

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN



BÙI VIỆT ANH  
NGUYỄN PHƯƠNG NAM  
TRẦN TRỌNG TRƯỜNG

**NGHIÊN CỨU PHƯƠNG PHÁP TỰ ĐỘNG PHÂN LOẠI  
BÀI VIẾT MẠNG XÃ HỘI BẰNG HỌC MÁY**

Chuyên ngành: Công nghệ phần mềm

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC**

Hà Nội - 2025

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN



BÙI VIỆT ANH  
NGUYỄN PHƯƠNG NAM  
TRẦN TRỌNG TRƯỜNG

**NGHIÊN CỨU PHƯƠNG PHÁP TỰ ĐỘNG PHÂN LOẠI  
BÀI VIẾT MẠNG XÃ HỘI BẰNG HỌC MÁY**

Chuyên ngành Công nghệ phần mềm

Giảng viên hướng dẫn: ThS. Lê Ngọc An

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC**

Hà Nội - 2025

ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập – Tự do – Hạnh phúc

*Hà Nội, ngày.... tháng.... năm 2025*

### NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

Họ và tên: Bùi Việt Anh

Giới tính: Nam

Ngày sinh: 01/05/2001

Chuyên ngành: Công nghệ phần mềm

Mã SV: 19A10010261

Lớp hành chính: 1910A02

Họ và tên: Nguyễn Phương Nam

Giới tính: Nam

Ngày sinh: 13/08/2002

Chuyên ngành: Công nghệ phần mềm

Mã SV: 20A10010115

Lớp hành chính: 2010A05

Họ và tên: Trần Trọng Trường

Giới tính: Nam

Ngày sinh: 08/11/2002

Chuyên ngành: Công nghệ phần mềm

Mã SV: 20A10010189

Lớp hành chính: 2010A05

### 1. TÊN ĐỀ TÀI

Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng  
học máy

## **2. NHIỆM VỤ VÀ NỘI DUNG**

Nhiệm vụ của đồ án tốt nghiệp:

- Xây dựng một hệ thống có khả năng tự động phân loại bài viết trên mạng xã hội theo chủ đề dựa trên các phương pháp học máy, giúp giảm thiểu công sức phân loại thủ công và nâng cao độ chính xác.
- Cung cấp một mô hình học máy có thể áp dụng trên các nền tảng mạng xã hội để nhận diện nội dung, phân loại bài viết theo các nhóm chủ đề như tin tức, giải trí, giáo dục, quảng cáo, v.v. Hệ thống bao gồm các bước thu thập, tiền xử lý dữ liệu, huấn luyện và đánh giá mô hình.
- Triển khai mô hình thực nghiệm, đánh giá độ chính xác và hiệu suất của hệ thống trên tập dữ liệu thực tế, đồng thời đề xuất hướng cải thiện để nâng cao chất lượng phân loại bài viết.

**3. NGÀY GIAO NHIỆM VỤ: ..../..../2025**

**4. NGÀY HOÀN THÀNH NHIỆM VỤ: ..../..../2025**

**5. PHÂN CÔNG CÔNG VIỆC:**

| Công việc   | Người thực hiện   |
|---|-------------------|
| Xây dựng hệ thống tích hợp hai model sử dụng website, xây dựng dataset    | Bùi Việt Anh      |
| Xây dựng mô hình xử lý văn bản phoBERT, xây dựng dataset                  | Nguyễn Phương Nam |
| Xây dựng mô hình sử dụng các kỹ thuật crawl đưa vào bài, xây dựng dataset | Trần Trọng Trường |

**6. GIẢNG VIÊN HƯỚNG DẪN: ThS. Lê Ngọc An**

Nội dung và đề cương Đồ án đã được Hội đồng chuyên ngành thông qua.

Hà Nội, ngày.....tháng.....năm 2025

**GIẢNG VIÊN HƯỚNG DẪN**

**KHOA CÔNG NGHỆ THÔNG TIN**

  
Lê Ngọc An

## PHÂN CÔNG NHIỆM VỤ ĐỒ ÁN

| THÀNH VIÊN        | PHÂN CHIA CÔNG VIỆC   |
|-------------------|---|
| Bùi Việt Anh      | <ul style="list-style-type: none"><li>- Tổng quan về tin giả mạo</li><li>- Nghiên cứu các mô hình học máy và học sâu</li><li>- Làm dataset 7000 bài viết trên tổng 21000 bài viết</li><li>- Mockups</li><li>- Lập trình frontend bằng React JS</li><li>- Tích hợp giữa frontend và backend</li><li>- Kiểm thử và sửa lỗi</li></ul>  |
| Nguyễn Phương Nam | <ul style="list-style-type: none"><li>- Tổng quan về tin giả mạo</li><li>- Nghiên cứu các mô hình học máy và học sâu</li><li>- Xác định các tiêu chí và đặc trưng</li><li>- Làm dataset 7000 bài viết trên tổng 21000 bài viết</li><li>- Nghiên cứu mô hình xử lý văn bản bằng PhoBERT và chức năng phân loại bài viết</li><li>- Thiết kế hệ thống và vẽ sơ đồ của hệ thống</li><li>- Lập trình Backend bằng Fast API</li><li>- Tích hợp giữa frontend và backend</li><li>- Kiểm thử và sửa lỗi</li></ul> |

Trần Trọng Trường

- Tổng quan về tin giả mạo
- Nghiên cứu các mô hình học máy và học sâu
- Xác định các tiêu chí và đặc trưng
- Làm dataset 7000 bài viết trên tổng 21000 bài viết
- Mockups
- Code chức năng crawl bài viết từ fanpage facebook
- Kiểm thử và sửa lỗi

## LỜI NÓI ĐẦU

Trong kỷ nguyên số hóa, mạng xã hội đã trở thành một phần không thể thiếu trong đời sống hàng ngày, nơi hàng tỷ người dùng chia sẻ thông tin, ý kiến và quan điểm. Tuy nhiên, sự gia tăng nhanh chóng của khối lượng dữ liệu trên các nền tảng này cũng đi kèm với thách thức lớn: sự lan truyền của thông tin sai lệch, tin giả, và nội dung không phù hợp. Những vấn đề này không chỉ ảnh hưởng đến nhận thức của cộng đồng mà còn đặt ra yêu cầu cấp thiết về việc phát triển các giải pháp công nghệ để quản lý và phân loại nội dung một cách hiệu quả. Trong bối cảnh đó, việc tự động phân loại bài viết mạng xã hội bằng các phương pháp học máy đã trở thành một hướng nghiên cứu đầy triển vọng, mang lại giá trị khoa học và thực tiễn cao.

Đề tài "Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy" được thực hiện nhằm mục tiêu xây dựng một hệ thống có khả năng thu thập, xử lý và phân loại bài viết từ các nền tảng mạng xã hội. Hệ thống hướng đến việc nhận diện các loại nội dung khác nhau, chẳng hạn như tin tức chính thống, thông tin sai lệch về xã hội, hay các bài viết có tính chất nhạy cảm về chính trị. Thông qua việc ứng dụng các kỹ thuật học máy và xử lý ngôn ngữ tự nhiên, đề tài không chỉ góp phần nâng cao hiệu quả quản lý thông tin mà còn hỗ trợ các tổ chức, cơ quan truyền thông, và cộng đồng trong việc kiểm soát chất lượng nội dung trực tuyến.

Là sinh viên chuyên ngành Công nghệ Phần mềm, tôi nhận thức rõ tầm quan trọng của việc kết hợp giữa phát triển phần mềm và các phương pháp học máy để giải quyết các bài toán thực tiễn. Đề tài này là cơ hội để tôi vận dụng kiến thức về thiết kế hệ thống, lập trình, và quản lý dữ liệu, đồng thời khám phá các kỹ thuật học máy hiện đại. Hệ thống được xây dựng với mục tiêu đảm bảo tính chính xác, hiệu quả, và khả năng ứng dụng trong thực tế, từ đó đóng góp một giải pháp có ý nghĩa cho cộng đồng và lĩnh vực công nghệ thông tin.

Tôi xin gửi lời cảm ơn sâu sắc đến Thầy Lê Ngọc An, giảng viên hướng dẫn của tôi, người đã tận tình chỉ bảo, định hướng và hỗ trợ tôi trong suốt quá

trình thực hiện đồ án. Những góp ý quý báu và sự động viên của thầy là nguồn động lực lớn để tôi hoàn thiện đề tài. Bên cạnh đó, tôi cũng xin bày tỏ lòng biết ơn đến nhà trường, các thầy cô trong khoa Công nghệ Thông tin, cùng bạn bè và gia đình, những người đã luôn đồng hành và tạo điều kiện thuận lợi cho tôi trong hành trình này.

Với tinh thần học hỏi và mong muốn đóng góp vào sự phát triển của chuyên ngành Công nghệ Phần mềm, tôi hy vọng rằng đồ án này sẽ mang lại những giá trị thiết thực, đồng thời nhận được những ý kiến đóng góp quý giá từ Hội đồng để ngày càng hoàn thiện hơn. Tôi xin chân thành cảm ơn!

**Sinh viên thực hiện**

Bùi Việt Anh

Nguyễn Phương Nam

Trần Trọng Trường

## MỤC LỤC

|   |    |
|---|----|
| DANH MỤC TỪ VIẾT TẮT.....                                       | 11 |
| CHƯƠNG 1. TỔNG QUAN VỀ PHÁT HIỆN TIN GIẢ MẠO .....              | 1  |
| 1.1. Tổng quan về tin giả mạo và mối nguy hiểm .....            | 1  |
| 1.2. Thực trạng tin giả thế giới.....                           | 1  |
| 1.3. Thực trạng tin giả Việt Nam .....                          | 2  |
| 1.4. Các bài báo nghiên cứu khoa học liên quan .....            | 4  |
| 1.4.1. Bài báo khoa học 1.....                                  | 4  |
| 1.4.2. Bài báo khoa học 2.....                                  | 5  |
| 1.4.3. Bài báo khoa học 3.....                                  | 6  |
| 1.5. Kết luận chương.....                                       | 7  |
| CHƯƠNG 2: CƠ SỞ LÝ THUYẾT .....                                 | 8  |
| 2.1. Các phương pháp truyền thống trong phát hiện tin giả ..... | 8  |
| 2.1.1. Giới thiệu .....   | 8  |
| 2.1.2. Các phương pháp phát hiện tin giả truyền thống.....      | 8  |
| 2.1.3. Ưu điểm .....  | 11 |
| 2.1.4. Nhược điểm.....  | 11 |
| 2.2. Các phương pháp phát hiện tin giả bằng học máy.....        | 11 |
| 2.2.1. Học Máy Có Giám Sát (Supervised Learning).....           | 11 |
| 2.2.2 Học Máy Không Giám Sát (Unsupervised Learning) .....      | 17 |
| 2.3. Các phương pháp phát hiện tin giả bằng học sâu .....       | 21 |
| 2.3.1 . Model BERT .....  | 21 |
| 2.3.2. RoBERTa.....   | 25 |
| 2.3.3. PhoBERT .....  | 26 |
| 2.4. So sánh PhoBERT, BERT và RoBERTa.....                      | 28 |
| 2.5. Kết luận chương hai .....                                  | 29 |
| CHƯƠNG 3. TRIỂN KHAI HỆ THỐNG .....                             | 31 |
| 3.1. Thu thập thông tin và thiết kế hệ thống .....              | 31 |
| 3.1.1. Triển khai khảo sát, thu thập thông tin .....            | 31 |

