TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Nguyễn Thành Nhật Tân - 52100106**

**Vương Thanh Huy - 52100038**

**XÂY DỰNG ỨNG DỤNG EMAIL TÍCH HỢP LỌC VÀ PHÂN LOẠI SPAM, LỪA ĐẢO - SỬ DỤNG MÔ HÌNH HỌC MÁY**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**KỸ THUẬT PHẦN MỀM**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Nguyễn Thành Nhật Tân - 52100106**

**Vương Thanh Huy - 52100038**

**XÂY DỰNG ỨNG DỤNG EMAIL TÍCH HỢP LỌC VÀ PHÂN LOẠI SPAM, LỪA ĐẢO - SỬ DỤNG MÔ HÌNH HỌC MÁY**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**KỸ THUẬT PHẦN MỀM**

Người hướng dẫn

**ThS. Huỳnh Ngọc Tú**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Chúng em muốn tỏ lòng biết ơn sâu sắc đến Cô **Huỳnh Ngọc Tú** và toàn bộ đội ngũ giáo viên trong Khoa Công nghệ thông tin tại Trường Đại học Tôn Đức Thắng. Sự hướng dẫn tận tâm và kiến thức sâu rộng của cô đã giúp chúng em không chỉ tiến bộ về chuyên môn mà còn về tư duy và khả năng sáng tạo trong môn học mẫu thiết kế.

Ngoài ra, chúng em cũng muốn bày tỏ lòng biết ơn đối với sự đồng hành của gia đình và bạn bè. Sự đồng lòng và hỗ trợ không điều kiện cùng những khoảnh khắc làm việc chung đã là nguồn động viên lớn, giúp chúng em vượt qua những thách thức và hoàn thiện công việc một cách tốt nhất.

Chúng em tự hào và biết ơn vì đã có cơ hội học tập và trải nghiệm những điều tuyệt vời như vậy trong quá trình học. Hy vọng chúng em sẽ tiếp tục nhận được sự hỗ trợ và hướng dẫn từ thầy cô, để không chỉ phát triển bản thân mà còn đóng góp tích cực vào cộng đồng và xã hội. Một lần nữa, chúng em xin chân thành cảm ơn tất cả mọi người!

*TP. Hồ Chí Minh, ngày 29 tháng 7 năm 2024*

*Tác giả*

*Nguyễn Thành Nhật Tân*

*Vương Thanh Huy*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ThS. Huỳnh Ngọc Tú. 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 Dự 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 Dự 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 29 tháng 7 năm 2024*

*Tác giả*

*Nguyễn Thành Nhật Tân*

*Vương Thanh Huy*

**XÂY DỰNG ỨNG DỤNG EMAIL TÍCH HỢP LỌC VÀ PHÂN LOẠI SPAM, LỪA ĐẢO - SỬ DỤNG MÔ HÌNH HỌC MÁY**

**TÓM TẮT**

Đề tài dự án xây dựng một ứng dụng email tích hợp khả năng lọc và phân loại email spam và lừa đảo, sử dụng các mô hình học máy (ML). Mục tiêu chính của dự án là tạo ra một hệ thống hiệu quả trong việc nhận diện và loại bỏ các email không mong muốn và nguy hiểm, từ đó nâng cao trải nghiệm người dùng và bảo vệ họ khỏi các mối đe dọa mạng.

Trong bài báo cáo, sẽ trình bày tóm tắt quá trình thực hiện và những vấn đề đã được nghiên cứu xuyên suốt quá trình xây dựng, bao gồm:

Nghiên cứu dữ liệu và xây dựng mô hình nhận dạng, phân loại các giống chó cảnh từ hình ảnh để áp dụng vào hệ thống quản lý dự án.

Phân tích và đặc tả tác nhân, hệ thống, thiết kế các sơ đồ usecase, các loại mô hình phục vụ cho việc thiết kế hệ thống như: mô hình tuần tự, mô hình hoạt động, mô hình lớp, mô hình quan hệ để xây dựng ứng dụng.

Nghiên cứu và hiện thực những cách huấn luyện mô hình khác nhau để xây dựng mô hình nhận diện tốt nhất cho ứng dụng di động.

Tìm hiểu phương pháp tích hợp mô hình trên hệ thống và ứng dụng trên thiết bị di động đã xây dựng. Bên cạnh đó, tiến hành xây dựng trang web admin để có thể quản lý các sản phẩm cho cửa hàng

**MỤC LỤC**

[DANH MỤC HÌNH VẼ v](#_Toc143173056)

[DANH MỤC BẢNG BIỂU vii](#_Toc143173057)

[DANH MỤC CÁC CHỮ VIẾT TẮT viii](#_Toc143173058)

[CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1](#_Toc143173059)

[1.1 Lý do chọn đề tài 1](#_Toc143173060)

[1.2 Mục tiêu thực hiện đề tài 1](#_Toc143173061)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 2](#_Toc143173062)

[2.1 Mạng neural hồi quy 2](#_Toc143173063)

[2.1.1 Recurrent Neural Network (RNN) 2](#_Toc143173064)

[2.1.2 Long Short-term Memory (LSTM) 2](#_Toc143173065)

