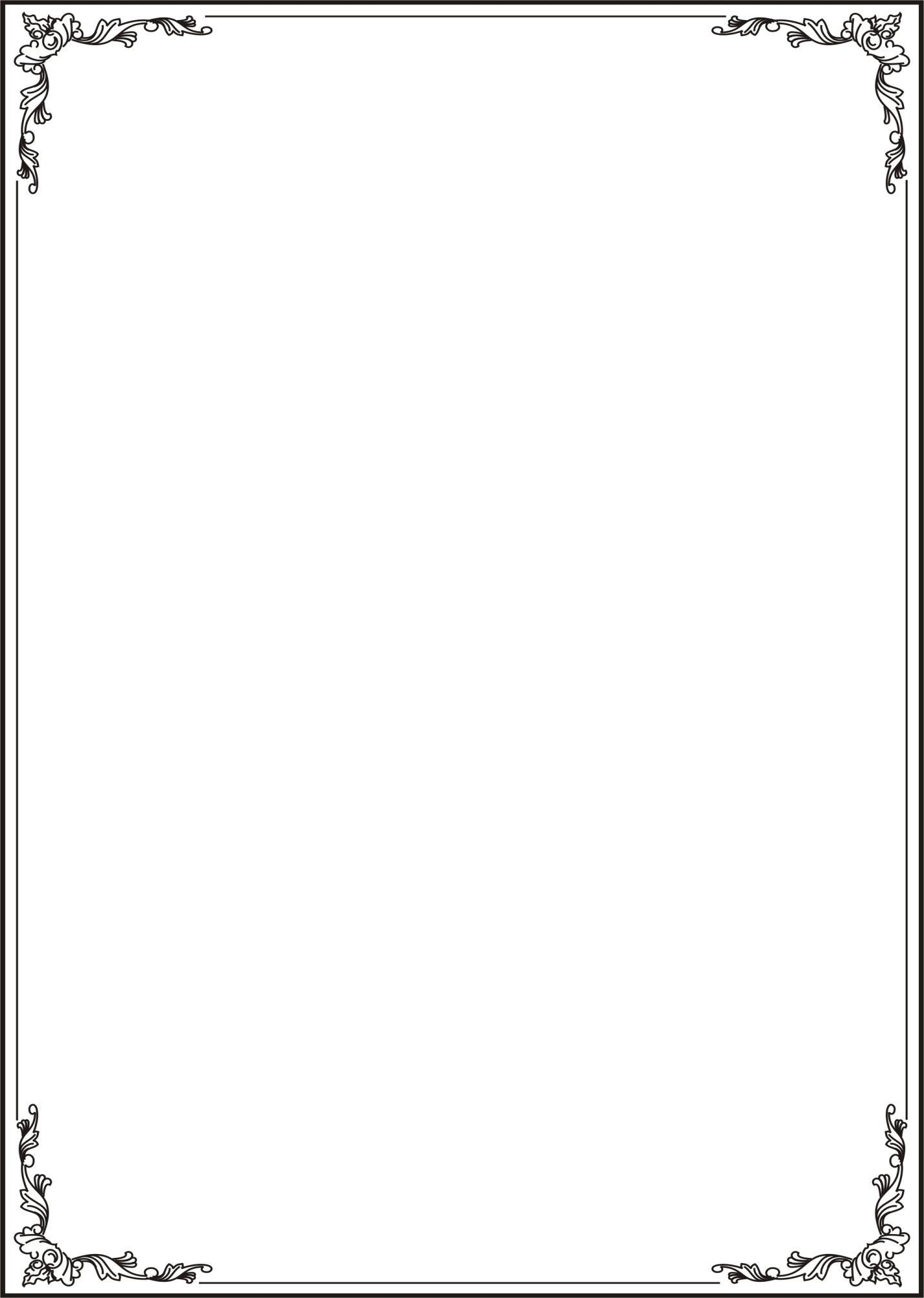
**.**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**

**BỘ MÔN ĐIỆN TOÁN ĐÁM MÂY**

-----🙞🙜🕮🙞🙜-----

****

**XÂY DỰNG ỨNG DỤNG NHẬN DIỆN TÊN TRONG CHỨNG MINH NHÂN DÂN ỨNG DỤNG AWS SAGEMAKER**

Khóa 2019-2022

**GVHD: TS. HUỲNH XUÂN PHỤNG**

**SVTH: MSSV:**

**THÁI HOÀNG PHÚC 20110543**

**PHẠM NGUYỄN PHÚ QUÍ 20110079**

Khóa 2019 -2023

**TP.HCM 12/2022**

**MỤC LỤC**

[LỜI CẢM ƠN 3](#_Toc121498554)

[MỞ ĐẦU 4](#_Toc121498555)

[**1.** **Lí do chọn đề tài** 4](#_Toc121498556)

[**2.** **Đối tượng nghiên cứu** 4](#_Toc121498557)

[**3.** **Phạm vi nghiên cứu** 4](#_Toc121498558)

[**4.** **Kết quả dự kiến** 4](#_Toc121498559)

[NỘI DUNG 5](#_Toc121498560)

[Chương 1: CƠ SỞ LÝ THUYẾT 5](#_Toc121498561)

[**1.** **AWS SageMaker** 5](#_Toc121498562)

[**1.1.** **Khái quát** 5](#_Toc121498563)

[**1.2.** **Lợi ích và tính năng** 6](#_Toc121498564)

[**1.3.** **Amazon SageMaker Studio Lab** 6](#_Toc121498565)

[**2.** **Amazon S3 & Amazon EC2** 7](#_Toc121498566)

[Chương 2: XÂY DỰNG ỨNG DỤNG 9](#_Toc121498567)

[**1.** **Giới thiệu website** 9](#_Toc121498568)

[**2.** **Cài đặt** 9](#_Toc121498569)

[**2.1.** **Cài đặt AWS S3** 9](#_Toc121498570)

[**2.2.** **Cài đặt AWS EC2** 11](#_Toc121498571)

[**2.3.** **Cài đặt AWS SageMaker** 13](#_Toc121498572)

[**3.** **Chức năng thực hiện được** 13](#_Toc121498573)

[KẾT LUẬN 16](#_Toc121498574)

# **LỜI CẢM ƠN**

Đầu tiên, em xin gửi lời cảm ơn chân thành đến Trường Đại học Sư phạm kỹ thuật thành phố Hồ Chí Minh đã đưa môn học Điện toán đám mây vào chương trình giảng dạy. Đặc biệt, em xin gửi lời cảm ơn sâu sắc đến giảng viên bộ môn - thầy Huỳnh Xuân Phụng đã dạy dỗ, truyền đạt những kiến thức quý báu cho em trong suốt thời gian học tập vừa qua. Đây chắc chắn sẽ là những kiến thức quý báu, là hành trang để em có thể vững bước sau này.

Bộ môn Điện toán đám mây là môn học thú vị, vô cùng bổ ích và có tính thực tế cao. Tuy nhiên, do vốn kiến thức còn nhiều hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ. Mặc dù em đã cố gắng hết sức nhưng chắc chắn báo cáo khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong thầy xem xét và góp ý để bài báo cáo của em được hoàn thiện hơn.

Em xin chân thành cảm ơn!

# **MỞ ĐẦU**

1. **Lí do chọn đề tài**

Hằng ngày văn thư tại các cơ quan nhà nước, doanh nghiệp nhận được rất nhiều loại văn bản cần phải nhập lại vào phần mềm. Tuy nhiên việc gõ văn bản với nhiều form khác nhau được thực hiện thủ công tốn nhiều thời gian công sức nhưng vẫn tiềm ẩn nguy cơ sai sót. Với những thực trạng trên, phần mềm nhận dạng chữ viết tay cũng như các phần mềm trích xuất hình ảnh đã ra đời và giải quyết.

Amazon SageMaker là dịch vụ machine learning dạng mô-đun được quản lý toàn phần, cho phép các nhà phát triển và nhà khoa học dữ liệu xây dựng, đào tạo và triển khai các mô hình machine learning với quy mô bất kỳ.