|  |    |
|--|----|
| 3.1.2. Sơ đồ hệ thống .....  | 50 |
| 3.1.3. Kiến trúc mô hình .....   | 52 |
| 3.1.4. Biểu đồ phân cấp chức năng - BFD và sơ đồ luồng dữ liệu - DFD ..... | 54 |
| 3.1.5. Sơ đồ quan hệ dữ liệu - ERD .....                                   | 60 |
| 3.1.6 CÁC BẢN THIẾT KẾ GIAO DIỆN .....                                     | 62 |
| 3.2. Triển khai hệ thống .....   | 64 |
| 3.2.1 Chức năng Phân loại bài viết .....                                   | 64 |
| 3.2.2 So sánh hiệu suất với các mô hình liên quan .....                    | 69 |
| 3.2.3 Chức năng crawl dữ liệu từ bài viết .....                            | 70 |
| 3.3. Kết quả hệ thống .....  | 71 |
| 3.3.1. Giao diện AI & Màn hình (Người dùng - Cuối) .....                   | 71 |
| 3.3.2. Giao diện Admin panel - thống kê .....                              | 73 |
| 3.3.3. Giao diện Admin panel - danh sách các bài viết .....                | 73 |
| 3.3.4. Giao diện Admin panel - crawl và lọc .....                          | 75 |
| TÀI LIỆU THAM KHẢO .....   | 78 |
| PHỤ LỤC .....  | 80 |

### DANH MỤC TỪ VIẾT TẮT

| STT | Tên viết tắt | Tên đầy đủ  | Dịch ra tiếng Việt                |
|-----|--------------|---|-----------------------------------|
| 1   | CNN          | Convolutional Neural Network                            | Mạng Nơ-ron Tích Chập             |
| 2   | RNN          | Recurrent Neural Network                                | Mạng Nơ-ron Hồi Quy               |
| 3   | LSTM         | Long Short-Term Memory                                  | Bộ Nhớ Dài Ngắn Hạn               |
| 4   | SVM          | Support Vector Machine                                  | Máy Vector Hỗ Trợ                 |
| 5   | BERT         | Bidirectional Encoder Representations from Transformers | Mã hóa hai chiều từ Bộ biến đổi   |
| 6   | KNN          | K-Nearest Neighbors                                     | là một thuật toán học có giám sát |
| 7   | NLP          | Natural Language Processing                             | Xử lý Ngôn ngữ Tự nhiên           |
| 8   | CLS          | Classification Token                                    | Vector Biểu diễn của Câu          |
| 9   | RoBERTa      | Robustly Optimized BERT approach                        |                                   |
| 10  | AI           | Artificial Intelligence                                 | Trí tuệ nhân tạo                  |
| 11  | OCR          | Optical Character Recognition                           | Nhận dạng ký tự quang học         |
| 12  | HTTP         | Hypertext Transfer Protocol                             | Giao thức truyền siêu văn bản     |
| 15  | MLM          | Masked Language Model                                   | Mô hình ngôn ngữ ẩn               |
| 16  | SEP          | Separation Token  |                                   |
| 17  | NSP          | Next Sentence Prediction                                | Dự đoán câu tiếp theo             |

## **DANH MỤC HÌNH ẢNH**

|  |    |
|--|----|
| Hình 1.1: Bài báo khoa học 1.....  | 4  |
| Hình 1.2: Bài báo khoa học 2.....  | 5  |
| Hình 1.3: Bài báo khoa học 3.....  | 6  |
| Hình 2.1 : Công thức Naive Bayes .....   | 12 |
| Hình 2.2 : Công thức tối ưu hóa lè.....  | 13 |
| Hình 2.1: Kiến trúc BERT based và BERT large .....   | 22 |
| Hình 2.2: Toàn bộ tiến trình pre-training và fine-tuning của BERT .....                              | 23 |
| Hình 2.3: Sơ đồ kiến trúc BERT cho nhiệm vụ MLM .....  | 23 |
| Hình 3.1: Bộ dữ liệu đã được xây dựng trên PostgreSQL .....  | 49 |
| Hình 3.2: Sơ đồ hệ thống .....   | 50 |
| Hình 3.3: Sơ đồ kiến trúc hệ thống .....   | 52 |
| Hình 3.4: Biểu đồ phân cấp chức năng - BFD.....  | 55 |
| Hình 3.5: Ký hiệu sử dụng .....  | 56 |
| Hình 3.6: Sơ đồ mức ngữ cảnh - DFD level 0 .....   | 56 |
| Hình 3.7: Sơ đồ luồng dữ liệu mức đỉnh - DFD level 1 phân rã từ DFD level 0.....                     | 57 |
| Hình 3.8: Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phân rã từ (1) thu thập dữ liệu.....       | 58 |
| Hình 3.9 : Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phân rã từ (2) xử lý dữ liệu .....        | 59 |
| .....  | 60 |
| Hình 3.10: Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phân rã từ (3) Hiển thị và thống kê ..... | 60 |
| .....  | 60 |
| Hình 3.11: Cơ sở dữ liệu - hệ thống crawl bài tự động .....  | 60 |
| Hình 3.12: Cơ sở dữ liệu - hệ thống huấn luyện model AI .....  | 61 |
| Hình 3.13: Cơ sở dữ liệu - hệ thống dashboard hiển thị cho người dùng ...                            | 61 |
| Hình 3.14: Giao diện AI và Màn hình ( Người dùng - Cuối ) .....                                      | 62 |
| Hình 3.15: Giao diện Admin panel - thống kê.....   | 62 |

|  |    |
|--|----|
| Hình 3.16: Giao diện Admin panel - danh sách các bài viết .....                | 63 |
| Hình 3.17: Giao diện Admin panel - crawl và lọc .....                          | 63 |
| Hình 3.18: Các chỉ số đánh giá .....   | 66 |
| Hình 3.19: Công thức Accuracy .....  | 67 |
| Hình 3.20 : Công thức Percision.....   | 67 |
| Hình 3.21: Công thức Recall.....   | 68 |
| Hình 3.22: Công thức F1 Score .....  | 68 |
| Hình 3.23: Kết quả đánh giá Dataset của hệ thống.....                          | 68 |
| Hình 3.24: Kết quả đánh giá Dataset của hệ thống.....                          | 72 |
| Hình 3.25: Kết quả thống kê của Admin Panel .....                              | 73 |
| Hình 3.26: Kết quả danh sách bài viết Admin Panel.....                         | 74 |
| Hình 3.27: Kết quả giao diện nhập URL của người dùng .....                     | 76 |
| Hình 3.28: Kết quả crawl của Admin Panel .....                                 | 76 |
| Hình 3.29: Kết quả lọc các bài viết URL fanpage Facebook của Admin Panel ..... | 77 |

## **DANH MỤC BẢNG**

|   |    |
|---|----|
| Bảng 2.1 : Bảng so sánh các thuật toán học có giám sát .....  | 16 |
| Bảng 2.2 : So sánh K-means và DBSCAN.....                     | 21 |
| Bảng 2.3: Bảng so sánh các mô hình học sâu.....               | 29 |
| Bảng 3.1: Bảng gom nhóm chức năng.....                        | 54 |
| Bảng 3.2: Bảng thuật toán trong mạng nơ ron của PhoBERT ..... | 66 |

## CHƯƠNG 1. TỔNG QUAN VỀ PHÁT HIỆN TIN GIẢ MẠO

### 1.1. Tổng quan về tin giả mạo và mối nguy hiểm

Tin giả mạo là những thông tin sai lệch, chưa được kiểm chứng hoặc bị bóp méo nhằm gây hiểu lầm. Chúng lan truyền nhanh trên mạng xã hội, ảnh hưởng tiêu cực đến cá nhân, xã hội và quốc gia.

#### Mối Nguy Hiểm Của Tin Giả

- Gây hoang mang, mất niềm tin: Tin giả làm nhiễu loạn thông tin, khiến người dân khó phân biệt thật - giả.
- Ảnh hưởng an ninh, kinh tế: Gây bất ổn xã hội, làm suy yếu nền kinh tế, doanh nghiệp có thể bị tổn hại.
- Tác động chính trị: Tin giả có thể xuyên tạc, bôi nhọ lãnh đạo, kích động xung đột.
- Gây nguy hại sức khỏe cộng đồng: Tin giả y tế dẫn đến các hành vi sai lầm, nguy hiểm.
- Chịu trách nhiệm pháp lý: Đăng tải, lan truyền tin giả có thể bị xử phạt hành chính hoặc truy cứu trách nhiệm hình sự.

### 1.2. Thực trạng tin giả thế giới

Theo báo cáo của UNESCO (2019) và Liên minh Châu Âu (EU), tin giả đang trở thành mối đe dọa toàn cầu, đặc biệt trong không gian mạng. Một số vấn đề chính bao gồm:

- Lan truyền mạnh trên mạng xã hội: Facebook, YouTube, TikTok là môi trường phát tán tin giả nhanh chóng, gây hoang mang dư luận (UNESCO, 2019).
- Tác động đến chính trị: Tin giả được sử dụng để thao túng bầu cử, làm mất niềm tin vào chính phủ (EU, 2019).

- Gây hại sức khỏe cộng đồng: Tổ chức Y tế Thế giới (WHO) cảnh báo tin giả về COVID-19, vắc-xin đã dẫn đến nhiều hiểu lầm nguy hiểm (WHO, 2020).
- Thiệt hại kinh tế: Theo báo cáo của Forbes (2022), tin giả gây tổn thất hàng tỷ USD mỗi năm do ảnh hưởng tiêu cực đến doanh nghiệp.

### **Cơ Sở Nhận Biết Và Kiểm Chứng Tin Giả**

Theo Cẩm nang phòng chống tin giả của Bộ Thông tin và Truyền thông Việt Nam (2022), để nhận diện tin giả, cần dựa vào:

- Kiểm tra nguồn tin: Đối chiếu với báo chí chính thống, các tổ chức uy tín.
- Kiểm tra hình ảnh, video: Sử dụng công cụ tìm kiếm hình ảnh ngược để phát hiện chỉnh sửa.
- Kiểm tra thời gian, sự kiện: Xác minh thông tin phản ánh đúng bối cảnh thực tế.
- Tham khảo chuyên gia, cơ quan chức năng: Đặc biệt với tin tức về y tế, chính trị.
- Không chia sẻ thông tin chưa kiểm chứng: Tránh tiếp tay lan truyền tin giả.