[2.2 Mô hình Transformer 3](#_Toc143173066)

[2.2.1 Encoder và Decoder 3](#_Toc143173067)

[2.2.2 Attention 3](#_Toc143173068)

[CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT 4](#_Toc143173069)

[CHƯƠNG 4. THỰC NGHIỆM 4](#_Toc143173070)

[4.1 Dữ liệu thực nghiệm 4](#_Toc143173071)

[4.2 Cài đặt thực nghiệm 4](#_Toc143173072)

[CHƯƠNG 5. KẾT LUẬN 5](#_Toc143173073)

[5.1 Kết luận 5](#_Toc143173074)

[5.2 Hướng phát triển 5](#_Toc143173075)

[TÀI LIỆU THAM KHẢO 6](#_Toc143173076)

# DANH MỤC HÌNH VẼ

[Hình 2.2.1: Scaled Dot-Product Attention 5](#_Toc173263696)

[Hình 3.3.1 Sơ đồ usecase tổng quát 8](#_Toc173263697)

# DANH MỤC BẢNG BIỂU

[Bảng 3.4.1 Đặc tả usecase tạo tài khoản 9](#_Toc173263586)

[Bảng 3.4.2 Đặc tả usecase đăng nhập 10](#_Toc173263587)

[Bảng 3.4.3 Bảng đặc tả usecase gửi email 11](#_Toc173263588)

[Bảng 3.4.4 Bảng đặc tả usecase nhận email 12](#_Toc173263589)

[Bảng 3.4.5 Đặc tả usecase xem email chi tiết 13](#_Toc173263590)

[Bảng 3.4.6 Đặc tả usecase đánh dấu sao email 13](#_Toc173263591)

[Bảng 3.4.7 Đặc tả usecase xem email có gắn dấu sao 14](#_Toc173263592)

[Bảng 3.4.8 Đặc tả usecase xem chi tiết email có gắn dấu sao 15](#_Toc173263593)

[Bảng 3.4.9 Đặc tả usecase xem email spam 16](#_Toc173263594)

[Bảng 3.4.10 Đặc tả usecase xem email lừa đảo 17](#_Toc173263595)

[Bảng 3.4.11 Đặc tả usecase chuyển email vào thùng rác 19](#_Toc173263596)

[Bảng 3.4.12 Đặc tả usecase xóa email 20](#_Toc173263597)

[Bảng 3.4.13 Đặc tả usecase xem email ở thùng rác 21](#_Toc173263598)

[Bảng 3.4.14 Đặc tả usecase xem chi tiết email ở thùng rác 22](#_Toc173263599)

[Bảng 3.4.15 Đặc tả usecase đăng xuất 22](#_Toc173263600)

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| ML | Machine Learning |
| NLP | Natural Language Processing |
|  |  |
|  |  |
|  |  |

# MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

## Lý do chọn đề tài

Trong bối cảnh hiện nay, email vẫn là một trong những phương tiện giao tiếp chính trong công việc và cuộc sống hàng ngày. Tuy nhiên, sự phổ biến của email cũng kéo theo một loạt các vấn đề về bảo mật, trong đó nổi bật là các email spam và lừa đảo. Email spam thường chứa các nội dung quảng cáo không mong muốn, trong khi email lừa đảo có thể chứa các liên kết độc hại hoặc yêu cầu thông tin cá nhân, gây nguy cơ mất an toàn thông tin và tài sản cho người dùng.

Dự án này nhằm mục đích phát triển một ứng dụng email tích hợp công nghệ học máy (ML) và xử lý ngôn ngữ tự nhiên (NLP) để tự động phát hiện và phân loại các email spam và lừa đảo. Bằng cách này, giúp người dùng giảm thiểu rủi ro từ các email nguy hiểm, đồng thời nâng cao hiệu quả và trải nghiệm sử dụng email. Để giải quyết vấn đề này, nhiều giải pháp đã được đưa ra, trong đó, việc ứng dụng các kỹ thuật học máy (ML) và xử lý ngôn ngữ tự nhiên (NLP) vào việc lọc và phân loại email đang được xem là một hướng đi đầy triển vọng. Các mô hình học máy có khả năng học hỏi từ lượng lớn dữ liệu, tự động nhận diện các đặc trưng của email spam/lừa đảo và đưa ra quyết định phân loại chính xác.

Bên cạnh đó, việc tích hợp các mô hình học máy và NLP vào một ứng dụng email cũng đặt ra nhiều thách thức về mặt kỹ thuật, bao gồm việc thu thập và xử lý dữ liệu, huấn luyện mô hình, và đảm bảo tính bảo mật và hiệu suất của hệ thống. Tuy nhiên, chúng em tin rằng với những nỗ lực này sẽ đem lại giá trị lớn cho người dùng, giúp họ an tâm hơn khi sử dụng email và tập trung vào những công việc quan trọng khác.