Việc triển khai mô hình machine learning từ lý thuyết vào giai đoạn sản xuất thường phức tạp và cần nhiều thời gian. Cần phải quản lý lượng lớn dữ liệu để đào tạo mô hình, chọn thuật toán tốt nhất để đào tạo mô hình, quản lý năng lực điện toán trong khi đào tạo, cuối cùng triển khai mô hình đó vào môi trường sản xuất. Amazon SageMaker giảm thiểu sự phức tạp này bằng cách làm cho việc xây dựng và triển khai các mô hình ML dễ dàng hơn nhiều. Sau khi bạn chọn thuật toán và framework phù hợp từ hàng loạt lựa chọn có sẵn, dịch vụ này quản lý tất cả cơ sở hạ tầng nền tảng để đào tạo mô hình của bạn ở quy mô petabyte và triển khai mô hình đó vào giai đoạn sản xuất.

Vì thế nhóm quyết định lựa chọn SageMaker để hỗ trợ thực hiện ứng dụng nhận diện tên trong chứng minh nhân dân trên web.

1. **Đối tượng nghiên cứu**

Đối với đề tài này, đối tượng nghiên cứu là SageMaker. Đồng thời kết hợp với các kiến thức về Front-end và Back-end để tạo một trang web cụ thể.

1. **Phạm vi nghiên cứu**

Đề tài nghiên cứu chủ yếu về SageMaker cùng với sự kết hợp của S3 và EC2 để tạo trang web hỗ trợ nhận diện văn bản thông qua hình ảnh.

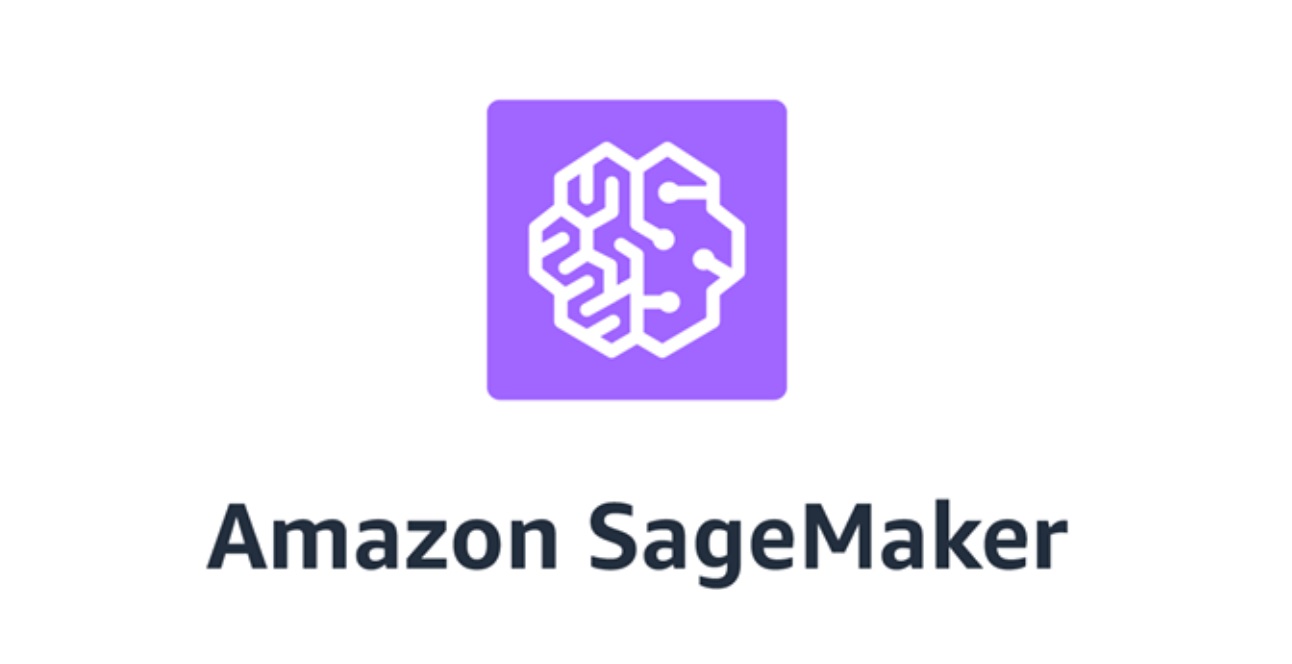
1. **Kết quả dự kiến**

Trang web có chức năng nhận diện được văn bản có trong hình ảnh, cụ thể là tên trong chứng minh nhân dân.

# **NỘI DUNG**

# **Chương 1: CƠ SỞ LÝ THUYẾT**

1. **AWS SageMaker**
   1. **Khái quát**



Amazon SageMaker là một dịch vụ được quản lý trong đám mây công cộng Amazon Web Services (AWS). Nó cung cấp các công cụ để xây dựng, đào tạo và triển khai các mô hình học máy (ML) cho các ứng dụng phân tích dự đoán. Nền tảng này tự động hóa công việc tẻ nhạt trong việc xây dựng hệ thống trí tuệ nhân tạo (AI) sẵn sàng đưa vào sản xuất.

Amazon SageMaker cho phép nhiều đối tượng người dùng tạo ra các mô hình ML, bao gồm các môi trường không lập trình trong Amazon SageMaker Canvas dành cho các nhà phân tích kinh doanh không có kinh nghiệm về ML, cũng như môi trường ML không cài đặt, miễn phí để sinh viên học tập và thử nghiệm với ML nhanh chóng hơn.

Với Amazon SageMaker, khách hàng có thể giảm thời gian huấn luyện từ hàng giờ xuống còn vài phút với cơ sở hạ tầng được tối ưu hóa. Cuối cùng, khách hàng có thể tự động hóa và tiêu chuẩn hóa các thực hành vận hành học máy (MLOps) trong tổ chức để xây dựng, huấn luyện, triển khai và quản lý các mô hình trên quy mô lớn.

Amazon SageMaker được xem là giải pháp phần mềm có chức năng và chi phí phù hợp cho các doanh nghiệp từ nhỏ và vừa (SMEs) tới các doanh nghiệp lớn. Phần mềm Amazon SageMaker được đánh giá cao bởi cả người dùng lẫn chuyên gia trong lĩnh vực AI Platforms Software.

* 1. **Lợi ích và tính năng**
* *Labeling raw data with active learning*

Amazon SageMaker Ground Truth sử dụng mô hình máy học để tự động cố gắn nhãn cho dữ liệu đào tạo. Bất kỳ dữ liệu nào không thể được mô hình dán nhãn một cách tự tin sẽ tự động được gửi đến người dán nhãn. Dữ liệu do con người dán nhãn được cung cấp lại cho mô hình để mô hình có thể liên tục học hỏi và cải thiện.

* *Highly accurate training datasets*

Các mô hình học tập tích cực từ Amazon SageMaker Ground Truth mang lại độ nhất quán và độ chính xác rất cao cho tập dữ liệu đào tạo mà không phải chịu gánh nặng kiểm tra và yêu cầu cùng một dữ liệu được gắn nhãn nhiều lần để loại bỏ các giá trị ngoại lệ.

* *Fully-managed notebook instances*

Đối với quá trình tiền xử lý và khám phá dữ liệu đào tạo, Amazon SageMaker cung cấp các phiên bản được quản lý hoàn toàn chạy sổ ghi chép Jupyter bao gồm mã ví dụ cho các bài tập lưu trữ và đào tạo mô hình phổ biến.

* *Highly-optimized machine learning algorithms*

Amazon SageMaker cài đặt các thuật toán machine learning hiệu suất cao, có thể thay đổi quy mô được tối ưu hóa về tốc độ, quy mô và độ chính xác để chạy trên tập dữ liệu đào tạo cực lớn.

* *One-click training*

Khi bạn đã sẵn sàng đào tạo trong Amazon SageMaker, chỉ cần cho biết loại và số lượng phiên bản bạn cần và bắt đầu đào tạo chỉ bằng một cú nhấp chuột.