Việc nâng cao nhận thức và có trách nhiệm khi tiếp nhận, chia sẻ thông tin là biện pháp quan trọng để ngăn chặn tác hại của tin giả

### **1.3. Thực trạng tin giả Việt Nam**

Theo báo cáo của Bộ Thông tin và Truyền thông (2022) và Trung tâm Xử lý tin giả Việt Nam (VAFC), tình trạng tin giả tại Việt Nam đang diễn biến phức tạp, đặc biệt trên các nền tảng mạng xã hội. Một số vấn đề chính gồm:

- Lan truyền nhanh chóng trên mạng xã hội: Facebook, TikTok, YouTube là những kênh phát tán tin giả phổ biến, gây hoang mang dư luận (Bộ Thông tin và Truyền thông, 2022).

- Ảnh hưởng đến an ninh quốc gia: Các đối tượng xấu sử dụng tin giả để xuyên tạc, kích động, gây mất ổn định xã hội (VAFC, 2023).
- Tác động tiêu cực đến kinh tế: Theo báo cáo của VCCI (2022), nhiều doanh nghiệp bị ảnh hưởng nặng nề bởi tin giả, làm giảm uy tín thương hiệu và gây thiệt hại tài chính.
- Gây nguy hại sức khỏe cộng đồng: Tin giả về dịch COVID-19, thực phẩm chức năng tràn lan, gây hiểu lầm trong dư luận, ảnh hưởng đến sức khỏe người dân (Bộ Y tế, 2021).

### **Cơ Sở Nhận Biết Và Kiểm Chứng Tin Giả**

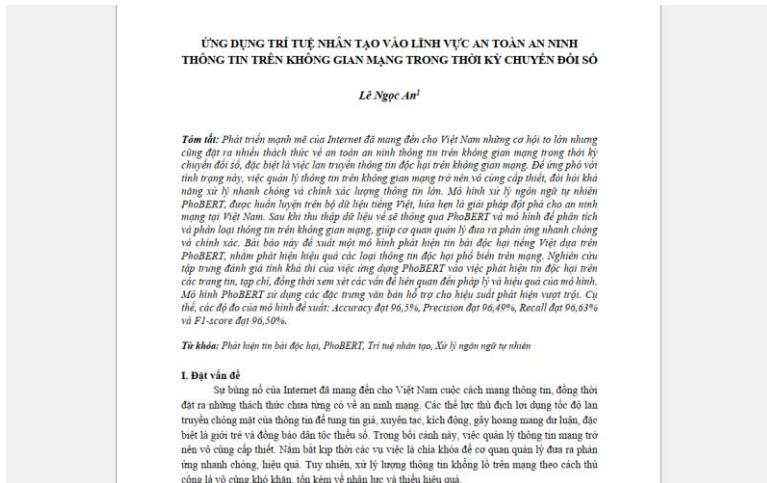
Theo Cẩm nang phòng chống tin giả của Bộ Thông tin và Truyền thông (2022) [4], để nhận diện tin giả, cần dựa vào:

- Kiểm tra nguồn tin: Đối chiếu với báo chí chính thống, cơ quan chức năng.
- Kiểm tra hình ảnh, video: Sử dụng công cụ tìm kiếm hình ảnh ngược để phát hiện nội dung bị chỉnh sửa.
- Kiểm tra thời gian, sự kiện: Xác minh xem tin tức có phù hợp với bối cảnh thực tế không.
- Tham khảo ý kiến chuyên gia, cơ quan chức năng: Đối với tin tức y tế, kinh tế, chính trị.
- Không chia sẻ thông tin chưa được kiểm chứng: Tránh tiếp tay lan truyền tin giả, gây hậu quả nghiêm trọng.

Nhận thức đúng và trách nhiệm khi tiếp nhận thông tin là yếu tố quan trọng để phòng chống tin giả tại Việt Nam.

## 1.4. Các bài báo nghiên cứu khoa học liên quan

### 1.4.1. Bài báo khoa học 1



Hình 1.1: Bài báo khoa học 1

Tên bài báo: *Ứng dụng trí tuệ nhân tạo vào lĩnh vực an toàn an ninh thông tin trên không gian mạng trong thời kỳ chuyển đổi số*.

Tác giả: ThS. Lê Ngọc An.

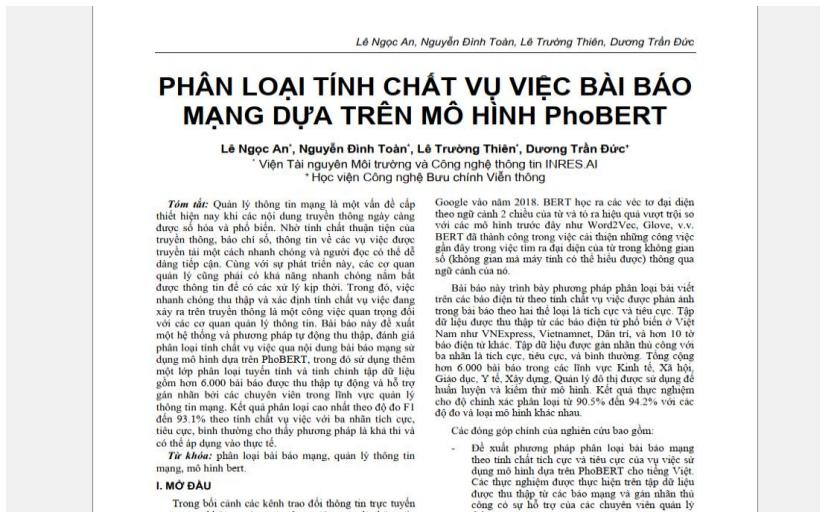
Bài báo khoa học này tập trung vào bài toán phát hiện tin độc hại bằng các phương pháp xử lý ngôn ngữ tự nhiên (NLP), đặc biệt là việc ứng dụng mô hình PhoBERT để từ đó cải thiện độ chính xác trong phân loại tin tức. Nghiên cứu đã khảo sát các tiếp cận trước đó trong lĩnh vực phát hiện tin độc hại, bao gồm các mô hình machine learning truyền thống như Naïve Bayes, SVM, Decision Tree và các mô hình deep learning như LSTM và CNN.

Mô hình PhoBERT, một biến thể của BERT dành riêng cho tiếng Việt, được sử dụng để mã hóa các đặc trưng ngôn ngữ từ tin tức. Dữ liệu sử dụng trong bài báo bao gồm tài nguyên ngôn ngữ lớn và bộ dữ liệu gắn nhãn cho tin độc hại. Các phương pháp tiền xử lý dữ liệu như loại bỏ ký tự đặc biệt, chuẩn hóa dữ liệu và tokenization đã được áp dụng để cải thiện độ chính xác.

Kết quả nghiên cứu chỉ ra rằng PhoBERT vượt trội so với các phương pháp trước đó về độ chính xác, precision, recall và F1-score. Mô hình này giúp giảm thiểu tình trạng dương tính giả và âm tính giả, từ đó đề xuất khả năng ứng dụng trong thực tế.

Tổng kết lại, nghiên cứu này góp phần cung cấp tầm quan trọng của việc ứng dụng mô hình tiền huấn luyện như PhoBERT trong việc phát hiện tin độc hại tiếng Việt. Nghiên cứu đề xuất hướng phát triển tiếp theo là kết hợp với các mô hình phân tích ngôn ngữ khác và ứng dụng trong hệ thống thực tế như các trang web, mạng xã hội.

#### 1.4.2. Bài báo khoa học 2



Hình 1.2: Bài báo khoa học 2

Tên bài báo: Phân loại tính chất vụ việc bài báo mạng dựa trên mô hình PhoBERT. [7]

Tác giả: ThS. Lê Ngọc An, Nguyễn Đình Toàn, Lê Trường Thiên, Dương Trần Đức.

Hiện nay, với sự lan truyền nhanh chóng của tin tức trên mạng xã hội, việc phát hiện và phân loại tin giả trở thành một thách thức lớn. Các phương

pháp truyền thống như SVM , Random Forest, Naïve Bayes đã được áp dụng, nhưng các nghiên cứu cho thấy PhoBERT có khả năng trích xuất đặc trưng ngữ nghĩa tốt hơn, giúp tăng độ chính xác trong phân loại tin thật - giả.

Ngoài ra, việc kết hợp PhoBERT với các mô hình phân loại như SVM , Logistic Regression đã cho kết quả khả quan. Một số nghiên cứu còn áp dụng phương pháp phân cụm dữ liệu, giúp xác định chủ đề tin giả và hiểu rõ hơn về xu hướng lan truyền.

Tuy nhiên, việc phát hiện tin giả vẫn gặp nhiều thách thức, như sự thay đổi liên tục trong cách tạo dựng tin giả và mất cân bằng dữ liệu huấn luyện. Do đó, các nghiên cứu đang tập trung vào tăng cường dữ liệu và phát triển mô hình thích ứng tốt hơn.

Nghiên cứu này góp phần quan trọng trong việc nâng cao hiệu quả phân loại bài báo mạng, hỗ trợ phát hiện tin giả và giảm thiểu tác động tiêu cực của thông tin sai lệch.

### **1.4.3. Bài báo khoa học 3**

ISSN 1859-1531 - TẠP CHÍ KHOA HỌC VÀ CÔNG NGHỆ - ĐẠI HỌC ĐÀ NẴNG, VOL. 20, NO. 3, 2022

71

#### **PHÁT HIỆN TỰ ĐỘNG TIN GIẢ: THÀNH TỰU VÀ THÁCH THỨC** AUTOMATIC FAKE NEWS DETECTION: ACHIEVEMENTS AND CHALLENGES

**Võ Trung Hùng<sup>1\*</sup>, Ninh Khánh Chi<sup>2</sup>, Trần Anh Kiệt<sup>3</sup>**

<sup>1</sup>Trường Đại học Sư phạm Kỹ thuật - Đại học Đà Nẵng  
<sup>2</sup>Trường Đại học CNTT & Truyền thông Việt-Hàn - Đại học Đà Nẵng  
<sup>3</sup>Đại học Đà Nẵng

\*Tác giả liên hệ: vthung@ute.udn.vn

(Nhận bài: 06/01/2022; Chấp nhận đăng: 27/02/2022)

**Tóm tắt** - Trong bài báo này, nhóm tác giả trình bày một cách tổng quan các vấn đề liên quan đến khái niệm, phân loại, cách xác định thủ công và xác định tự động các tin giả. Đặc biệt, nhóm tác giả đã trình bày hai kỹ thuật được ứng dụng rộng rãi hiện nay đó là kỹ thuật học máy và kỹ thuật học sâu. Hai kỹ thuật này đều dựa trên phân tích nội dung bản tin và bước đầu đã mang lại những kết quả tích cực. Tuy nhiên, đây là bài báo mang tính chất nghiên cứu tổng quan nên nhóm tác giả chỉ dừng ở mức tổng hợp, phân tích, nhận định và trình bày lại những kết quả nghiên cứu đã có trước đó. Đồng góp chính trong bài báo này là chỉ ra được những thách thức và hướng nghiên cứu sắp đến cho tiếng Việt trong lĩnh vực phát hiện tin giả.