## Mục đích nghiên cứu

Nghiên cứu này nhằm phát triển một giải pháp hiệu quả để phát hiện và loại bỏ các email spam và lừa đảo. Việc ứng dụng học máy và xử lý ngôn ngữ tự nhiên không chỉ giúp nâng cao độ chính xác của việc phát hiện mà còn có khả năng học hỏi và thích nghi với các mẫu spam và lừa đảo mới. Mục đích chính là xây dựng một hệ thống có thể bảo vệ người dùng một cách toàn diện, đồng thời giảm thiểu thời gian và công sức của người dùng trong việc quản lý và lọc email.

## Mục tiêu dự án

Mục tiêu chính của dự án bao gồm:

1. Xây dựng một mô hình học máy có khả năng phân loại email thành các nhóm spam, lừa đảo, và email hợp lệ.
2. Tích hợp các mô hình NLP để phân tích nội dung email và phát hiện các dấu hiệu của spam và lừa đảo.
3. Phát triển một giao diện người dùng thân thiện cho ứng dụng email, cho phép người dùng dễ dàng quản lý và kiểm tra các email bị phân loại.
4. Đảm bảo tính bảo mật và hiệu suất cao cho hệ thống, nhằm cung cấp một trải nghiệm người dùng mượt mà và an toàn.

## Phương pháp tiếp cận

Dự án sẽ sử dụng các kỹ thuật học máy và xử lý ngôn ngữ tự nhiên để xây dựng và huấn luyện các mô hình phát hiện spam và lừa đảo. Quá trình phát triển sẽ bao gồm các bước sau:

* Thu thập và xử lý dữ liệu email, bao gồm cả email hợp lệ, spam, và lừa đảo.
* Tiền xử lý dữ liệu để chuẩn bị cho quá trình huấn luyện mô hình, bao gồm việc loại bỏ các thông tin không cần thiết và chuẩn hóa dữ liệu.
* Sử dụng các thuật toán học máy như Naive Bayes, Support Vector Machines (SVM), và mạng nơ-ron để huấn luyện các mô hình phân loại.
* Ứng dụng các kỹ thuật NLP như phân tích từ vựng, phân tích cú pháp và phân tích ngữ nghĩa để cải thiện độ chính xác của việc phát hiện.
* Triển khai và tích hợp mô hình vào ứng dụng email, đảm bảo tính tương thích và hiệu suất cao.

## Kết quả dự kiến

Kết quả dự kiến của dự án bao gồm một ứng dụng email thông minh với khả năng lọc và phân loại email spam và lừa đảo một cách hiệu quả. Hệ thống sẽ có khả năng phát hiện và loại bỏ các email không mong muốn trước khi chúng tiếp cận hộp thư đến của người dùng, giúp nâng cao mức độ bảo mật và giảm thiểu nguy cơ bị lừa đảo. Ứng dụng sẽ cung cấp một giao diện người dùng thân thiện, cho phép người dùng dễ dàng kiểm tra và quản lý các email bị phân loại. Với việc áp dụng ML và NLP, hệ thống sẽ liên tục học hỏi và thích nghi với các mẫu spam và lừa đảo mới, đảm bảo hiệu quả bảo vệ cao nhất cho người dùng.

# CƠ SỞ LÝ THUYẾT

## Mạng neural hồi quy

### Recurrent Neural Network (RNN)

Về mặt toán học, mạng RNN có thể được mô tả như trong công thức (2.1) và (2.2). Công thứ (2.1), là vectơ đầu vào tại bước thứ . là trạng thái ẩn tại bước thứ . là một hàm phi tuyến tính (nonlinear function), thường là hàm tanh hay ReLu. là ma trận trọng số cho trạng thái ẩn ở bước trước đó , là ma trận trọng số cho đầu vào. Như vậy có thể thấy rằng, trạng thái ẩn ở bước thứ được tính dựa trên trạng thái ẩn ở bước trước đó và dữ liệu đầu vào ở bước hiện tại.

|  |  |  |
| --- | --- | --- |
|  |  | (2.1) |
|  |  | (2.2) |

Trong công thức (2.2), là ma trận trọng số cho đầu ra của mạng, là một phân phối xác suất trên từ điển tại bước thứ . Trạng thái ẩn được xem như là bộ nhớ của RNN, nó lưu trữ thông tin tính toàn được thực hiện ở các bước trước đó. Không giống như các mạng nơ-ron truyền thống, RNN chia sẽ chung bộ trọng số [, , ] cho tất cả các bước, nghĩa là các phép tính toán sẽ được thực hiện tương tự nhau cho tất cả các bước lặp chỉ các dữ liệu đầu vào. Đây là một ưu điểm của RNN giúp giảm số lượng tham số cần học cho mô hình.

### Long Short-term Memory (LSTM)

Mạng RNN chia sẽ chung một bộ trọng số giữa các bước lặp nên giảm đáng kể số lượng tham số, tuy nhiên nó vẫn là một mạng rất sâu. Trong quá trình lan truyền ngược (backward), phải thực hiện nhiều bước để có tính được đạo hàm cho những đầu vào đầu tiên của một chuỗi dữ liệu dài, do đó kết quả đạo hàm thường lớn hơn hoặc nhỏ hơn 1 đáng kể, dẫn đến giá trị của đạo hàm thường bị bùng nổ (exploding) hoặc mất mát (vanishing). Long Short-term Memory là một biến thể cải tiếng của RNN được đề xuất bởi (Hochreiter & Schmidhuber, 1997) nhằm khắc phục hai vấn đề trên.