* *Deployment without engineering effort*

Sau khi đào tạo, SageMaker cung cấp các tạo phẩm mô hình và hình ảnh chấm điểm cho bạn để triển khai lên Amazon EC2 hoặc bất kỳ nơi nào khác.

* 1. **Amazon SageMaker Studio Lab**



Amazon SageMaker Studio Lab là một ứng dụng web trực tuyến miễn phí để học và thử nghiệm khoa học dữ liệu và máy học bằng cách sử dụng sổ ghi chép Jupyter.

1. **Amazon S3 & Amazon EC2**

* *Amazon S3*

Một dịch vụ lưu trữ đối tượng cung cấp khả năng mở rộng, tính sẵn có của dữ liệu, bảo mật và hiệu suất hàng đầu trong ngành.

Được sử dụng nhằm ngừng sử dụng cơ sở hạ tầng vật lý và lưu trữ dữ liệu với các lớp lưu trữ S3 giúp lưu giữ các đối tượng lâu dài ở mức giá thấp nhất.

* *Amazon EC2*

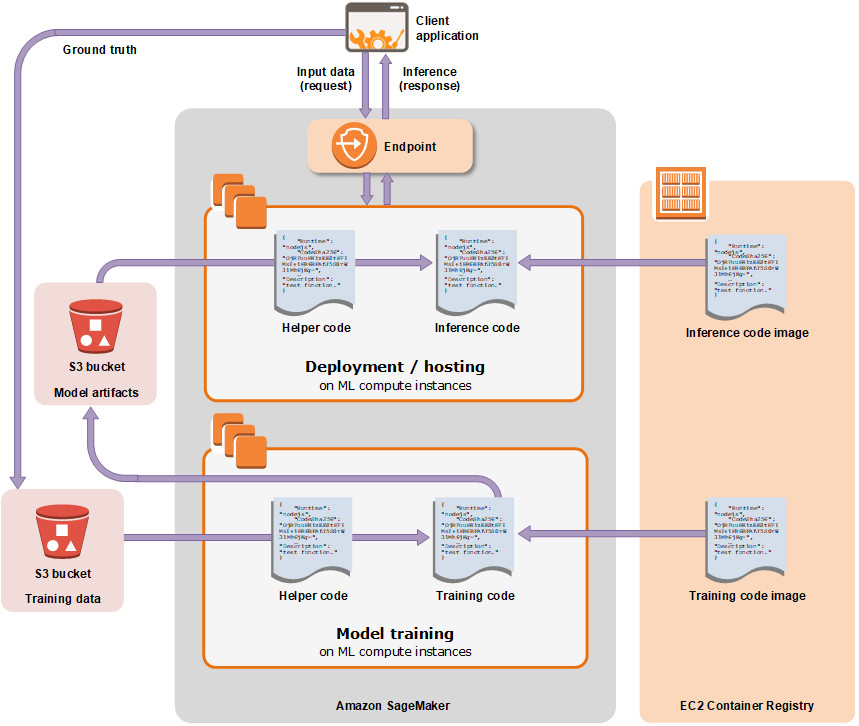
Mang đến nền tảng điện toán sâu rộng nhất, cùng với đó là hơn 500 phiên bản và tuyển tập bộ xử lý, dung lượng lưu trữ, kết nối mạng, hệ điều hành và mô hình mua hàng mới nhất để giúp bạn đáp ứng tốt nhất những nhu cầu về khối lượng công việc của mình.

Nhằm cung cấp cơ sở hạ tầng điện toán bảo mật, đáng tin cậy, hiệu năng cao và tiết kiệm chi phí nhằm đáp ứng các nhu cầu kinh doanh khắt khe.

# **Chương 2: XÂY DỰNG ỨNG DỤNG**

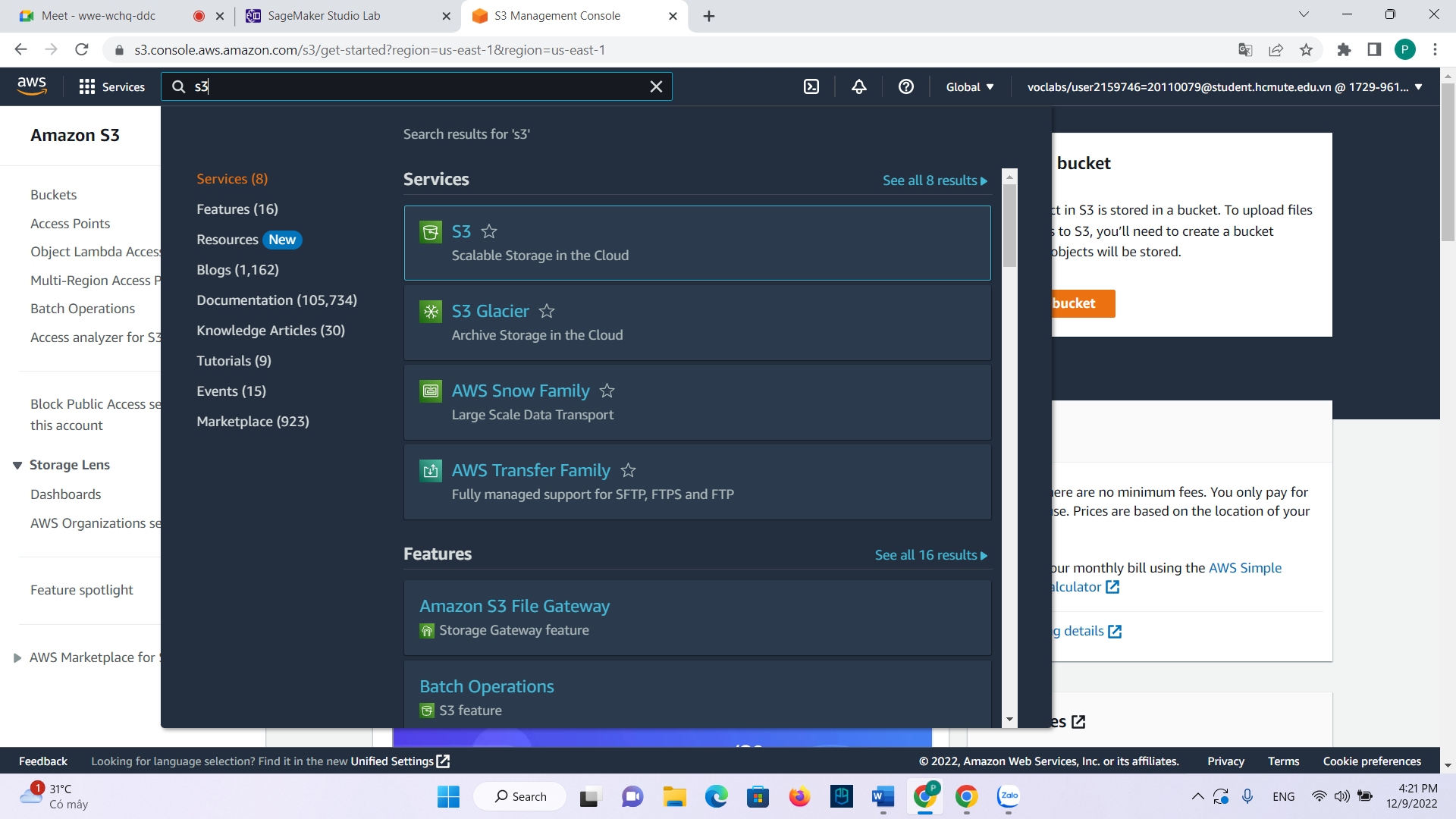
1. **Giới thiệu website**

Trang web được thiết kế chạy trên dịch vụ EC2 đồng thời dùng Bucket của S3 để lưu trữ dữ liệu.