**Từ khóa** - Tin giả; phát hiện tự động; mạng nơ-ron; học máy; học sâu

**Abstract** - In this paper, the authors present an overview of issues related to the concept, classification, manual detection and automatic detection of fake news. In particular, the authors present two widely applied techniques today: Traditional machine learning and deep learning. These two techniques are based on content analysis and initially offered positive results. However, this article is of an overview research, therefore we only stop at the level of synthesizing, analyzing, commenting and presenting previous research results. Our main contribution in this paper is to point out the challenges and upcoming research directions for Vietnamese in the field of fake news detection.

**Key words** - Fake news; automatic detection; neural networks; machine learning; deep learning

#### *Hình 1.3: Bài báo khoa học 3*

Tên bài báo: Phát hiện tự động tin giả: thành tựu và thách thức, 2022

Tác giả: Võ Trung Hùng, Ninh Khánh Chi , Trần Anh Kiệt

Bài báo "Phát hiện tự động tin giả: Thành tựu và thách thức" (2022) của Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng, trình bày tổng quan về các phương pháp phát hiện tin giả, bao gồm phương pháp thủ công và phương pháp tự động. Các thuật toán học máy như Naïve Bayes, Decision Tree, SVM và KNN được sử dụng để phân loại tin tức dựa trên nội dung đã gán nhãn. Trong khi đó, các mô hình học sâu như CNN, RNN và LSTM giúp phân tích sâu hơn, cải thiện độ chính xác trong việc nhận diện tin giả. Bài báo cũng chỉ ra những thách thức trong việc phát hiện tin giả, đặc biệt với ngôn ngữ tiếng Việt, đồng thời đề xuất các hướng nghiên cứu trong tương lai nhằm nâng cao hiệu quả của các phương pháp tự động.

### **1.5. Kết luận chương**

Việc ứng dụng trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) giúp phát hiện tin giả hiệu quả, trong đó PhoBERT đóng vai trò quan trọng nhờ khả năng xử lý ngữ nghĩa chính xác.

PhoBERT đạt hiệu suất cao trong phân loại văn bản, nhận diện thực thể (NER) và trả lời câu hỏi, hỗ trợ tốt cho phát hiện tin giả, phân tích cảm xúc và chatbot.

Tuy nhiên, thách thức vẫn còn như sự thay đổi của tin giả và mất cân bằng dữ liệu. Do đó, cần mở rộng dữ liệu, cải tiến mô hình và kết hợp nhiều phương pháp để nâng cao độ chính xác.

Nghiên cứu này góp phần quan trọng trong bảo vệ không gian mạng và phát triển AI trong xử lý tiếng Việt.

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

### 2.1. Các phương pháp truyền thông trong phát hiện tin giả

#### 2.1.1. Giới thiệu

Phương pháp truyền thông trong phát hiện tin giả chủ yếu dựa vào quá trình kiểm duyệt thủ công và xác minh nguồn tin thông qua các tiêu chí xác thực do con người thực hiện. Không giống như các phương pháp tự động sử dụng công nghệ AI hay học máy, các phương pháp này tận dụng kinh nghiệm và khả năng phân tích của con người để đánh giá tính chính xác của thông tin. Dù đòi hỏi nhiều thời gian và công sức, chúng vẫn được đánh giá cao nhờ khả năng hiểu sâu sắc ngữ cảnh, sắc thái ngôn ngữ và ý đồ ẩn trong bài viết—những yếu tố mà công nghệ hiện tại đôi khi chưa thể nắm bắt trọn vẹn. Trong bối cảnh tin giả ngày càng phổ biến và phức tạp, các phương pháp truyền thông vẫn giữ vai trò quan trọng trong việc đảm bảo độ tin cậy của thông tin.

#### 2.1.2. Các phương pháp phát hiện tin giả truyền thống

Kiểm duyệt bởi chuyên gia

- Mô tả: Các tổ chức báo chí hoặc cơ quan kiểm duyệt tin tức thường sử dụng đội ngũ chuyên gia được đào tạo để phân tích và đánh giá nội dung tin tức. Những chuyên gia này không chỉ xem xét thông tin bề mặt mà còn đi sâu vào ngữ cảnh, cách diễn đạt và ý nghĩa tổng thể.
- Quy trình: Chuyên gia đối chiếu tin tức với các nguồn đáng tin cậy, kiểm tra tính logic và phát hiện các dấu hiệu bất thường trước khi đưa ra kết luận.
- Ví dụ: Một chuyên gia có thể nhận ra rằng một bài viết sử dụng ngôn ngữ phóng đại không phù hợp với sự kiện được đề cập, từ đó nghi ngờ tính xác thực.

So sánh với nguồn tin chính thống

- Mô tả: Phương pháp này liên quan đến việc đối chiếu nội dung tin tức với các báo chí chính thống hoặc trang tin chính thức (ví dụ: BBC, CNN, hoặc các tờ báo lớn tại Việt Nam như Tuổi Trẻ, Thanh Niên).

- Quy trình: Nếu thông tin không xuất hiện trên các nguồn uy tín, đặc biệt là trong các sự kiện quan trọng, thì cần đặt nghi vấn về tính xác thực.
- Ví dụ: Một tin tức về thiên tai lớn nhưng không được các hãng tin lớn xác nhận có thể là tin giả.

#### Kiểm tra nguồn gốc tin tức

- Mô tả: Việc xác minh danh tính của tác giả và đơn vị phát hành tin tức là bước quan trọng để đánh giá độ tin cậy.
- Quy trình: Nếu nguồn tin không rõ ràng (ví dụ: không có thông tin tác giả) hoặc có lịch sử đăng tải nội dung sai lệch, đó là dấu hiệu đáng nghi ngờ.
- Ví dụ: Một bài viết từ một trang web không tên tuổi, không có thông tin liên hệ rõ ràng thường bị coi là thiếu uy tín.

#### Phân tích nội dung và dấu hiệu đáng ngờ

- Mô tả: Tin giả thường có những đặc điểm dễ nhận diện thông qua phân tích nội dung.
  - Dấu hiệu cụ thể:
    - Tiêu đề giật gân: Sử dụng ngôn ngữ gây sốc, kích động (ví dụ: "Cả thế giới sững sờ vì bí mật này!").
    - Lỗi ngôn ngữ: Nhiều lỗi chính tả, cú pháp lộn xộn hoặc cách viết thiếu chuyên nghiệp.
    - Thiếu dẫn chứng: Không trích dẫn nguồn cụ thể hoặc sử dụng thông tin mơ hồ, không kiểm chứng được.
- Ví dụ: Một bài viết với tiêu đề "Chữa khỏi ung thư trong 3 ngày" nhưng không có bằng chứng khoa học thường là tin giả.

#### Xác minh thông tin từ nhiều nguồn khác nhau

- Mô tả: Phương pháp này kiểm tra xem tin tức có được nhiều nguồn độc lập cùng xác nhận hay không.
- Quy trình: Nếu chỉ một trang hoặc một nhóm liên kết lan truyền thông tin, tính xác thực cần được xem xét kỹ lưỡng. Ngược lại, sự đồng thuận từ nhiều nguồn uy tín tăng độ tin cậy.

- Ví dụ: Tin đồn về một sự kiện chính trị chỉ xuất hiện trên một trang mạng xã hội mà không được báo chí chính thống đề cập có thể là giả mạo.

Dựa vào tổ chức kiểm chứng tin tức (Fact-Checking)

- Mô tả: Các tổ chức chuyên kiểm chứng như Snopes, FactCheck.org (quốc tế) hoặc các chuyên mục của báo chí Việt Nam (như VnExpress, Tuổi Trẻ) đóng vai trò quan trọng trong việc xác minh tin tức.
- Quy trình: Họ sử dụng các phương pháp chuyên sâu, kết hợp chuyên gia và dữ liệu để phân tích và công bố kết quả.
- Ví dụ: Một tin tức về "vaccine gây đột biến gen" có thể bị các tổ chức này bác bỏ dựa trên bằng chứng khoa học.

Kiểm tra thời gian và địa điểm của sự kiện

- Mô tả: Nhiều tin giả tái sử dụng thông tin cũ hoặc không liên quan, sau đó trình bày như tin mới.
- Quy trình: Kiểm tra ngày tháng, bối cảnh và địa điểm của sự kiện để đảm bảo thông tin phù hợp và không bị bóp méo.
- Ví dụ: Một bức ảnh từ trận lụt năm 2010 được dùng để minh họa cho lụt lội năm 2023 là dấu hiệu của tin giả.

### 2.1.3. Ưu điểm

- Độ chính xác cao: Khi được thực hiện bởi các chuyên gia có kinh nghiệm, phương pháp truyền thống mang lại kết quả đáng tin cậy nhờ khả năng phân tích chi tiết.
- Hiểu sắc thái và ý đồ: Con người có thể nhận diện các yếu tố tinh vi như giọng điệu, ẩn ý hoặc thao túng cảm xúc—những điều mà AI hiện tại khó làm được.
- Xử lý tin giả phức tạp: Với các tin giả được ngụy tạo khéo léo, phương pháp thủ công vẫn vượt trội hơn các thuật toán chưa hoàn thiện.

### 2.1.4. Nhược điểm

- Tốn thời gian và nhân lực: Việc kiểm duyệt thủ công không thể xử lý kịp thời lượng tin tức khổng lồ sinh ra mỗi ngày, đặc biệt trên mạng xã hội.
- Yếu tố chủ quan: Quan điểm cá nhân hoặc thiếu thông tin của người kiểm duyệt có thể dẫn đến sai sót trong đánh giá.
- Khó theo kịp tốc độ lan truyền: Tin giả thường lan nhanh hơn khả năng xác minh thủ công, khiến việc ngăn chặn trở nên kém hiệu quả.

## 2.2. Các phương pháp phát hiện tin giả bằng học máy

Học máy (Machine Learning) đã trở thành công cụ quan trọng trong việc tự động hóa quá trình phát hiện tin giả nhờ khả năng phân tích dữ liệu lớn và rút trích đặc trưng phức tạp. Các phương pháp này có thể được chia thành hai nhóm chính: Học có giám sát (Supervised Learning) và Học không giám sát (Unsupervised Learning). Trong phần này, chúng tôi tập trung vào các thuật toán học có giám sát phổ biến, bao gồm Naïve Bayes, SVM (Support Vector Machine) và Random Forest.

### 2.2.1. Học Máy Có Giám Sát (Supervised Learning)

Học có giám sát là phương pháp huấn luyện mô hình dựa trên tập dữ liệu đã được gắn nhãn (ví dụ: "tin thật" hoặc "tin giả"). Mô hình học cách ánh xạ các đặc trưng đầu vào (features) đến nhãn đầu ra (label) thông qua việc tối ưu hóa hàm mất mát. Quy trình này bao gồm hai giai đoạn chính:

- Huấn luyện: Mô hình học từ dữ liệu lịch sử.
- Dự đoán: Áp dụng mô hình đã huấn luyện để phân loại dữ liệu mới.

Các thuật toán tiêu biểu gồm Naïve Bayes, SVM và Random Forest, mỗi phương pháp có ưu nhược điểm riêng trong xử lý văn bản.