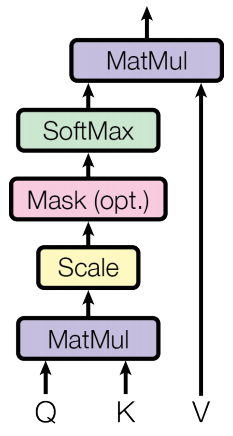
## Mô hình Transformer

### Encoder và Decoder

### Attention

Attention của mô hình Transformer bao gồm 02 kỹ thuật là Scaled Dot-Product Attention và Multi-head Attention.

#### Scaled Dot-Product Attention



Hình 2.2.1: Scaled Dot-Product Attention

(Nguồn: (Vaswani et al., 2023))

Sẽ có 03 ma trận , và được sử dụng trong kỹ thuật attention này, tương ứng với các khái niệm query, key và value. Các dòng trong trong ma trận và sẽ có kích thước , các dòng trong ma trận sẽ có kích thước . Các ma trận này được tạo ra từ việc cho vector embedding đầu vào đi qua 03 bộ trọng số , và .

#### Multi-head Attention

# PHÂN TÍCH THIẾT KẾ YÊU CẦU

## Giới thiệu chung

Ứng dụng Email Thông Minh là một ứng dụng mộ phỏng lại hệ thống email. Nó được thiết kế để cung cấp khả năng lọc và phân loại email spam và lừa đảo. Hệ thống này sử dụng các công nghệ học máy (ML) và xử lý ngôn ngữ tự nhiên (NLP) để cải thiện độ chính xác và hiệu quả trong việc phát hiện các email không mong muốn và nguy hiểm, bảo vệ người dùng và nâng cao trải nghiệm sử dụng email.

## Đặc tả

### Đặc tả hệ thống

Gửi và nhận email: Hệ thống phải cho phép người dùng gửi và nhận email một cách dễ dàng và hiệu quả. Người dùng có thể soạn thảo, đính kèm tệp, và gửi email tới một hoặc nhiều địa chỉ liên lạc. Chức năng này cũng bao gồm việc nhận và lưu trữ email từ các dịch vụ email khác nhau, đảm bảo rằng tất cả email đến đều được tổ chức và hiển thị trong hộp thư đến của người dùng.

Mục có gắn dấu sao: Người dùng có thể gắn dấu sao cho các email quan trọng để dễ dàng theo dõi và tìm kiếm sau này. Các email có gắn dấu sao sẽ được hiển thị trong một thư mục riêng biệt để tiện quản lý.

Lọc email spam: Hệ thống sẽ tự động phát hiện và lọc các email spam, chuyển chúng vào một thư mục riêng biệt. Mô hình học máy sẽ được huấn luyện để nhận diện các đặc điểm của email spam, như các từ khóa thường xuất hiện, cấu trúc email, và các mẫu gửi email hàng loạt. Người dùng cũng có thể cấu hình các tiêu chí lọc spam của riêng mình, thêm hoặc xóa các địa chỉ email vào danh sách chặn hoặc danh sách cho phép.

Phân loại email lừa đảo: Hệ thống sẽ sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên để phân tích nội dung email và phát hiện các dấu hiệu của email lừa đảo. Các dấu hiệu này có thể bao gồm các yêu cầu cung cấp thông tin cá nhân, các liên kết đến các trang web đáng ngờ, hoặc các thông báo khẩn cấp giả mạo. Khi phát hiện email lừa đảo, hệ thống sẽ cảnh báo người dùng và chuyển email này vào một thư mục riêng để kiểm tra thêm.

Thùng rác: Các email bị xóa sẽ được chuyển vào thùng rác, nơi người dùng có thể khôi phục hoặc xóa vĩnh viễn các email này. Thùng rác sẽ tự động xóa các email cũ sau một khoảng thời gian nhất định để tiết kiệm không gian lưu trữ.

Quản lý hộp thư: Ứng dụng cung cấp giao diện trực quan và thân thiện để người dùng dễ dàng quản lý các email trong hộp thư của mình. Người dùng có thể tạo các thư mục tùy chỉnh, di chuyển email giữa các thư mục, đánh dấu email quan trọng, và gắn nhãn cho email để dễ dàng tìm kiếm sau này. Ngoài ra, hệ thống còn hỗ trợ các thao tác như xóa, lưu trữ, và đánh dấu đã đọc/chưa đọc cho các email.

### Kiến trúc hệ thống

**Lớp giao diện người dùng:**

* Ứng dụng Windows với giao diện trực quan, thân thiện, cho phép người dùng dễ dàng tương tác và quản lý email.