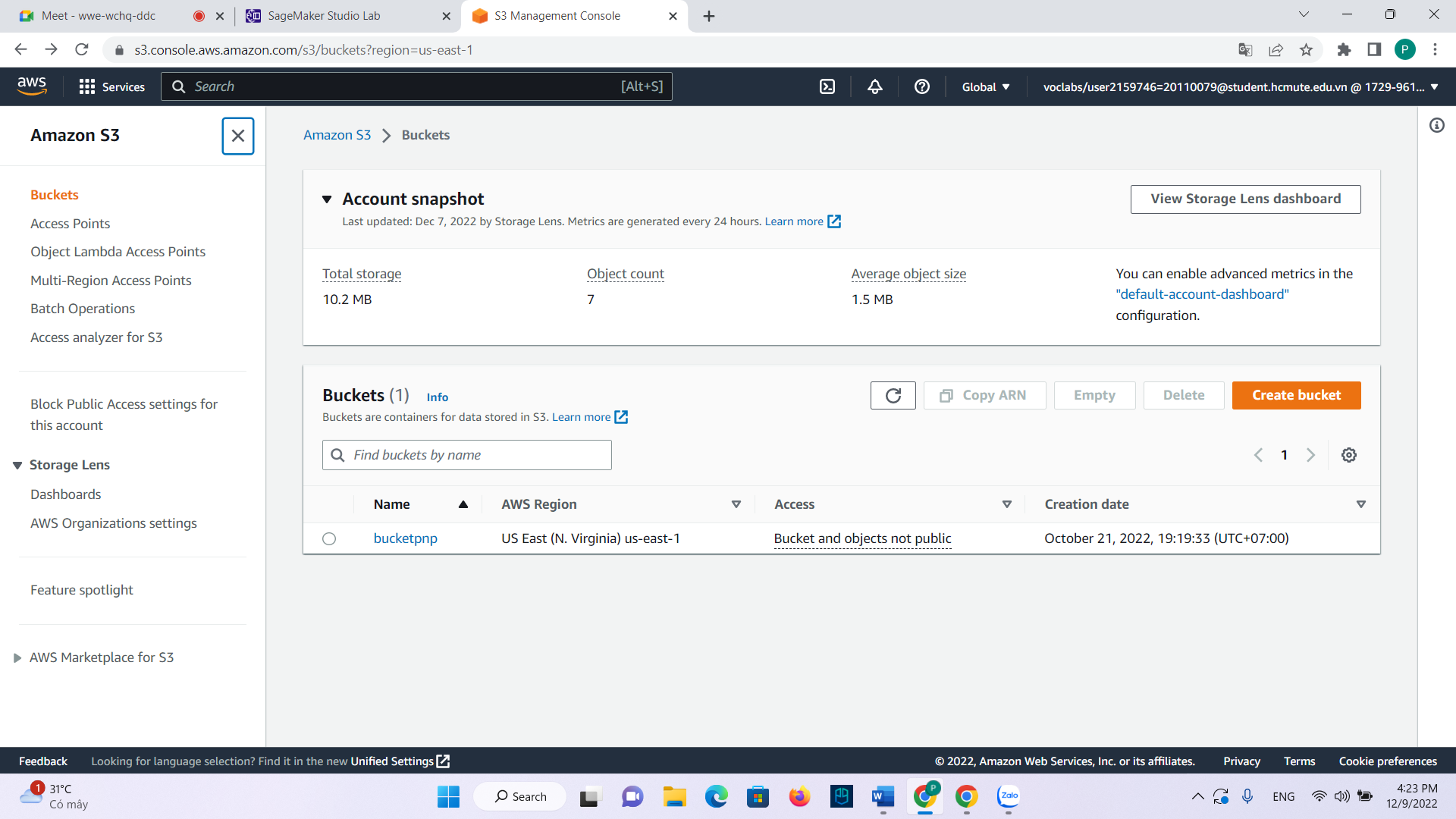
Cùng với sự hỗ trợ các dịch vụ của SageMaker giúp cài đặt, xây dựng mô hình nhận diện chữ viết thông qua hình ảnh. Đầu tiên, Data Wrangling giúp tiền xử lý dự liệu tiếp đó train model với Training và tạo Endpoint. Cuối cùng Deploy streamlit webapp với EC2 và Inference.