### 2.2.1.1 Mô hình Naive Bayes

Nguyên lý hoạt động:

Naïve Bayes dựa trên Định lý Bayes với giả định các đặc trưng độc lập với nhau (independence assumption). Công thức tính xác suất hậu nghiệm (posterior probability) của lớp C khi có đặc trưng X:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

*Hình 2.1 : Công thức Naive Bayes*

Trong đó:

- $P(C|X)$ : Xác suất hậu nghiệm của lớp C (tích cực hoặc tiêu cực) khi có đặc trưng X.
- $P(X|C)$ : Xác suất có điều kiện của đặc trưng X trong lớp C.
- $P(C)$ : Xác suất tiên nghiệm của lớp C.
- $P(X)$ : Xác suất của đặc trưng X (thường được bỏ qua khi so sánh các lớp).

Ví dụ:

Giả sử một bài viết chứa từ "khẩn cấp" ( X). Nếu  $P(\text{tin giả}|X) > P(\text{tin thật}|X)$ , mô hình sẽ phân loại bài viết này là tin giả.

#### Ưu điểm:

- Tốc độ cao: Huấn luyện và dự đoán nhanh, phù hợp với dữ liệu lớn.
- Đơn giản: Dễ triển khai, ít yêu cầu tinh chỉnh siêu tham số.

- Hiệu quả với văn bản: Hoạt động tốt trên dữ liệu text nhờ mô hình hóa tần suất từ (ví dụ: phân loại spam email).

#### Nhược điểm:

- Giả định độc lập không thực tế: Trong thực tế, các từ trong câu có quan hệ ngữ nghĩa (ví dụ: "khẩn cấp" thường đi kèm "lừa đảo").
- Không xử lý được đặc trưng tương quan: Hiệu suất giảm khi các đặc trưng phụ thuộc lẫn nhau.

Ứng dụng trong phát hiện tin giả:

Naïve Bayes được sử dụng để phân loại sơ bộ dựa trên từ khóa, ví dụ:

- Phát hiện tin giả y tế bằng cách đếm tần suất các từ như "miễn phí", "thần kỳ".
- Phân tích cảm xúc để xác định bài viết mang tính giật gân.

#### 2.2.1.2 Support Vector Machine (SVM)

Nguyên lý hoạt động:

SVM tìm siêu phẳng (hyperplane) tối ưu trong không gian đa chiều để phân tách hai lớp dữ liệu sao cho lề (margin) giữa chúng là lớn nhất. Các điểm dữ liệu gần siêu phẳng nhất được gọi là support vectors.

#### Công thức tối ưu:

$$\text{Tối đa hóa: } \frac{2}{\|\mathbf{w}\|} \quad \text{Điều kiện: } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall i$$

Hình 2.2 : Công thức tối ưu hóa lề

Trong đó:

- $\mathbf{w}$ : Vector pháp tuyến của siêu phẳng.
- $\mathbf{x}_i$ : Vector đặc trưng của văn bản thứ  $i$ .
- $y_i$ : Nhãn lớp của văn bản (+1 cho tích cực, -1 cho tiêu cực).
- $b$ : Hệ số chặn.
- $\|\mathbf{w}\|$ : Độ lớn của vector  $\mathbf{w}$ .

Ví dụ:

Với dữ liệu phi tuyến, SVM sử dụng kernel trick (ví dụ: kernel RBF) để ánh xạ dữ liệu sang không gian chiều cao hơn, giúp phân tách dễ dàng.

### Kết quả

Văn bản mới được phân loại dựa trên vị trí của nó so với siêu phẳng (bên nào của ranh giới).

### Giải thích

SVM rất mạnh mẽ trong phân loại văn bản với dữ liệu lớn và nhiều chiều, đặc biệt khi kết hợp với kernel (như kernel RBF) để xử lý dữ liệu phi tuyến tính. Ví dụ, SVM có thể phân biệt tốt giữa tin tích cực và tiêu cực dựa trên sự khác biệt rõ ràng trong từ vựng. Tuy nhiên, nó tồn tại nhiều tài nguyên tính toán và khó diễn giải hơn so với Naive Bayes hay Hồi quy Logistic.

### Ưu điểm

- Tự động hóa cao: Học máy giúp tự động phân loại hàng loạt tin tức mà không cần sự can thiệp thủ công.
- Tốc độ xử lý nhanh: Mô hình học máy có thể xử lý một lượng lớn văn bản trong thời gian ngắn, hiệu quả hơn so với phương pháp truyền thống.
- Độ chính xác cao hơn khi có dữ liệu tốt: Nếu được huấn luyện trên một tập dữ liệu lớn và chất lượng, các mô hình học máy có thể đạt độ chính xác cao trong việc phân loại tin tức.
- Linh hoạt: Có thể áp dụng trên nhiều loại ngôn ngữ và nguồn tin tức khác nhau.

### Nhược điểm

- Phụ thuộc vào chất lượng dữ liệu: Nếu tập dữ liệu huấn luyện không cân bằng hoặc chứa nhiều dữ liệu nhiễu, mô hình có thể cho kết quả sai lệch.
- Khả năng giải thích hạn chế: Một số thuật toán học máy, như SVM hay Random Forest, khó diễn giải được tại sao một bài báo lại được xếp vào loại tích cực hoặc tiêu cực.

- Cần nhiều công đoạn tiền xử lý: Việc làm sạch và chuẩn hóa dữ liệu văn bản trước khi huấn luyện mô hình có thể tốn nhiều thời gian.
- Không xử lý tốt các ngữ cảnh phức tạp: Các mô hình học máy truyền thống không hiểu được ngữ cảnh sâu sắc như các mô hình học sâu.

#### *2.2.1.3 Random Forest*

Nguyên lý hoạt động:

Random Forest là ensemble method kết hợp nhiều cây quyết định (Decision Trees) thông qua kỹ thuật Bootstrap Aggregating (Bagging). Mỗi cây được huấn luyện trên tập con ngẫu nhiên của dữ liệu và đặc trưng, kết quả cuối cùng dựa trên bỏ phiếu đa số (phân loại) hoặc trung bình (hồi quy).

Quy trình huấn luyện:

- Bootstrap Sampling: Tạo N tập con từ dữ liệu gốc bằng cách lấy mẫu có hoàn lại.
- Xây dựng cây: Mỗi cây được huấn luyện trên một tập con, chọn đặc trưng ngẫu nhiên tại mỗi nút.
- Tổng hợp kết quả: Dự đoán dựa trên kết quả của tất cả cây.

Ví dụ:

Giả sử 100 cây trong rừng, nếu 60 cây phân loại một bài viết là "tin giả", kết quả cuối cùng sẽ là "tin giả".

#### **Ưu điểm:**

- Giảm overfitting: Độ chính xác ổn định nhờ kết hợp nhiều cây.
- Xử lý đặc trưng phức tạp: Hiệu quả với dữ liệu có nhiều biến đầu vào.
- Không yêu cầu chuẩn hóa dữ liệu: Có thể làm việc trực tiếp với dữ liệu thô.

#### **Nhược điểm:**

- Tốn thời gian: Huấn luyện chậm khi số lượng cây lớn.
- Khó diễn giải: Mô hình phức tạp, khó hiểu cách ra quyết định.

- Nhạy cảm với nhiễu: Cần làm sạch dữ liệu trước khi huấn luyện.

#### **Ứng dụng thực tế:**

- Phân loại bài viết dựa trên kết hợp nhiều đặc trưng như metadata (tác giả, nguồn), từ khóa, và cấu trúc câu.
- Xử lý dữ liệu mêtô bằng cách điều chỉnh trọng số các lớp.

#### *2.2.1.4 So sánh các thuật toán học có giám sát*

| Tiêu chí            | Naive Bayes        | SVM                  | Random Forest        |
|---------------------|--------------------|----------------------|----------------------|
| Tốc độ              | Rất nhanh          | Chậm với dữ liệu lớn | Trung bình           |
| Xử lý nhiễu         | Kém                | Trung bình           | Tốt                  |
| Giải thích mô hình  | Dễ Dàng            | Khó                  | Khó                  |
| Hiệu suất phi tuyến | Kém ( cần Kernel ) | Tốt ( với Kernel )   | Tốt                  |
| Ứng dụng điển hình  | Phân loại sơ bộ    | Phân lớp chính xác   | Dữ liệu đa đặc trưng |

Bảng 2.1 : Bảng so sánh các thuật toán học có giám sát

Các phương pháp học có giám sát đều có điểm mạnh và hạn chế riêng:

- Naïve Bayes phù hợp cho hệ thống yêu cầu tốc độ cao nhưng dữ liệu đơn giản.
- SVM cho độ chính xác cao ở dữ liệu phi tuyến nhưng đòi hỏi tài nguyên tính toán.
- Random Forest linh hoạt với dữ liệu phức tạp nhưng khó tối ưu.

Việc lựa chọn thuật toán phụ thuộc vào đặc điểm dữ liệu, yêu cầu tốc độ và khả năng giải thích mô hình. Trong các nghiên cứu gần đây, SVM và Random

Forest thường được kết hợp với kỹ thuật trích xuất đặc trưng sâu (như BERT) để nâng cao hiệu suất

### **2.2.2 Học Máy Không Giám Sát (Unsupervised Learning)**

Học không giám sát tập trung vào việc khám phá cấu trúc ẩn trong dữ liệu không có nhãn, giúp phát hiện các mẫu bất thường hoặc nhóm dữ liệu tự nhiên. Trong phát hiện tin giả, phương pháp này hữu ích để nhận diện nội dung không phù hợp thông qua phân cụm (clustering) hoặc phát hiện ngoại lai (anomaly detection). Hai thuật toán phổ biến là K-means và DBSCAN, mỗi phương pháp có ưu điểm riêng biệt.

#### **2.2.2.1 K-means**

*K-means là một thuật toán phân cụm không giám sát (unsupervised clustering) phổ biến, chia dữ liệu thành K nhóm dựa trên sự tương đồng về khoảng cách. Mục tiêu là tối ưu hóa sao cho các điểm trong cùng cụm gần nhau nhất và cách xa các cụm khác.*

#### **Cơ chế hoạt động:**

- Khởi tạo: Chọn ngẫu nhiên K điểm làm tâm cụm (centroids) ban đầu (thường dùng phương pháp K-means++ để cải thiện khởi tạo).
- Gán cụm: Tính khoảng cách Euclidean từ mỗi điểm dữ liệu đến tất cả các tâm cụm, sau đó gán điểm đó vào cụm có tâm gần nhất.
- Cập nhật tâm cụm: Tính lại tâm cụm mới bằng cách lấy trung bình tọa độ của tất cả các điểm trong cụm đó.
- Lặp lại: Quay lại bước 2 và 3 cho đến khi tâm cụm không thay đổi nữa (hoặc đạt số lần lặp tối đa).

#### **Ưu điểm:**

- Tốc độ nhanh: Thuật toán đơn giản, tính toán hiệu quả với dữ liệu lớn nếu K không quá lớn.
- Dễ triển khai: Yêu cầu ít tham số (chỉ cần chọn K) và dễ hiểu.