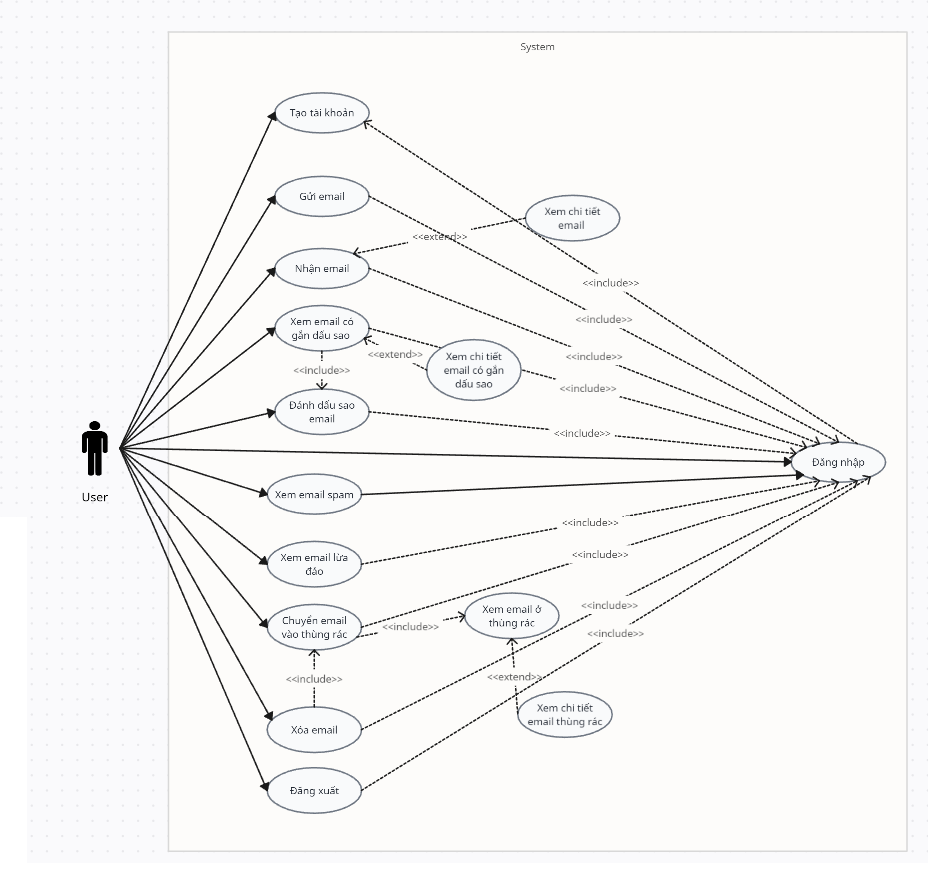
**Lớp ứng dụng:**

* Quản lý Email: Xử lý logic nghiệp vụ liên quan đến việc gửi, nhận và quản lý email.
* Lọc và Phân loại: Sử dụng các mô hình học máy và NLP để phát hiện và phân loại email spam và lừa đảo.

**Lớp dữ liệu:**

* Cơ sở dữ liệu: Lưu trữ thông tin người dùng, dữ liệu email, và các mô hình học máy. Sử dụng Firebase Storage.
* Dữ liệu Huấn luyện: Lưu trữ dữ liệu email được sử dụng để huấn luyện các mô hình học máy.

## Sơ đồ usecase



Hình 3.3.1 Sơ đồ usecase tổng quát

## Đặc tả usecase

### Đặc tả usecase tạo tài khoản

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Tạo tài khoản | |
| Tác nhân | User | |
| Mô tả | User muốn tạo tài khoản | |
| Điều kiện tiên quyết | User phải có kết nối internet ổn định.  User chưa có tài khoản trên hệ thống. | |
| Điều kiện sau | Tạo tài khoản thành công | |
| Điều kiện thoát | User quyết định không tạo tài khoản và thoát khỏi quá trình đăng ký.  Hệ thống gặp lỗi và không thể xử lý yêu cầu tạo tài khoản. | |
| Luồng | Actor | System |
|  | 1. User nhập thông tin đăng nhập, mật khẩu và xác nhận mật khẩu  2. Nhấn nút đăng ký | 2.1 Hệ thống kiểm tra tài khoản và mật khẩu và trả về kết quả sau cùng  2.2 Kết thúc usecase. |
| Điều kiện ngoại lệ | 2.1 Nếu thông tin đăng ký không hợp lệ (VD: tên đăng nhập đã tồn tại, mật khẩu không đủ mạnh,…) thì hệ thống sẽ báo lỗi và yêu cầu nhập lại | |

Bảng 3.4.1 Đặc tả usecase tạo tài khoản

### Đặc tả usecase đăng nhập

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Đăng nhập | |
| Tác nhân | User | |
| Mô tả | User muốn đăng nhập vào hệ thống | |
| Điều kiện tiên quyết | User phải có kết nối internet ổn định.  User đã có tài khoản trên hệ thống. | |
| Điều kiện sau | Đăng nhập vào hệ thống thành công | |
| Điều kiện thoát | User đăng nhập thành công và muốn thoát khỏi hệ thống | |
| Luồng | Actor | System |
|  | 1. User nhập thông tin đăng nhập, mật khẩu  2. Nhấn nút đăng nhập | 2.1 Hệ thống kiểm tra tài khoản và mật khẩu và trả về kết quả sau cùng  2.2 Kết thúc usecase. |
| Điều kiện ngoại lệ | 2.1 Nếu thông tin đăng nhập không hợp lệ (VD: tên đăng nhập không tồn tại, mật khẩu sai,…) thì hệ thống sẽ báo lỗi và yêu cầu nhập lại | |