1. **Cài đặt**
   1. **Cài đặt AWS S3**

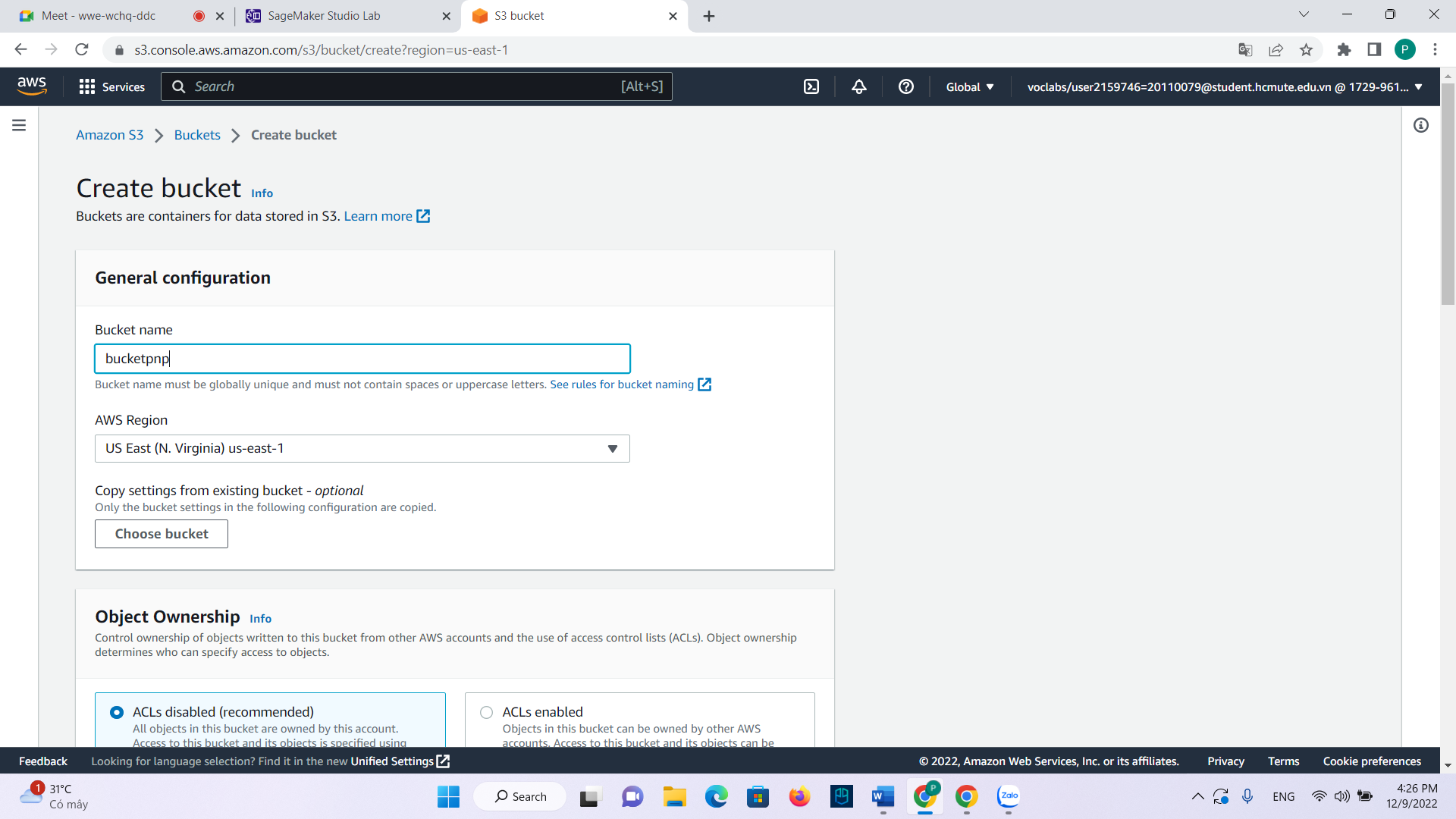
* Chọn dịch vụ S3



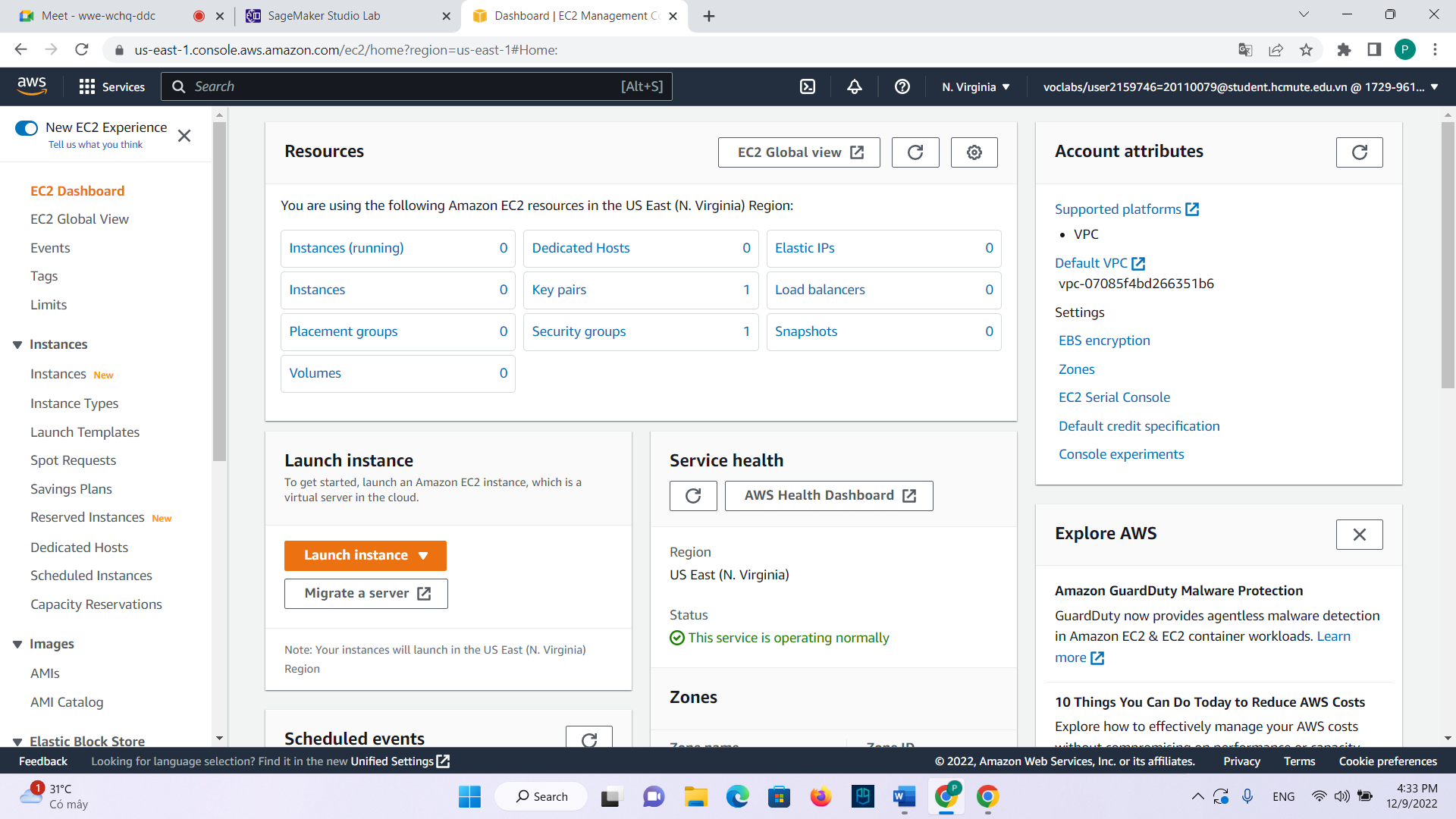
* Giao diện S3.



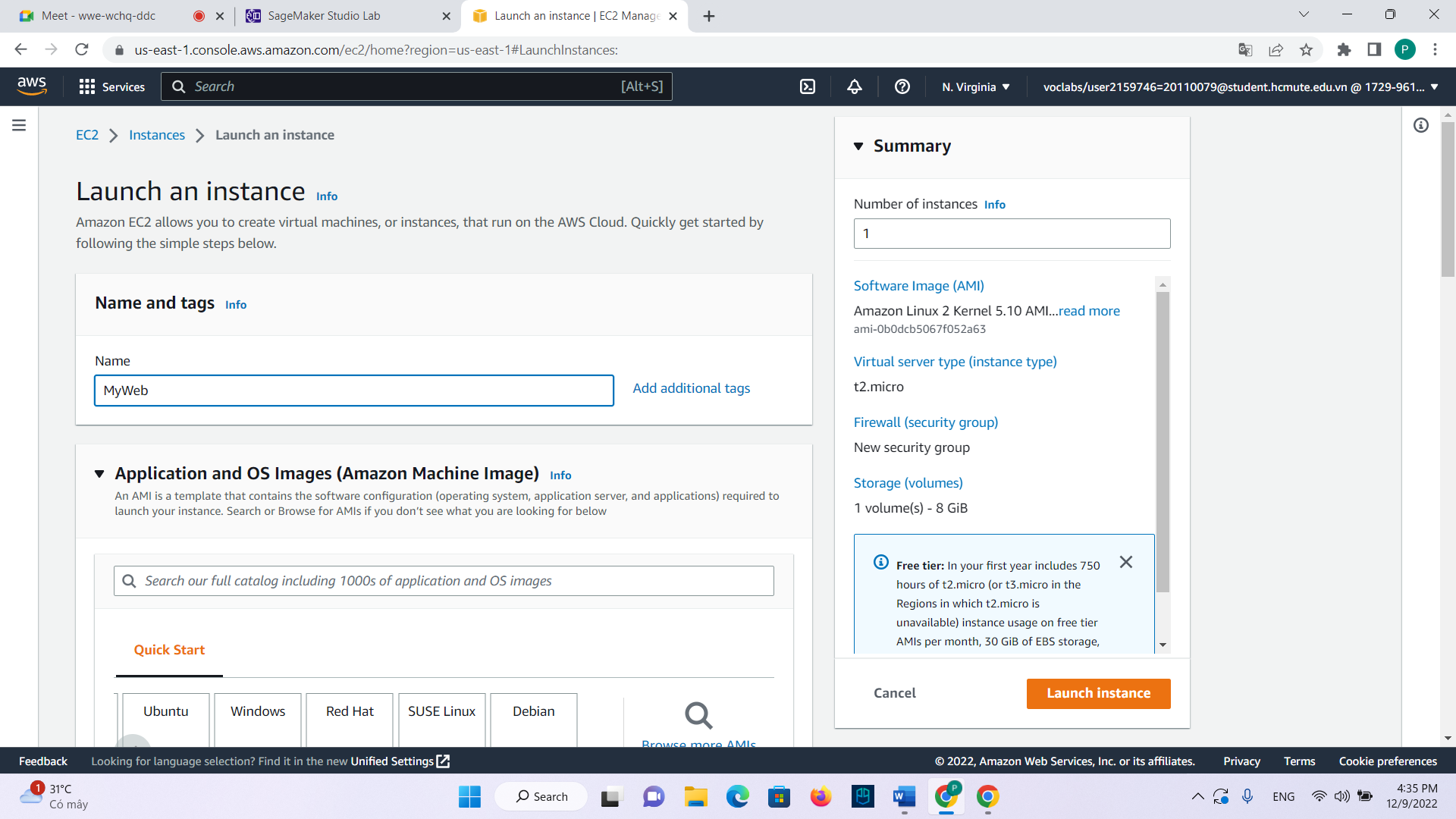
* Chọn Create bucket. Nhập tên, để các lựa chọn mặc định và chọn create bucket.