- Hiệu quả với dữ liệu hình cầu: Hoạt động tốt khi các cụm có dạng tròn hoặc phân bố đồng đều.

#### **Nhược điểm:**

- Phải chọn K trước: Việc xác định số cụm tối ưu (K) là thách thức lớn, thường cần thử nghiệm hoặc dùng phương pháp như Elbow Method.
- Nhạy cảm với ngoại lai: Các điểm bất thường (outliers) có thể làm lệch tâm cụm, ảnh hưởng đến kết quả phân cụm.
- Hạn chế với dữ liệu phức tạp: Không hiệu quả nếu cụm có hình dạng không đều (non-spherical) hoặc kích thước khác nhau.
- Phụ thuộc khởi tạo: Kết quả có thể thay đổi tùy thuộc vào tâm cụm ban đầu, đôi khi rơi vào cực tiểu cục bộ (local minima).

#### **Ứng dụng thực tế:**

- Phát hiện tin bất thường:
  - + Phân cụm các bài viết hoặc bài đăng trên mạng xã hội (như X posts) dựa trên đặc trưng văn bản (từ khóa, độ dài, cảm xúc) hoặc metadata (lượt chia sẻ, thời gian đăng).
  - + Các bài viết không nằm trong cụm lớn hoặc ở xa tâm cụm có thể được coi là tin bất thường, chẳng hạn tin giả hoặc nội dung gây tranh cãi.
  - + Ví dụ: Phân cụm tin tức thành các chủ đề lớn (chính trị, y tế); tin giả thường không khớp với cụm chính thống.
- Ngoài tin giả:
  - + Phân khúc khách hàng trong marketing (nhóm người dùng theo hành vi).
  - + Nén ảnh (giảm số lượng màu bằng cách phân cụm pixel).
  - + Phân tích dữ liệu sinh học (nhóm gen tương đồng).

#### **2.2.2.2 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**

*DBSCAN là một thuật toán phân cụm không giám sát dựa trên mật độ điểm dữ liệu. Nó nhóm các điểm gần nhau thành cụm dựa trên khái niệm "mật độ" và*

tự động xác định các điểm thừa thớt (noise) làm ngoại lai, không cần chỉ định số cụm trước.

### Cơ chế hoạt động:

- Tham số đầu vào:
- + Epsilon ( $\epsilon$ ): Khoảng cách tối đa để hai điểm được coi là "gần nhau".
- + MinPts: Số điểm tối thiểu trong bán kính  $\epsilon$  để một điểm được xem là "điểm lõi" (core point).
- Phân loại điểm:
  - + Điểm lõi: Có ít nhất MinPts điểm trong bán kính  $\epsilon$  (bao gồm chính nó).
  - + Điểm biên: Nằm trong bán kính  $\epsilon$  của điểm lõi nhưng không đủ MinPts để tự trở thành lõi.
  - + Điểm nhiễu: Không thuộc cụm nào, không gần điểm lõi nào (ngoại lai).
- Tạo cụm:
  - + Bắt đầu từ một điểm lõi bất kỳ, mở rộng cụm bằng cách thêm tất cả các điểm trong bán kính  $\epsilon$  của nó (bao gồm cả điểm lõi khác).
  - + Lặp lại cho đến khi không còn điểm lõi nào để mở rộng.
  - + Các điểm nhiễu được đánh dấu riêng.
  - + Kết quả là các cụm có hình dạng bất kỳ và một tập hợp điểm nhiễu (noise).

### Ưu điểm:

- Không cần chọn số cụm: Tự động xác định số cụm dựa trên mật độ, linh hoạt hơn K-means.
- Phát hiện ngoại lai tốt: Các điểm nhiễu được tách biệt, rất hữu ích để tìm tin bất thường.
- Xử lý hình dạng bất kỳ: Không yêu cầu cụm có dạng tròn, phù hợp với dữ liệu thực tế phức tạp.

### Nhược điểm:

- Nhạy cảm với tham số: Việc chọn  $\epsilon$  và MinPts rất quan trọng; sai tham số có thể dẫn đến cụm quá lớn hoặc quá nhỏ.
- Không hiệu quả với mật độ không đồng đều: Nếu dữ liệu có vùng mật độ cao xen lẫn vùng thưa, DBSCAN có thể bỏ sót cụm hoặc gộp sai.
- Tốn tài nguyên: Với dữ liệu lớn, tính toán khoảng cách giữa các điểm đòi hỏi bộ nhớ và thời gian lớn (độ phức tạp  $O(n^2)$  nếu không tối ưu).

### **Ứng dụng thực tế**

- Phát hiện tin bất thường:
- + Phân tích bài đăng trên mạng xã hội (như X) dựa trên đặc trưng như số lượt tương tác, từ khóa, hoặc nguồn đăng.
- + Các bài viết bị đánh dấu là "điểm nhiễu" (ít tương đồng với cụm chính) có thể là tin giả, tin đồn, hoặc nội dung bất thường.
- + Ví dụ: Phát hiện bài viết y tế sai lệch (như quảng cáo thuốc giả) không khớp với cụm tin tức y tế đáng tin cậy.
- Ngoài tin giả:
- + Phát hiện gian lận tài chính (giao dịch bất thường không thuộc cụm giao dịch thông thường).
- + Phân tích dữ liệu địa lý (nhóm vùng dân cư, xác định vùng ô nhiễm bất thường).
- + Phát hiện lỗi trong cảm biến IoT (dữ liệu bất thường từ thiết bị).

#### **2.2.2.3 So sánh K-means và DBSCAN**

| Tiêu chí          | K-means          | DBSCAN                        |
|-------------------|------------------|-------------------------------|
| Số cụm            | Cần chỉ định K   | Tự động xác định              |
| Xử lý nhiễu       | Kém              | Tốt                           |
| Hình dạng cụm     | Chỉ cụm hình cầu | Cụm bất kỳ                    |
| Độ phức tạp       | $O(n)$           | $O(n \mu^2)$                  |
| Phụ thuộc tham số | Phụ thuộc K      | Phụ thuộc $\epsilon$ , MinPts |

|                    |                  |                     |
|--------------------|------------------|---------------------|
| Ứng dụng điển hình | Phân nhóm chủ đề | Phát hiện ngoại lai |
|--------------------|------------------|---------------------|

Bảng 2.2 : So sánh K-means và DBSCAN

### Kết luận

- K-means phù hợp khi dữ liệu có cấu trúc rõ ràng và cần phân nhóm nhanh. Ví dụ: Phân loại tin tức thành nhóm chính trị, kinh tế, giải trí.
- DBSCAN vượt trội trong phát hiện tin giả nhờ khả năng tách biệt ngoại lai, đặc biệt khi tin giả có đặc trưng khác biệt (ví dụ: dùng từ ngữ kích động, lượt chia sẻ bất thường).

Để tăng hiệu quả, nên kết hợp cả hai phương pháp:

- Dùng K-means phân cụm sơ bộ.
- Áp dụng DBSCAN để tìm điểm nhiễu trong từng cụm.
- Kết hợp với mô hình học sâu (như BERT) để phân tích ngữ nghĩa sâu hơn

## 2.3. Các phương pháp phát hiện tin giả bằng học sâu

Học sâu (Deep Learning) đã cách mạng hóa lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) nhờ khả năng học biểu diễn ngữ nghĩa sâu từ dữ liệu. Trong phát hiện tin giả, các mô hình tiền huấn luyện (pre-trained models) như BERT, RoBERTa, và PhoBERT đã chứng minh hiệu quả vượt trội nhờ nắm bắt ngữ cảnh hai chiều và tối ưu hóa kiến trúc.

### 2.3.1 . Model BERT

#### 2.3.1.1. Giới thiệu của mô hình BERT

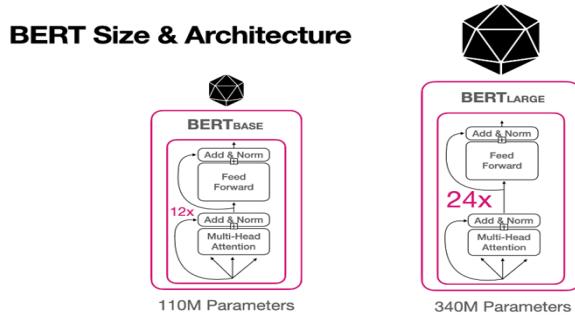
##### BERT (Bidirectional Encoder Representations from Transformers) [5]

Mô hình xử lý ngôn ngữ tự nhiên sử dụng Transformer, chỉ gồm phần Encoder.

Điểm đặc biệt của BERT là khả năng học ngữ cảnh hai chiều, hiểu nghĩa của từ dựa trên cả từ bên trái và phải trong câu.

BERT truyền toàn bộ từ trong câu vào mô hình cùng lúc, giúp huấn luyện không phụ thuộc vào thứ tự câu. Mô hình có nhiều phiên bản khác nhau, thay đổi dựa

trên số lượng layer, kích thước embedding vector, và số attention head trong multi-head attention.



Hình 2.1: Kiến trúc BERT based và BERT large

BERT có hai kiến trúc chính:

- BERT-base: 12 khối Transformer, embedding size 768, 12 attention heads, 110 triệu tham số.
- BERT-large: 24 khối Transformer, embedding size 1024, 16 attention heads, 340 triệu tham số.

BERT-large có số tầng gấp đôi, kích thước vector nhúng và số attention heads tăng 1.33 lần so với BERT-base.

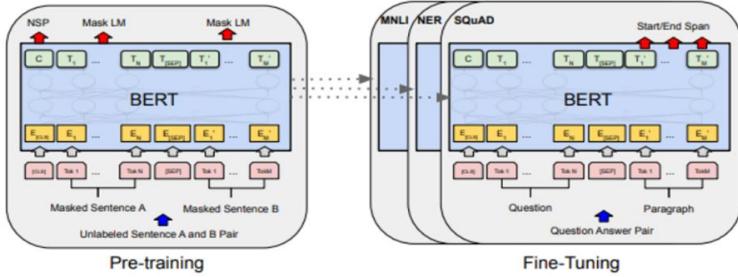
Mô hình được huấn luyện đồng thời hai nhiệm vụ:

- Masked Language Model (MLM) – Dự đoán từ bị che trong câu.
- Next Sentence Prediction (NSP) – Dự đoán câu tiếp theo có liên quan hay không.

Loss tổng là sự kết hợp của hai nhiệm vụ này và mô hình sẽ tối ưu để giảm loss chung.

### Fine tuning model BERT

BERT cho phép fine-tuning để tùy chỉnh mô hình pre-trained cho các tác vụ NLP khác nhau bằng cách cập nhật toàn bộ tham số.



Hình 2.2: Toàn bộ tiến trình pre-training và fine-tuning của BERT

#### Tiến trình Fine-Tuning

Bước 1: Tokenizer văn bản, thêm token [CLS] đầu câu và [SEP] giữa các câu (nếu là cặp câu).

Bước 2: Nhúng các token thành vector bằng pre-trained BERT.