Bảng 3.4.2 Đặc tả usecase đăng nhập

### Đặc tả usecase gửi email

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Gửi email | |
| Tác nhân | User | |
| Mô tả | User muốn gửi email đến người nhận | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống. | |
| Điều kiện sau | Email được gửi thành công đến người nhận | |
| Điều kiện thoát | User quyết định không gửi email và thoát.  Hệ thống gặp lỗi và không thể gửi email. | |
| Luồng | Actor | System |
|  | 1. User nhập địa chỉ email người nhận, tiêu đề, và nội dung email.  2. Nhấn nút "Gửi". | 2.1 Hệ thống kiểm tra địa chỉ email người nhận và nội dung email để xác nhận và gửi email đến người nhận  2.2 Kết thúc usecase. |
| Điều kiện ngoại lệ | 2. Hệ thống gặp lỗi kỹ thuật | |

Bảng 3.4.3 Bảng đặc tả usecase gửi email

### Đặc tả usecase nhận email

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Nhận email | |
| Tác nhân | User | |
| Mô tả | User nhận email từ người gửi | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống. | |
| Điều kiện sau | User xem được các email mới trong hộp thư. | |
| Điều kiện thoát | User quyết định không kiểm tra email và thoát.  Hệ thống gặp lỗi và không thể tải email. | |
| Luồng | Actor | System |
|  | 1.User nhấn vào tab “Hộp thư đến” và xem danh sách email đã nhận. | 2.1 Hệ thống nhận dữ liệu từ cơ sở dữ liệu và cập nhập lên giao diện người dùng  2.2 Kết thúc usecase. |
| Điều kiện ngoại lệ | 2. Hệ thống gặp lỗi kỹ thuật | |

Bảng 3.4.4 Bảng đặc tả usecase nhận email

### Đặc tả usecase xem email chi tiết

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem email chi tiết | |
| Tác nhân | User | |
| Mô tả | User xem email chi tiết từ người gửi | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống. | |
| Điều kiện sau | User xem được email chi tiết trong hộp thư | |
| Điều kiện thoát | User quyết định không xem email chi tiết và quay về hộp thư đến hoặc thoát.  Hệ thống gặp lỗi và không thể tải email. | |
| Luồng | Actor | System |
|  | 1.User ở tab “Hộp thư đến” và ấn vào email muốn xem chi tiết. | 2.1 Hệ thống nhận dữ liệu từ cơ sở dữ liệu  2.2 Hệ thống hiện thi thông tin chi tiết của email  2.3 Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải chi tiết email. | |

Bảng 3.4.5 Đặc tả usecase xem email chi tiết

### Đặc tả usecase đánh dấu sao email

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Đánh dấu sao email | |
| Tác nhân | User | |
| Mô tả | User đánh dấu sao những email quan trọng | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  User đã nhận được email và muốn đánh dấu sao. | |
| Điều kiện sau | Email được đánh dấu sao và hiển thị trong danh sách email đã đánh dấu sao. | |
| Điều kiện thoát | User không thể đánh dấu sao email do lỗi hệ thống | |
| Luồng | Actor | System |
|  | 1. User chọn một email từ danh sách email đã nhận.  2. User nhấn vào biểu tượng sao để đánh dấu email. | 1. Hệ thống ghi nhận trạng thái đánh dấu sao của email.  2. Hệ thống cập nhật trạng thái đánh dấu sao trong cơ sở dữ liệu và giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể đánh dấu sao email. | |

Bảng 3.4.6 Đặc tả usecase đánh dấu sao email

### Đặc tả usecase xem email có gắn dấu sao

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem email có gắn dấu sao email | |
| Tác nhân | User | |
| Mô tả | User xem danh sách các email đã được đánh dấu sao để quản lý và xử lý. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  User đã đánh dấu sao ít nhất một email. | |
| Điều kiện sau | User xem được danh sách các email đã được đánh dấu sao. | |
| Điều kiện thoát | User quyết định không xem danh sách email đã đánh dấu sao và thoát. | |
| Luồng | Actor | System |
|  | 1. User nhấn vào tab "Gắn dấu sao" hoặc chọn tùy chọn tương ứng trên giao diện. | 1. Hệ thống truy vấn cơ sở dữ liệu để lấy danh sách email đã được đánh dấu sao.  2. Hệ thống hiển thị danh sách email đã được đánh dấu sao trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải danh sách email đã đánh dấu sao. | |

Bảng 3.4.7 Đặc tả usecase xem email có gắn dấu sao

### Đặc tả usecase xem chi tiết email có gắn dấu sao

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem chi tiết email có gắn dấu sao email | |
| Tác nhân | User | |
| Mô tả | User xem chi tiết nội dung của một email đã được đánh dấu sao. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  User đã đánh dấu sao ít nhất một email. | |
| Điều kiện sau | User xem được nội dung chi tiết của email được chọn. | |
| Điều kiện thoát | User quyết định không xem chi tiết email và thoát.  Hệ thống gặp lỗi và không thể tải nội dung chi tiết email. | |
| Luồng | Actor | System |
|  | 1. User chọn một email đã được đánh dấu sao từ danh sách.  2. User nhấn vào email để mở chi tiết.. | 1. Hệ thống gửi yêu cầu lấy dữ liệu chi tiết của email từ cơ sở dữ liệu.  2. Hệ thống nhận dữ liệu và hiển thị nội dung chi tiết của email trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải chi tiết email. | |