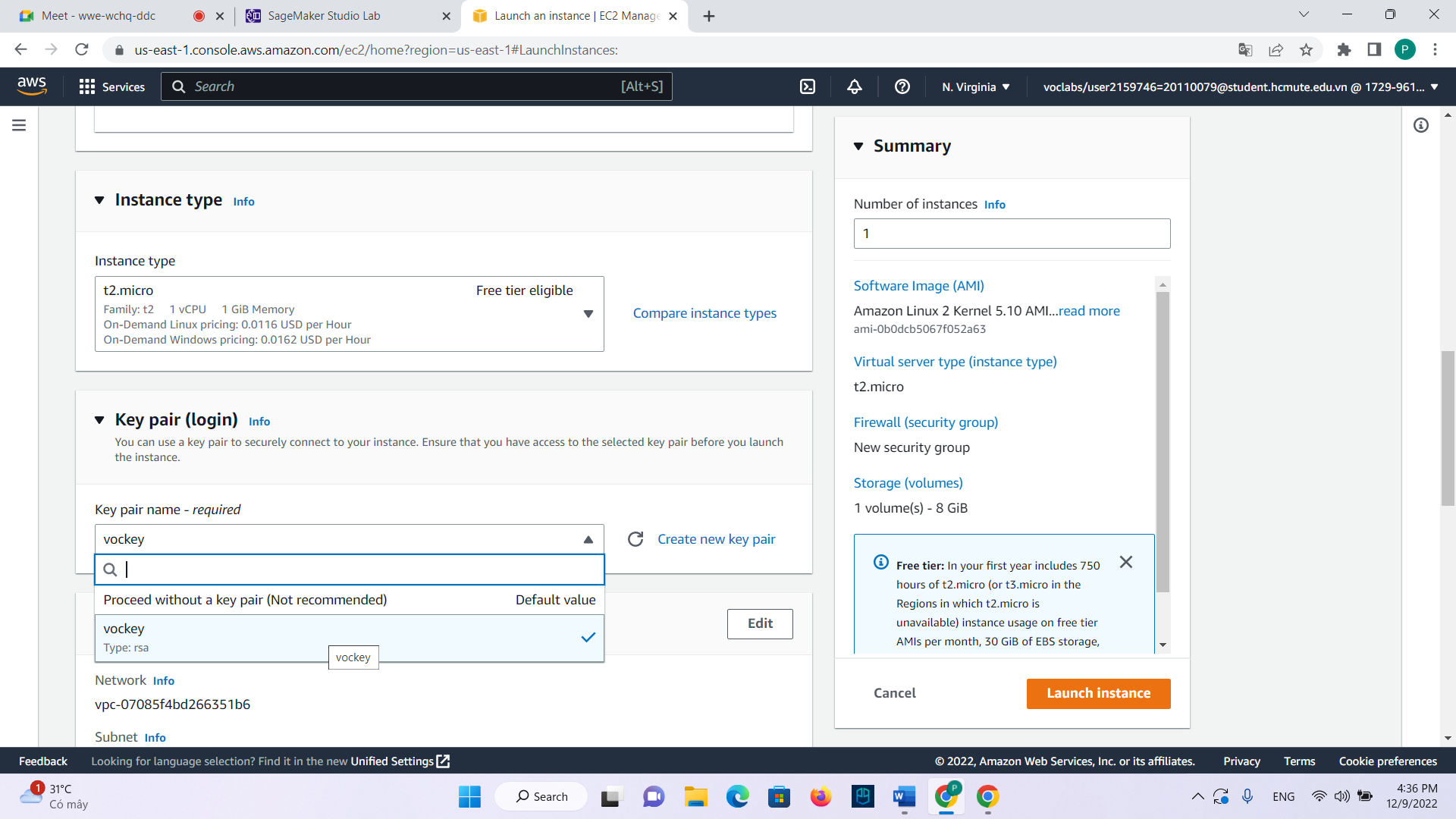
* Bucket sẽ được tạo.
  1. **Cài đặt AWS EC2**
* EC2



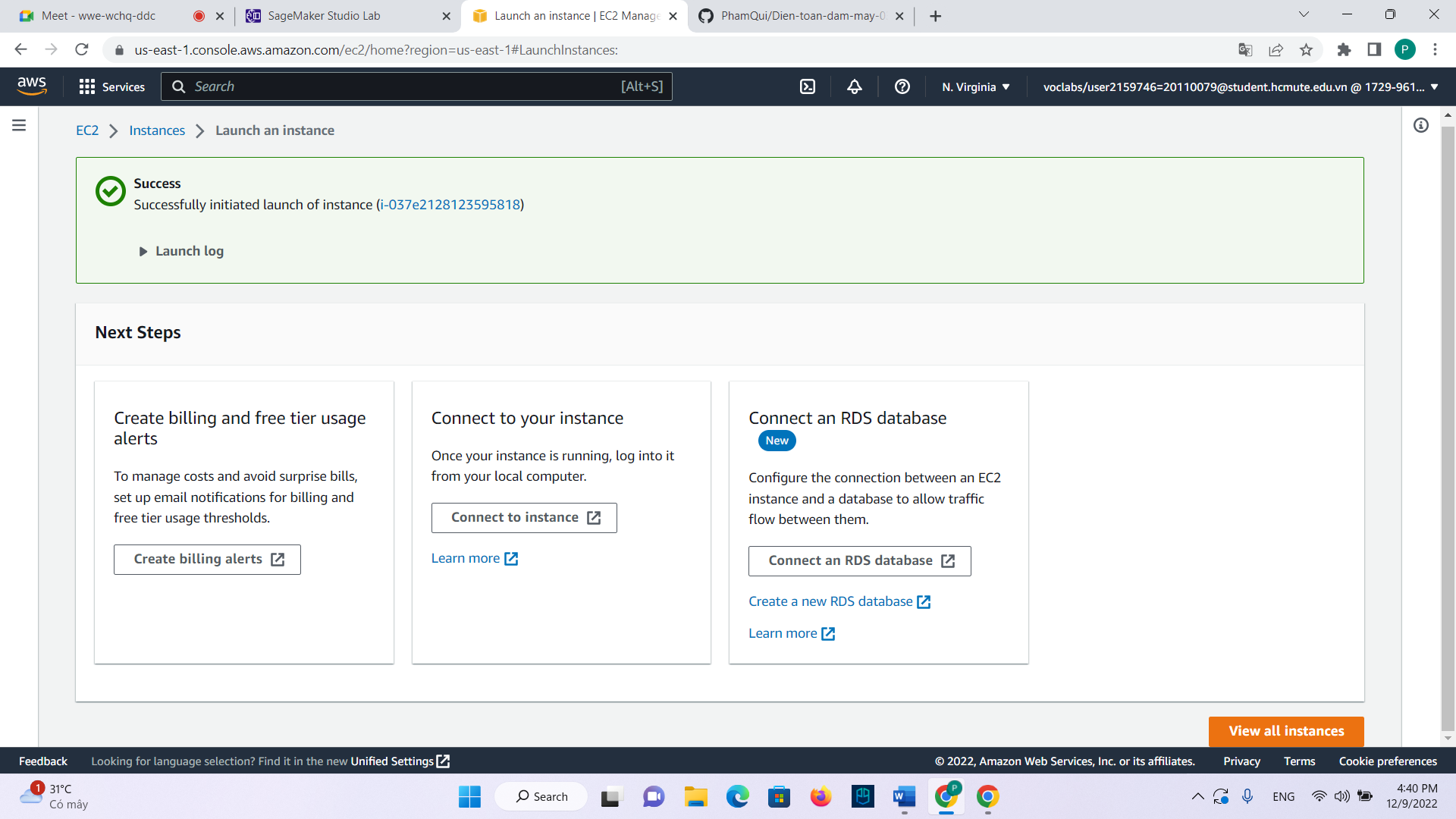
* Chọn launch instance. Nhập tên.



* Chọn Key pair. Sau đó chọn Launch instance



* Tạo thành công.