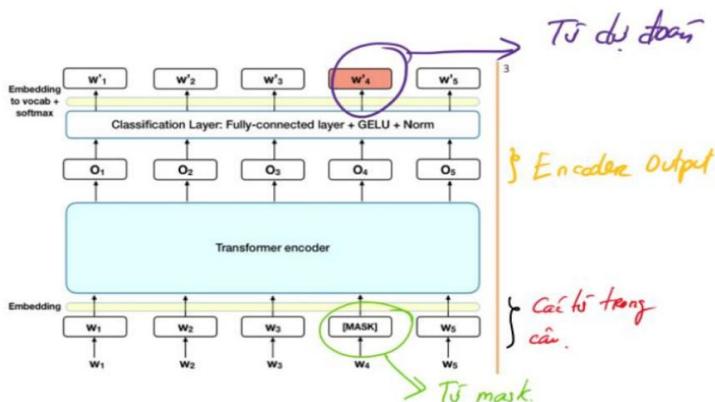
Bước 3: Truyền qua các lớp Transformer (multi-head attention, feed-forward).

Bước 4: Thêm một output layer phù hợp với tác vụ:

- Phân loại: Dùng fully connected layer và softmax.
- Hỏi - Đáp: Dự đoán vị trí bắt đầu/kết thúc câu trả lời.

Bước 5: Cập nhật tham số BERT bằng gradient descent để tối ưu hóa mô hình.

#### Mô hình ngôn ngữ ẩn (Masked Language Model)



Hình 2.3: Sơ đồ kiến trúc BERT cho nhiệm vụ MLM

- 15% từ trong câu bị thay bằng [MASK], mô hình dự đoán từ gốc dựa vào ngữ cảnh.
- Output của Transformer qua fully connected layer và softmax để tính xác suất từ gốc.
- Chỉ tính loss cho các vị trí bị che để cải thiện khả năng học ngữ cảnh.

#### **Next Sentence Prediction (NSP)**

- Xác định xem câu thứ hai có phải là câu tiếp theo của câu thứ nhất không.
- Ghép hai câu, thêm token [CLS] và [SEP].
- Thêm sentence embedding (0 cho câu đầu, 1 cho câu sau) và positional encoding.
- Lấy output tại [CLS], qua fully connected layer và softmax để phân loại IsNext / NotNext.

##### *2.3.1.2. Ưu điểm của mô hình BERT*

- Hiểu ngữ cảnh sâu : BERT sử dụng cơ chế self-attention và bidirectional encoding, giúp hiểu ý nghĩa của từ dựa trên toàn bộ câu.
- Chuyển giao tốt giữa các tác vụ NLP

Sau khi pretrain trên lượng dữ liệu lớn, BERT có thể fine-tune nhanh chóng cho nhiều tác vụ như phân loại văn bản, hỏi - đáp, dịch máy,...

Cải thiện độ chính xác vượt trội

BERT đạt kết quả tốt hơn nhiều so với các mô hình trước đó trong các bài toán NLP tiêu chuẩn như GLUE, SQuAD.

- Xử lý tốt các bài toán có cặp câu

Các tác vụ như hỏi - đáp (QA) hoặc dự đoán câu tiếp theo (NSP) đều được hỗ trợ tốt nhờ token [SEP].

##### *2.3.1.3. Nhược điểm của mô hình BERT*

- Tốn tài nguyên : BERT có hàng trăm triệu tham số, cần GPU/TPU mạnh để train & infer.
- Chậm hơn mô hình truyền thống

So với các mô hình đơn giản như LSTM hay TF-IDF, BERT có độ trễ cao hơn khi dự đoán.

- Không phù hợp với văn bản dài

BERT giới hạn tối đa 512 token, gây khó khăn khi xử lý văn bản dài.

- Cần nhiều dữ liệu để fine-tune tốt

Nếu tập dữ liệu nhỏ, mô hình có thể overfitting hoặc không đạt hiệu suất cao.

### **2.3.2. RoBERTa**

#### *2.3.2.1. Giới thiệu RoBERTa*

*RoBERTa (Robustly Optimized BERT approach) là phiên bản cải tiến của BERT do Facebook AI phát triển vào năm 2019, nhằm tối ưu hiệu suất xử lý ngôn ngữ tự nhiên. RoBERTa được huấn luyện trên tập dữ liệu lớn hơn và loại bỏ nhiệm vụ "Next Sentence Prediction" (NSP), chỉ tập trung vào "Masked Language Model" (MLM) để dự đoán từ bị che chính xác hơn.*

Mô hình áp dụng "dynamic masking" thay vì "static masking" như BERT, giúp học ngữ cảnh tốt hơn. Ngoài ra, RoBERTa sử dụng batch lớn hơn, tăng số bước huấn luyện, nhưng vẫn giữ nguyên kiến trúc Transformer hai chiều. Nhờ những cải tiến này, RoBERTa vượt trội hơn BERT trong nhiều tác vụ NLP như phân loại văn bản, phân tích cảm xúc và trả lời câu hỏi, trở thành mô hình phổ biến trong nghiên cứu và ứng dụng thực tế.

#### *2.3.2.2. Ưu điểm của RoBERTa*

- Hiệu suất cao hơn BERT: Nhờ huấn luyện trên dữ liệu lớn hơn và tối ưu hóa quá trình học, RoBERTa đạt kết quả vượt trội trong nhiều bài toán NLP như phân loại văn bản, phân tích cảm xúc, và trả lời câu hỏi.
- Không bị giới hạn bởi NSP: Việc loại bỏ nhiệm vụ "Next Sentence Prediction" giúp mô hình tập trung hoàn toàn vào việc hiểu ngữ nghĩa của từng câu.
- Dynamic Masking: Giúp mô hình học được ngữ cảnh phong phú hơn trong mỗi epoch thay vì lặp lại cùng một mẫu che từ như BERT.

- Tăng cường huấn luyện: Sử dụng batch lớn hơn và thời gian huấn luyện dài hơn giúp mô hình có khả năng biểu diễn ngữ nghĩa tốt hơn.

#### *2.3.2.3 Nhược điểm của RoBERTa:*

- Yêu cầu tài nguyên lớn: Việc huấn luyện RoBERTa đòi hỏi GPU mạnh và bộ nhớ cao do sử dụng batch lớn hơn và nhiều bước huấn luyện hơn.
- Không tối ưu cho mọi tác vụ: Dù loại bỏ NSP giúp cải thiện hiệu suất trên nhiều bài toán, một số ứng dụng như liên kết câu hoặc tóm tắt văn bản có thể hưởng lợi từ NSP.
- Ít thân thiện khi triển khai: Kích thước mô hình lớn làm tăng chi phí lưu trữ và tính toán khi đưa vào ứng dụng thực tế.

#### **2.3.3. PhoBERT**

##### *2.3.3.1. Giới thiệu PhoBERT*

*PhoBERT [5] là mô hình NLP chuyên biệt cho tiếng Việt, do VinAI Research phát triển dựa trên RoBERTa. Mô hình được tối ưu hóa bằng cách loại bỏ nhiệm vụ "Next Sentence Prediction" và sử dụng "dynamic masking" để nâng cao khả năng hiểu ngữ cảnh.*

Mô hình có hai phiên bản: PhoBERTbase (12 transformer blocks) và PhoBERTlarge (24 transformer blocks). Do tiếng Việt có cấu trúc ngữ pháp phức tạp và nhiều dấu thanh, PhoBERT giúp xử lý tốt hơn so với BERT gốc. PhoBERT sử dụng Masked Language Modeling (MLM), trong đó một số từ trong câu bị che đi, buộc mô hình phải dự đoán từ dựa trên ngữ cảnh. Token [CLS] chứa thông tin ngữ nghĩa của toàn câu, hữu ích cho các bài toán như phân loại văn bản, tóm tắt văn bản và nhận diện tin giả.

##### **Tại sao cần PhoBERT?**

Tiếng Việt có cấu trúc ngữ pháp phức tạp, nhiều dấu thanh và khác biệt lớn so với tiếng Anh. BERT gốc, được huấn luyện trên dữ liệu tiếng Anh, khó nắm bắt chính xác ngữ nghĩa và cú pháp tiếng Việt. PhoBERT, được huấn luyện

hoàn toàn trên dữ liệu tiếng Việt, giúp tối ưu hóa khả năng hiểu và phân tích văn bản tiếng Việt.

### **Quá trình huấn luyện PhoBERT**

PhoBERT được huấn luyện trên dữ liệu lớn từ báo chí và mạng xã hội, sử dụng Masked Language Modeling (MLM) để đoán từ bị che, buộc mô hình hiểu ngữ cảnh tổng thể. PhoBERT không sử dụng Next Sentence Prediction (NSP), theo kiến trúc RoBERTa.

### **Token [CLS] và embedding trong PhoBERT**

PhoBERT sử dụng token [CLS] để biểu diễn toàn bộ câu. Vector embedding của [CLS] chứa thông tin ngữ nghĩa tổng thể, hỗ trợ hiệu quả cho các bài toán phân loại văn bản, như nhận diện tin giả.

#### *2.3.3.2. Ưu điểm của mô hình PhoBERT*

- Tối ưu cho tiếng Việt: Được huấn luyện hoàn toàn trên dữ liệu tiếng Việt, PhoBERT xử lý ngữ nghĩa và cú pháp tiếng Việt tốt hơn so với BERT gốc.
- Hiệu suất cao trên nhiều bài toán NLP : PhoBERT đạt kết quả vượt trội trong các nhiệm vụ như phân loại văn bản, trích xuất thông tin, tóm tắt, sinh văn bản, v.v.
- Dữ liệu huấn luyện lớn: Được train trên 20GB dữ liệu, bao gồm Wikipedia và báo chí, giúp mô hình hiểu tốt ngữ cảnh tiếng Việt.
- Tận dụng cải tiến của RoBERTa : Loại bỏ nhiệm vụ Next Sentence Prediction (NSP) và sử dụng dynamic masking, giúp mô hình mạnh hơn trong việc học ngữ cảnh.
- Hỗ trợ hai phiên bản: PhoBERT có cả phiên bản base (12 layers) và large (24 layers), phù hợp với nhu cầu tính toán khác nhau.

#### *2.3.3.3. Nhược điểm của mô hình phoBERT*

- Tốn tài nguyên tính toán: Mô hình lớn, đòi hỏi phần cứng mạnh (GPU/VRAM cao) để huấn luyện và suy luận nhanh.

- Không hỗ trợ đa ngôn ngữ: Chỉ tối ưu cho tiếng Việt, không thể sử dụng hiệu quả với các văn bản song ngữ hoặc đa ngữ.
- Thiếu dữ liệu hội thoại và văn nói: Chủ yếu được huấn luyện trên văn bản báo chí, PhoBERT có thể gặp khó khăn với ngữ cảnh hội thoại hoặc tiếng Việt không chuẩn mực.
- Không có sẵn mô hình tiếp tục huấn luyện (fine-tuning) chuyên biệt: Cần tinh chỉnh cho từng bài toán cụ thể, vì mô hình gốc chưa được tối ưu hóa sẵn cho các ứng dụng đặc thù như chatbot, phân loại tin giả, v.v.

