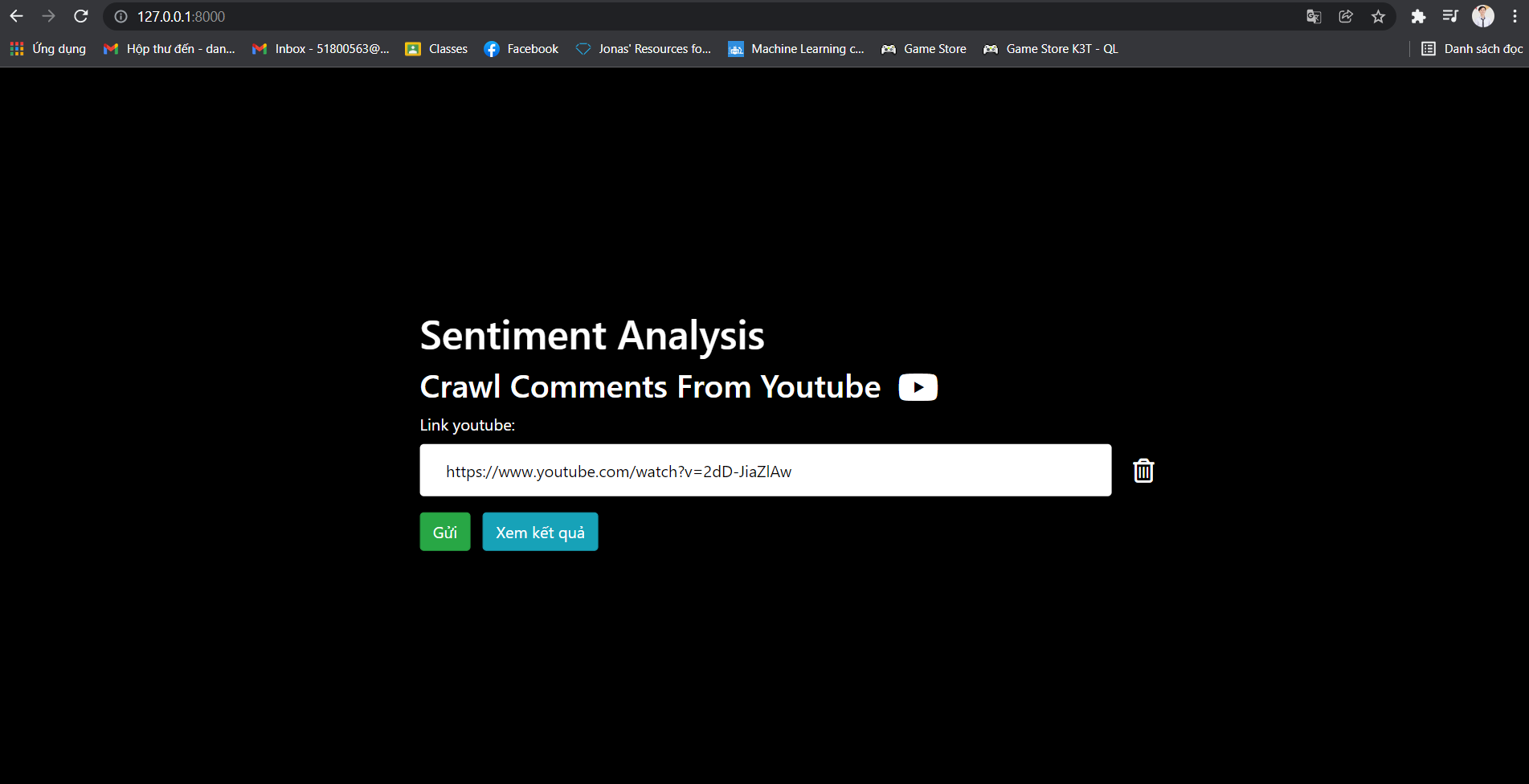


Hình 4.6 Server đã sẵn sàng



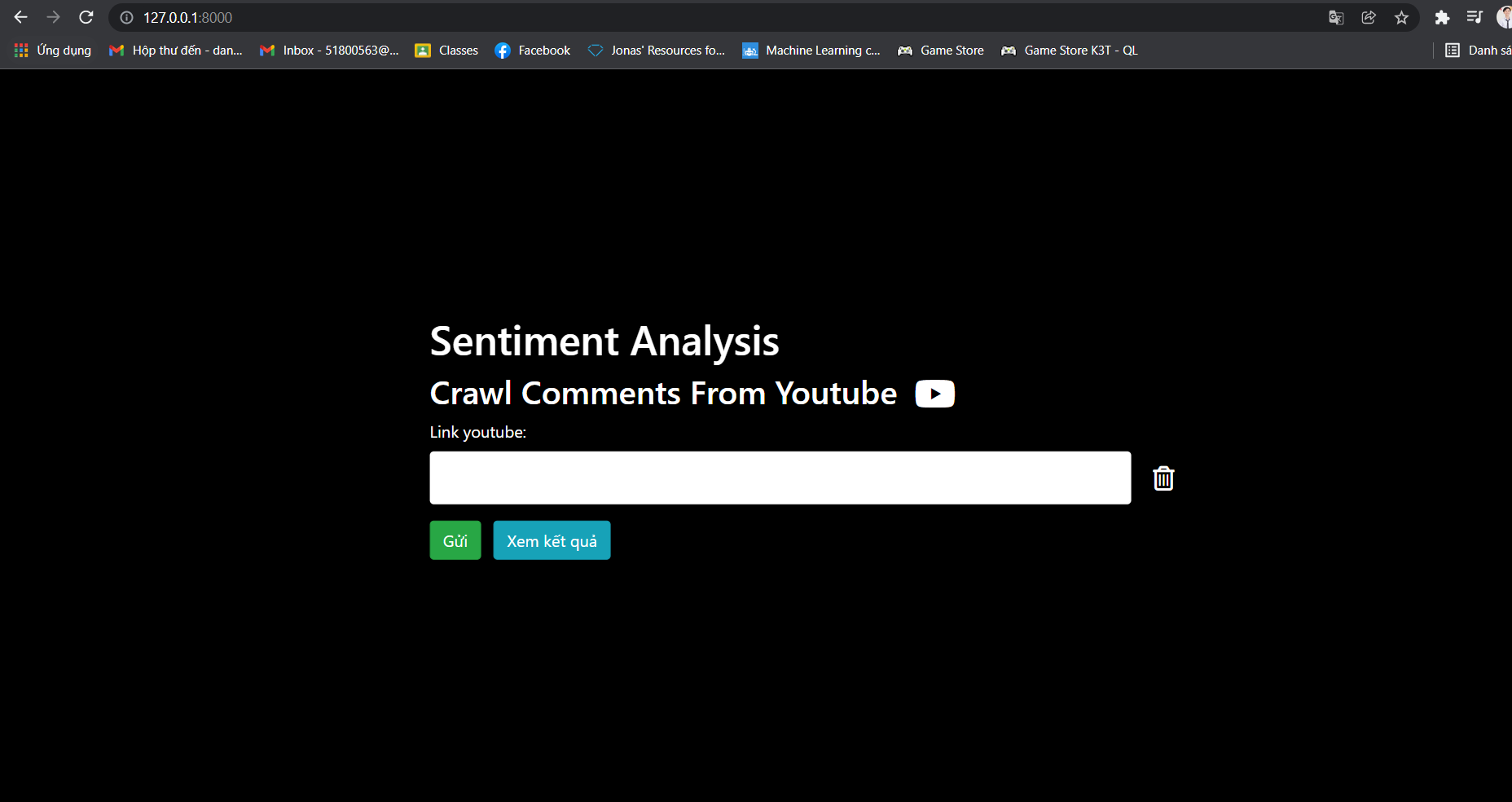
Hình 4.7 Giao diện trang web

Nhập link youtube vào ô input



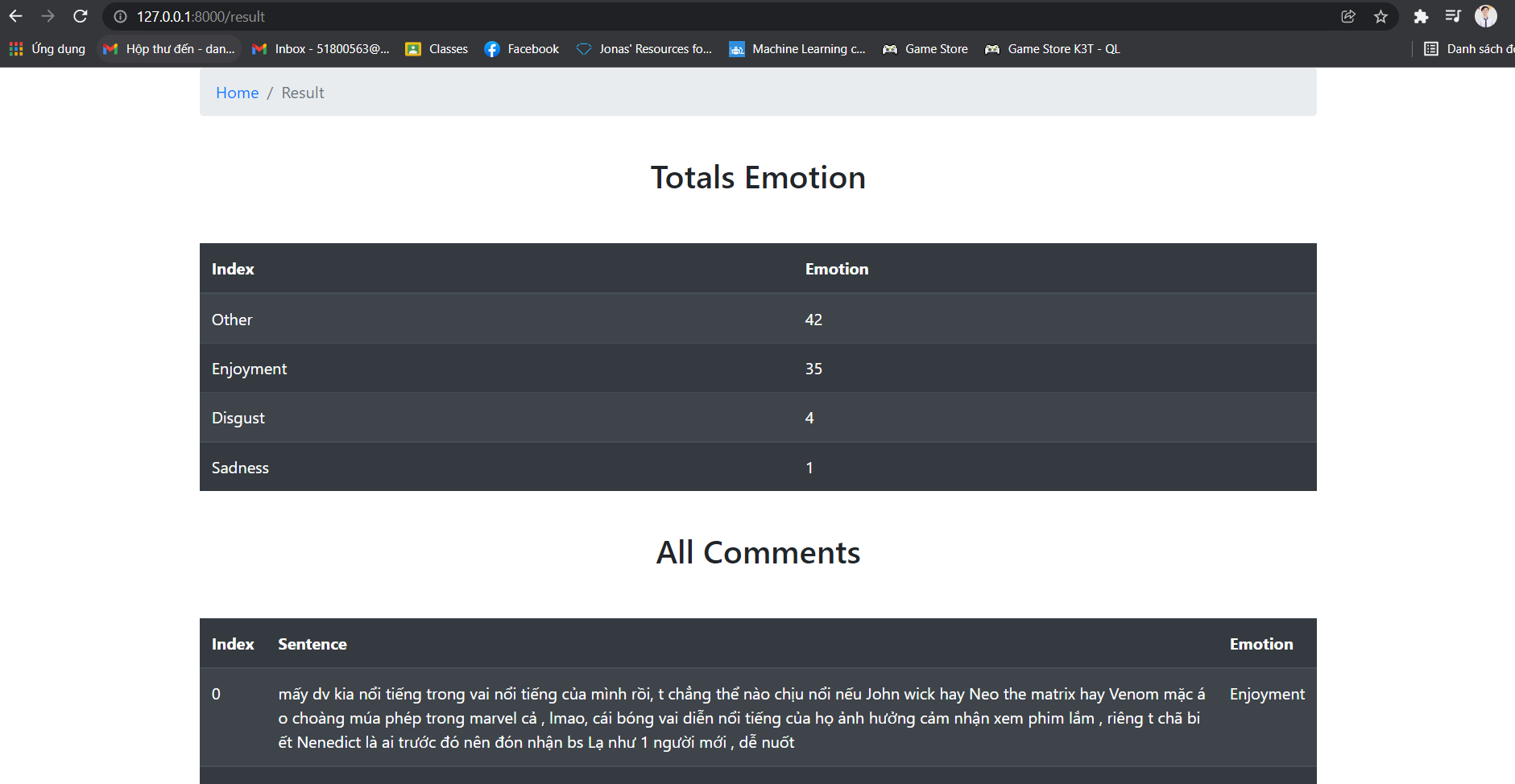