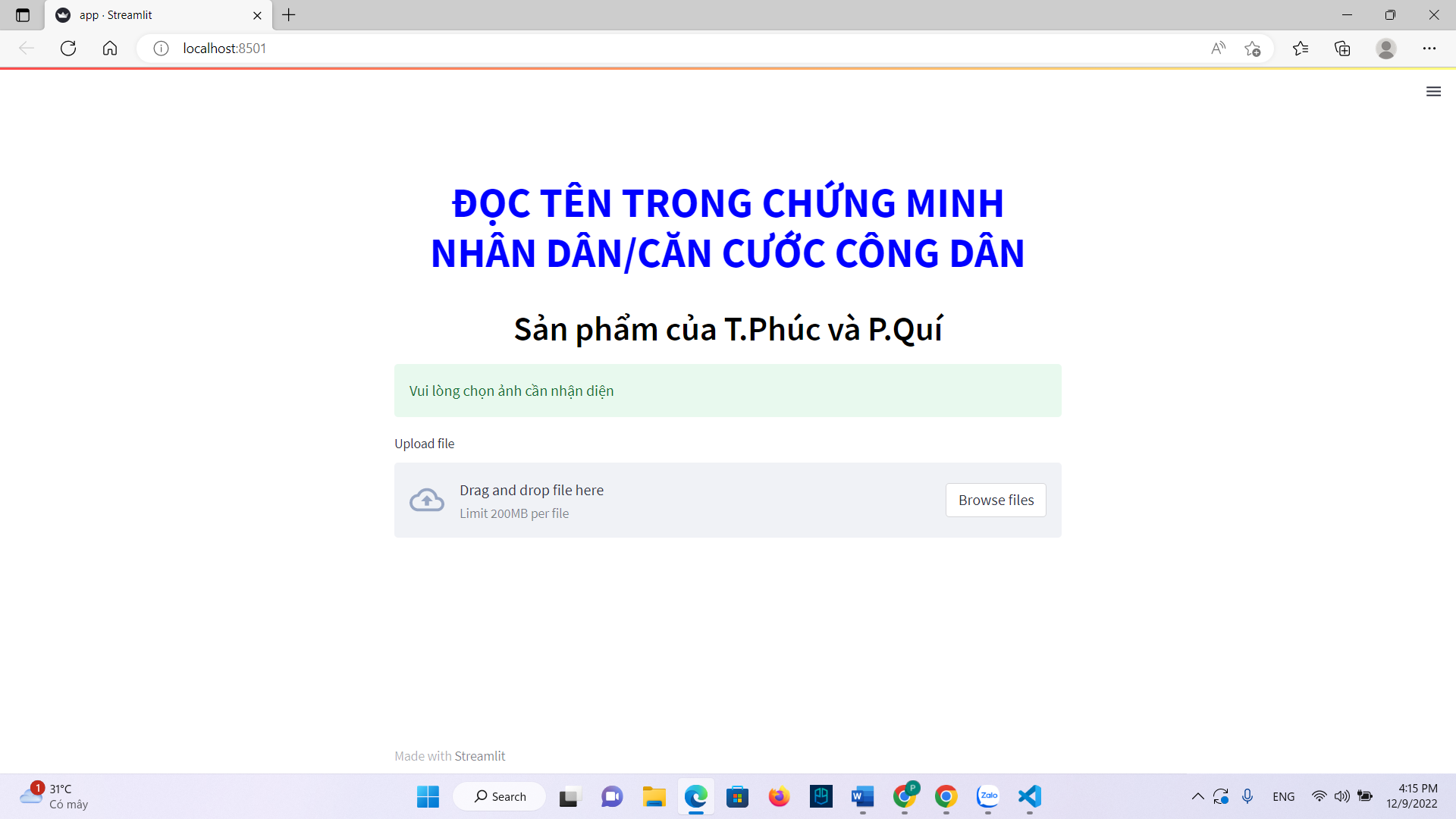


* 1. **Cài đặt AWS SageMaker**

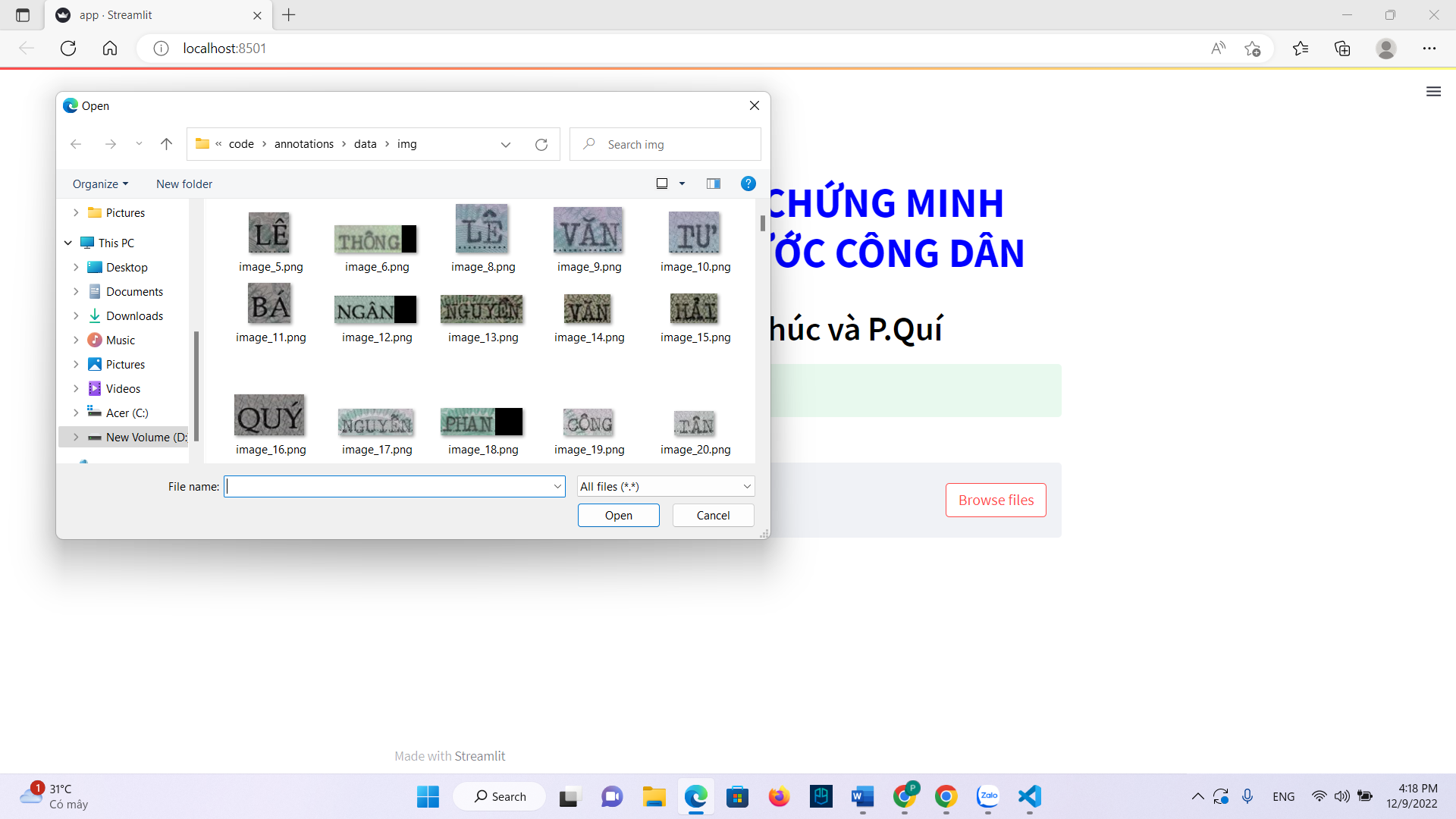
1. **Chức năng thực hiện được**

Nhận diện tên trong chứng minh nhân dân trên trang web.