Bảng 3.4.8 Đặc tả usecase xem chi tiết email có gắn dấu sao

### Đặc tả usecase xem email spam

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem email spam | |
| Tác nhân | User | |
| Mô tả | User xem danh sách các email được phân loại là spam để kiểm tra hoặc xóa. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Có email được hệ thống phân loại là spam. | |
| Điều kiện sau | User xem được danh sách các email được phân loại là spam. | |
| Điều kiện thoát | User quyết định không xem danh sách email spam và thoát. | |
| Luồng | Actor | System |
|  | 1. User nhấn vào tab "Spam" hoặc chọn tùy chọn tương ứng trên giao diện. | 1. Hệ thống truy vấn cơ sở dữ liệu để lấy danh sách email được phân loại là spam.  2. Hệ thống hiển thị danh sách email spam trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải danh sách email spam. | |

Bảng 3.4.9 Đặc tả usecase xem email spam

### Đặc tả usecase xem email lừa đảo

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem email lừa đảo | |
| Tác nhân | User | |
| Mô tả | User xem danh sách các email được phân loại là lừa đảo để kiểm tra hoặc báo cáo. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Có email được hệ thống phân loại là lừa đảo. | |
| Điều kiện sau | User xem được danh sách các email được phân loại là lừa đảo. | |
| Điều kiện thoát | User quyết định không xem danh sách email lừa đảo và thoát. | |
| Luồng | Actor | System |
|  | 1. User nhấn vào tab "Phishing" hoặc chọn tùy chọn tương ứng trên giao diện. | 1. Hệ thống truy vấn cơ sở dữ liệu để lấy danh sách email được phân loại là lừa đảo.  2. Hệ thống hiển thị danh sách email lừa đảo trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải danh sách email lừa đảo. | |

Bảng 3.4.10 Đặc tả usecase xem email lừa đảo

### Đặc tả usecase chuyển email vào thùng rác

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Chuyển email vào thùng rác | |
| Tác nhân | User | |
| Mô tả | User chuyển email vào thùng rác để loại bỏ khỏi hộp thư chính. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Có email mà User muốn chuyển vào thùng rác. | |
| Điều kiện sau | Email được chuyển vào thùng rác và không còn hiển thị trong hộp thư chính. | |
| Điều kiện thoát | User quyết định không chuyển email vào thùng rác hoặc hệ thống gặp lỗi không thể thực hiện hành động này. | |
| Luồng | Actor | System |
|  | 1. User chọn email cần chuyển vào thùng rác từ danh sách.  2. User nhấn vào icon thùng rác ở phía trên appbar để thực hiện | 1. Hệ thống cập nhật trạng thái của email trong cơ sở dữ liệu là "Trong thùng rác."  2. Hệ thống cập nhật giao diện người dùng, loại bỏ email khỏi hộp thư chính và đưa vào mục thùng rác.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể chuyển email vào thùng rác. | |

Bảng 3.4.11 Đặc tả usecase chuyển email vào thùng rác

### Đặc tả usecase xóa email

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xóa email | |
| Tác nhân | User | |
| Mô tả | User xóa hoàn toàn email từ thùng rác. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Email và user muốn xóa phải nằm ở mục thùng rác. | |
| Điều kiện sau | Email bị xóa hoàn toàn khỏi hệ thống và không thể khôi phục. | |
| Điều kiện thoát | User quyết định không xóa email hoặc hệ thống gặp lỗi không thể thực hiện hành động này. | |
| Luồng | Actor | System |
|  | 1. User chọn email cần chuyển vào thùng rác từ danh sách.  2. User nhấn vào icon thùng rác ở phía trên appbar để thực hiện | 1. Hệ thống xóa hoàn toàn email từ cơ sở dữ liệu.  2. Hệ thống cập nhật giao diện người dùng, loại bỏ email khỏi mọi thư mục.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể xóa email. | |

Bảng 3.4.12 Đặc tả usecase xóa email

### Đặc tả usecase xem email ở thùng rác

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem email ở thùng rác | |
| Tác nhân | User | |
| Mô tả | User xem danh sách các email đã được chuyển vào thùng rác để kiểm tra. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Có email trong thùng rác. | |
| Điều kiện sau | User xem được danh sách các email trong thùng rác. | |
| Điều kiện thoát | User quyết định không xem danh sách email trong thùng rác và thoát. | |
| Luồng | Actor | System |
|  | 1. User nhấn vào tab "Thùng rác" hoặc chọn tùy chọn tương ứng trên giao diện. | 1. Hệ thống truy vấn cơ sở dữ liệu để lấy danh sách email trong thùng rác.  2. Hệ thống hiển thị danh sách email trong thùng rác trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải danh sách email trong thùng rác. | |