Hình 4.8 Nhập link youtube vào ô input

Nhấn **Gửi** để Crawl Comments

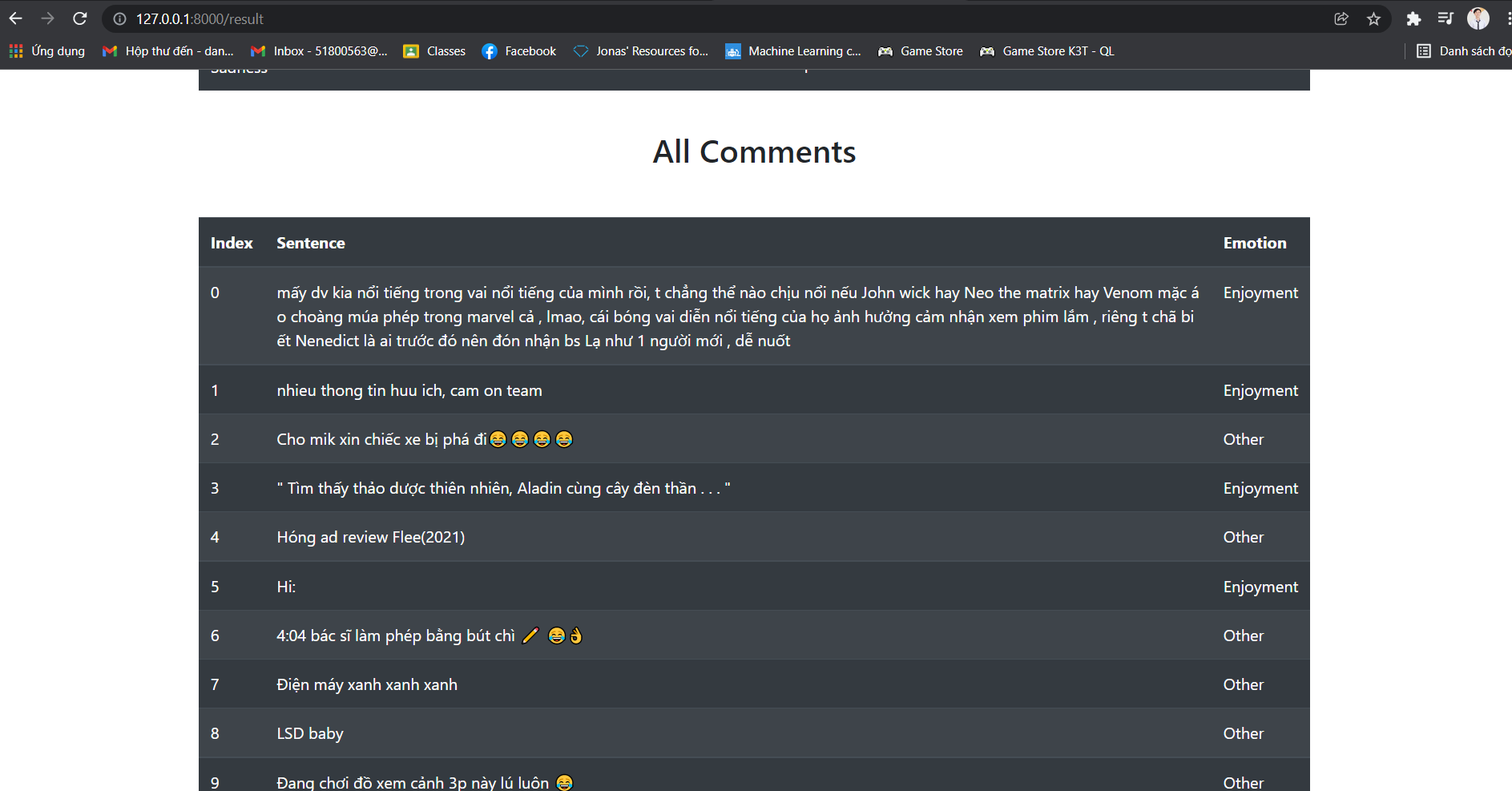


Hình 4.9 Sau khi đã Crawl Comment

Nhấn **Xem kết quả** và ta thấy tất cả Comment đã được đánh giá và gán nhãn



Hình 4.10 Tổng số Comments đã được phân tích tương ứng với từng nhãn



Hình 4.11 Comments đã được phân tích và gán nhãn

TÀI LIỆU THAM KHẢO

1. https[://www.digitalocean.com/community/tech\_talks/build-a-web-app-with-django](https://www.digitalocean.com/community/tech_talks/build-a-web-app-with-django)
2. https://product.vinbigdata.org/phan-loai-chu-so-viet-tay-voi-perceptron-nhieu-lop-mlp
3. https://paydayloanssqa.com/sentiment-analysis-la-gi
4. https://www.mindalife.vn/nlp