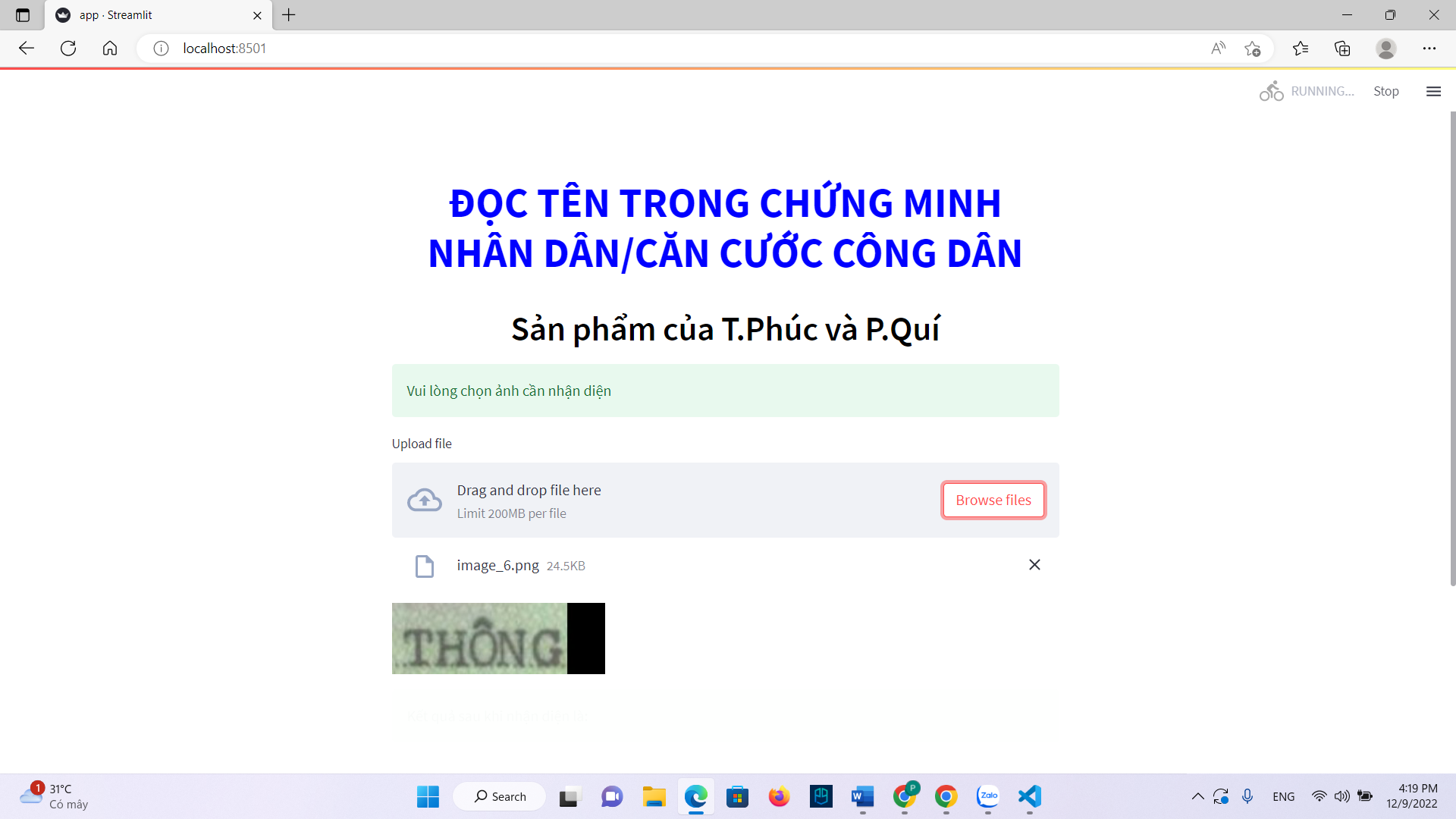
* Giao diện streamlit khi khởi chạy



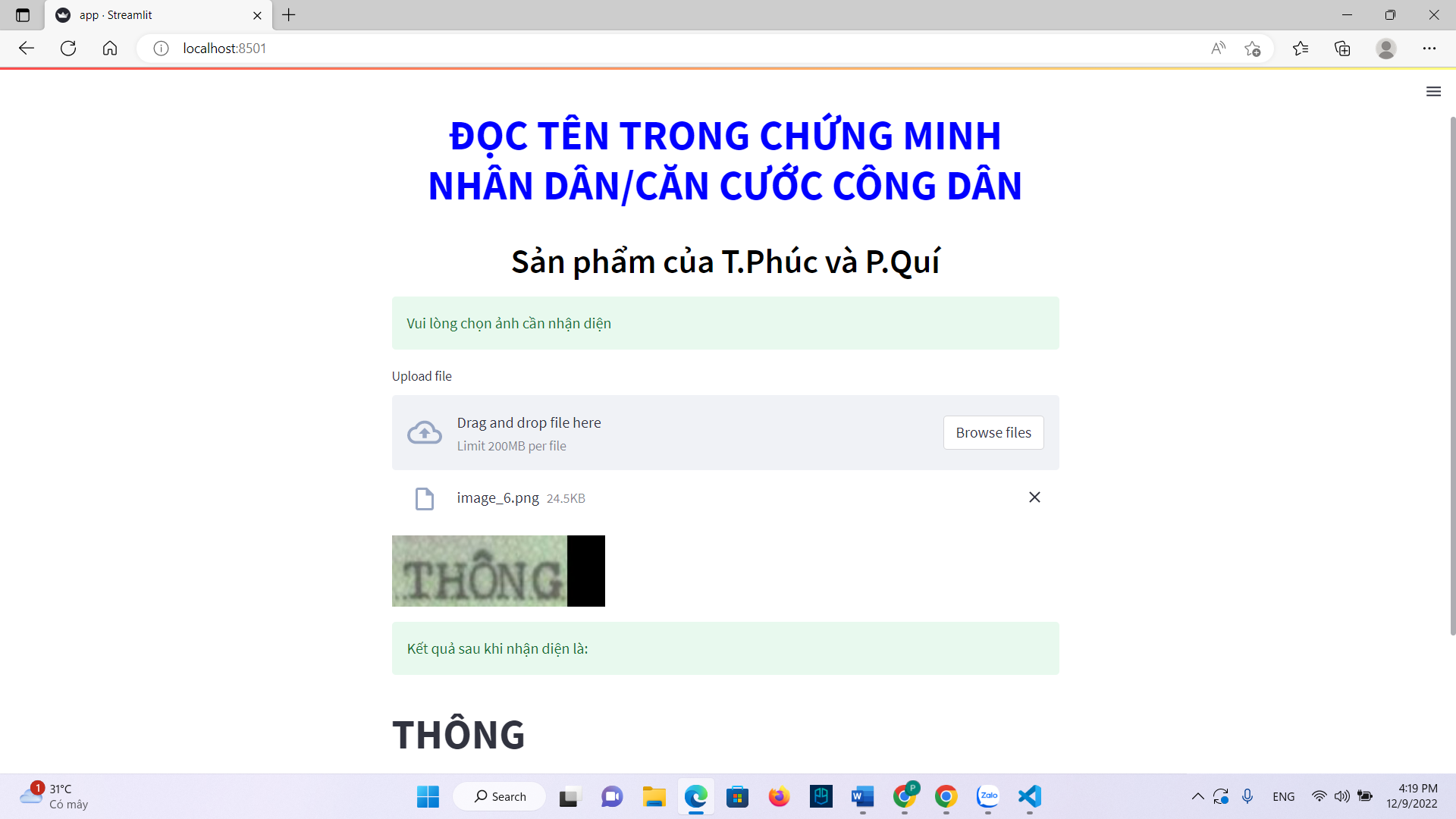
* Để nhận diện, ta chọn Browse files.



* Sau đó chọn hình ảnh cần nhận diện.



* Kết quả thu được.



# 

# **KẾT LUẬN**