Bảng 3.4.13 Đặc tả usecase xem email ở thùng rác

### Đặc tả usecase xem chi tiết mail ở thùng rác

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Xem chi tiết email ở thùng rác | |
| Tác nhân | User | |
| Mô tả | User xem chi tiết nội dung của một email đã được chuyển vào thùng rác. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống.  User đã có tài khoản trên hệ thống.  Có email trong thùng rác. | |
| Điều kiện sau | User xem được nội dung chi tiết của email được chọn. | |
| Điều kiện thoát | User quyết định không xem chi tiết email và thoát.  Hệ thống gặp lỗi và không thể tải nội dung chi tiết email. | |
| Luồng | Actor | System |
|  | 1. User chọn một email từ danh sách email trong thùng rác.  2. User nhấn vào email để mở chi tiết. | 1. Hệ thống gửi yêu cầu lấy dữ liệu chi tiết của email từ cơ sở dữ liệu.  2. Hệ thống nhận dữ liệu và hiển thị nội dung chi tiết của email trên giao diện người dùng.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể tải chi tiết email. | |

Bảng 3.4.14 Đặc tả usecase xem chi tiết email ở thùng rác

### Đặc tả usecase đăng xuất

|  |  |  |
| --- | --- | --- |
| Tên usecase: | Đăng xuất | |
| Tác nhân | User | |
| Mô tả | User thực hiện hành động đăng xuất khỏi hệ thống. | |
| Điều kiện tiên quyết | User đã đăng nhập vào hệ thống. | |
| Điều kiện sau | User được đăng xuất khỏi hệ thống và trở về giao diện đăng nhập. | |
| Điều kiện thoát | User hủy bỏ hành động đăng xuất. | |
| Luồng | Actor | System |
|  | 1. User nhấn vào icon "Đăng xuất" từ giao diện người dùng. | 1. Hệ thống kết thúc phiên làm việc của user.  2. Hệ thống cập nhật giao diện người dùng, đưa User trở về màn hình đăng nhập.  3. Kết thúc usecase. |
| Điều kiện ngoại lệ | 1. Hệ thống gặp lỗi kỹ thuật, không thể đăng xuất. | |

Bảng 3.4.15 Đặc tả usecase đăng xuất

## Công nghệ sử dụng

### Dart và Flutter:

**Dart**: Dart là một ngôn ngữ lập trình mạnh mẽ được phát triển bởi Google. Nó được sử dụng chủ yếu để phát triển các ứng dụng web và di động. Dart có cú pháp tương tự như các ngôn ngữ lập trình như JavaScript, giúp dễ dàng tiếp cận và học hỏi đối với các nhà phát triển.

**Flutter**: Flutter là một UI framework mã nguồn mở được phát triển bởi Google với mục đích tạo giao diện chất lượng cao đa nền tảng. Thành phần chính của Flutter bao gồm:

* SDK: Là một bộ những công cụ cho người dùng phát triển ứng dụng.
* Framework: Cho phép tập hợp các thành phần giao diện.
* Ưu điểm của Flutter:
* Sử dụng Dart, một ngôn ngữ lập trình được biên dịch thành mã máy, giúp nâng cao hiệu suất.
* Dựa vào thư viện đồ họa Skia để kết xuất, nâng cao trải nghiệm người dùng bởi tính nhất quán.
* Dễ sử dụng, dễ trực quan hóa và giải quyết các vấn đề tồn động của bố cục giao diện.
* Khả năng xây dựng ứng dụng đa nền tảng (iOS và Android) với một mã nguồn duy nhất.

### Firebase Storage:

**Firebase Storage** là dịch vụ lưu trữ được phát triển bởi Google để lưu trữ âm thanh, ảnh, video và các loại tệp khác. Dịch vụ này lưu trữ file trong Google Cloud và có thể dùng API Google Cloud Storage để quản lý việc tải lên.

Ưu điểm của Firebase Storage:

* Đánh dấu tiến trình khi có sự cố với internet.
* Tính bảo mật cao, cung cấp xác thực tệp và dữ liệu.
* Dễ dàng mở rộng quy mô.
* Tích hợp chặt chẽ với các dịch vụ khác của Firebase, giúp đơn giản hóa quá trình phát triển ứng dụng.

# THỰC NGHIỆM

## Dữ liệu thực nghiệm

Dữ liệu trong VLSP-2016 đã được chia sẵn thành hai tập huấn luyện và kiểm thử. Thống kê chi tiết của các kiểu thực thể và các tập dữ liệu được mô tả trong Bảng 4.1**.**

Bảng 4.1: Thống kê kiểu thực thể trong tập VLSP 2016

|  |  |  |
| --- | --- | --- |
| **Label** | **Train** | **Test** |
| **LOC** | 1210 | 1377 |
| **ORG** | 7478 | 274 |
| **PER** | 6230 | 1294 |
| **MISC** | 250 | 47 |
| **Total** | 15168 | 2992 |

## Cài đặt thực nghiệm

…

# KẾT LUẬN

## Kết luận

## Hướng phát triển

# TÀI LIỆU THAM KHẢO

Tiếng Việt

…

Tiếng Anh

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762