#### 2.4. So sánh PhoBERT, BERT và RoBERTa

| TIÊU CHÍ                       | BERT  | RoBERTa   | PhoBERT  |
|--------------------------------|---|---|--|
| <b>Khả năng xử lý ngữ cảnh</b> | Dùng Attention hai chiều để hiểu ngữ cảnh trong câu và đoạn văn, nhưng mô hình gốc huấn luyện bằng tiếng Anh nên không tối ưu cho tiếng Việt. | Cải tiến từ BERT với Dynamic Masking và loại bỏ Next Sentence Prediction (NSP), giúp học ngữ cảnh hiệu quả hơn. Tuy nhiên, vẫn chủ yếu huấn luyện bằng tiếng Anh. | Dựa trên RoBERTa nhưng được huấn luyện chuyên biệt trên dữ liệu tiếng Việt, giúp hiểu rõ cú pháp, từ ghép, dấu câu của tiếng Việt. |
| <b>Hiệu suất trên NLP</b>      | Mạnh với các bài toán NLP nhưng hiệu suất bằng tiếng Việt bị hạn chế nếu không fine-  | Vượt trội hơn BERT trong xử lý văn bản dài, hiệu quả cao hơn trên các bài toán NLP tổng quát. Tuy nhiên, vẫn  | Hiệu suất cao hơn đáng kể trong các bài toán NLP tiếng Việt như phân loại văn bản,   |

|                                     |   |  |  |
|-------------------------------------|---|--|--|
|                                     | tune với dữ liệu tiếng Việt.  | không tối ưu cho tiếng Việt.   | phân tích cảm xúc, phát hiện tin giả, v.v.   |
| <b>Tài nguyên &amp; Độ phức tạp</b> | Yêu cầu tài nguyên lớn để huấn luyện và triển khai.                   | Cần tài nguyên tính toán cao hơn BERT do cải tiến trong cách huấn luyện. | Tương tự RoBERTa về tài nguyên, nhưng do được tiền huấn luyện cho tiếng Việt, quá trình fine-tuning nhanh hơn và hiệu quả hơn. |
| <b>Ứng dụng thực tế</b>             | Phù hợp cho NLP tiếng Anh, cần tinh chỉnh khi áp dụng cho tiếng Việt. | Tốt hơn BERT cho NLP tổng quát, nhưng vẫn chưa tối ưu cho tiếng Việt.    | Lựa chọn tối ưu cho NLP tiếng Việt, đặc biệt khi cần xử lý ngữ cảnh sâu và dữ liệu phức tạp.                                   |

Bảng 2.3: Bảng so sánh các mô hình học sâu

Kết luận: Nếu làm NLP tiếng Anh, RoBERTa tốt hơn BERT nhờ tối ưu huấn luyện. Nhưng với tiếng Việt, PhoBERT là lựa chọn tốt nhất, vì nó được huấn luyện chuyên biệt trên dữ liệu tiếng Việt, giúp hiểu rõ ngữ pháp và từ vựng tiếng Việt hơn so với BERT hoặc RoBERTa .

## 2.5. Kết luận chương hai

Chương 2 trình bày các phương pháp học máy và NLP trong phát hiện tin giả, tập trung vào mô hình PhoBERT – tối ưu cho tiếng Việt. So với các mô

hình trước đây như BERT và RoBERTa , PhoBERT xử lý ngữ nghĩa tốt hơn, giúp nâng cao độ chính xác trong phân loại văn bản. Các kỹ thuật như Word Embedding, Transformer và Self-Attention đóng vai trò quan trọng trong trích xuất đặc trưng văn bản. Tuy nhiên, vẫn còn thách thức như mất cân bằng dữ liệu và yêu cầu tài nguyên lớn. Chương này cung cấp nền tảng lý thuyết quan trọng để phát triển hệ thống phát hiện tin giả hiệu quả

## CHƯƠNG 3. TRIỂN KHAI HỆ THỐNG

### 3.1. Thu thập thông tin và thiết kế hệ thống

#### 3.1.1. Triển khai khảo sát, thu thập thông tin

##### 3.1.1.1. Mục tiêu khảo sát

Mục tiêu chính của khảo sát là:

- Thu thập dữ liệu đa nguồn: Tập hợp thông tin về tin giả từ các nền tảng mạng xã hội (Facebook, Twitter, TikTok), trang tin tức, và diễn đàn trực tuyến.
- Nhận diện đặc trưng tin giả: Phân tích nội dung, cấu trúc, và metadata (nguồn đăng, thời gian, lượt tương tác) để xác định các dấu hiệu đặc trưng của tin giả.
- Đánh giá tác động: Đo lường mức độ lan truyền, ảnh hưởng đến nhận thức cộng đồng, và thiệt hại kinh tế - xã hội do tin giả gây ra.
- Đề xuất giải pháp: Xây dựng biện pháp nâng cao nhận thức người dùng và cải thiện hệ thống tự động phát hiện tin giả.

##### 3.1.1.2. Phương pháp thu thập thông tin

Quá trình thu thập thông tin được triển khai thông qua ba phương pháp chính, đảm bảo tính đa chiều và độ tin cậy của dữ liệu, bao gồm: thu thập dữ liệu tự động, khảo sát người dùng và phân tích đặc trưng bằng học máy. Cụ thể như sau:

- Thu thập dữ liệu tự động

Dữ liệu được thu thập từ các nguồn trực tuyến như mạng xã hội, trang tin tức và diễn đàn thông qua công cụ crawl chuyên dụng. Công cụ này được thiết kế để quét và trích xuất nội dung đa dạng, bao gồm bài viết, bình luận và hình ảnh, dựa trên các từ khóa liên quan đến chủ đề nghiên cứu. Quá trình này tuân thủ các nguyên tắc đạo đức và pháp lý, đảm bảo chỉ thu thập thông tin công khai, không vi phạm quyền riêng tư hoặc điều

khoản sử dụng của nền tảng. Dữ liệu sau đó được làm sạch và chuẩn hóa để loại bỏ nội dung trùng lặp hoặc không phù hợp trước khi phân tích.

- Khảo sát người dùng

Phương pháp định tính được áp dụng thông qua phỏng vấn và bảng hỏi nhằm đánh giá nhận thức, mức độ tin tưởng và khả năng nhận diện tin giả của người dùng. Đối tượng tham gia được lựa chọn đa dạng về độ tuổi, nghề nghiệp và mức độ tiếp cận thông tin để đảm bảo tính đại diện. Câu hỏi khảo sát được thiết kế kết hợp giữa câu hỏi đóng (lựa chọn) và câu hỏi mở, cho phép thu thập cả dữ liệu định lượng và định tính. Kết quả được mã hóa và phân tích theo chủ đề để xác định xu hướng chung và điểm bất thường trong nhận thức của cộng đồng.

- Phân tích đặc trưng bằng học máy

Dữ liệu thu thập được xử lý qua các bước trích xuất đặc trưng (feature extraction) để xác định các dấu hiệu đặc trưng của tin giả, như ngôn ngữ cảm tính, nguồn thông tin không xác thực, hoặc mâu thuẫn truyền bá thường. Mô hình học máy được huấn luyện trên tập dữ liệu đã gán nhãn (tin thật/tin giả), kết hợp kỹ thuật xác thực chéo để đảm bảo độ chính xác. Quá trình này bao gồm việc tối ưu hóa siêu tham số và đánh giá hiệu suất mô hình thông qua các chỉ số phân loại. Kết quả phân tích giúp xây dựng bộ quy tắc nhận diện tin giả, đồng thời cung cấp cơ sở để cải tiến mô hình theo thời gian.

#### Tính liên kết giữa các phương pháp

Dữ liệu từ crawl tự động cung cấp đầu vào cho mô hình học máy, trong đó kết quả khảo sát người dùng đóng vai trò hỗ trợ trong việc xác định đặc trưng cần phân tích. Sự kết hợp này cho phép nghiên cứu tiếp cận vấn đề từ cả góc độ kỹ thuật và xã hội, nâng cao tính toàn diện và khả năng ứng dụng thực tiễn của kết quả.

### 3.1.1.3. Phân loại tin giả

Theo “Cẩm nang phòng chống tin giả, tin sai sự thật trên không gian mạng” [4], tin giả được chia thành bốn nhóm chính: Tin giả về thương mại, Tin giả về chính trị, Tin giả về đời sống xã hội và Tin châm biếm, hài hước. Trong phạm vi khảo sát này, chúng tôi tập trung vào hai nhóm phô biến và có tác động lớn nhất: Tin giả về đời sống xã hội và Tin giả về chính trị. Việc lựa chọn hai nhóm này dựa trên mức độ ảnh hưởng rộng rãi, khả năng lan truyền cao và mức độ gây nhiễu loạn thông tin trên mạng xã hội.

Bộ dữ liệu khảo sát bao gồm 21.000 tin, trong đó:

- 10.000 tin thường ( 4.000 tin thường về đời sống xã hội và 6000 tin thường về chính trị )
- 4.500 tin giả về đời sống xã hội
- 6.500 tin giả về chính trị

#### ❖ Tin giả về đời sống xã hội

*Tin giả loại này thường tạo ra thông tin sai lệch liên quan đến đời sống, sức khỏe, tài chính và các vấn đề xã hội nhằm gây hoang mang hoặc thao túng niềm tin của công chúng.* Các đặc trưng chính gồm:

- + Phong tục : Nội dung dựa vào niềm tin dân gian, tín ngưỡng để bịa đặt thông tin, dễ gây hiểu lầm về truyền thống hoặc niềm tin tôn giáo. (Boolean/Số thực - Xác định từ khóa liên quan đến phong tục, tín ngưỡng).
- Cách thức hoạt động
- Bịa đặt gắn với yếu tố tâm linh:
  - Sử dụng các câu chuyện "có vẻ" hợp lý dựa trên truyền thuyết, lời đồn hoặc nghi lễ dân gian.
  - Ví dụ:
    - "Đêm mùng 1 Âm lịch, ai thắp hương sai cách sẽ bị ma ám cả năm."
    - "Ăn cơm chan canh theo cách này sẽ bị tổ tiên nguyên rùa."

**Commented [1]:** bổ sung thêm các yếu tố đặc trưng cho từng nhóm tin

- Lợi dụng sự thiếu hiểu biết:
  - Nhắm vào nhóm người lớn tuổi, vùng nông thôn ít tiếp cận thông tin khoa học.
  - Ví dụ: Tin đồn về "bùa ngải", "bói toán vận mệnh" kèm lời cảnh báo giật gân.
- Gắn với sự kiện tôn giáo:
  - Xuyên tạc nghi lễ, ngày lễ lớn (Phật đản, Giáng sinh) để gây hoang mang.
  - Ví dụ:  
 "Chùa Ba Vàng phát miễn phí 10 tờ tiền âm phủ, ai nhận sẽ gặp han."
  - Đặc điểm ngôn ngữ/nội dung
- Từ khóa thường gặp (Boolean/Số thực để nhận diện):
  - Tâm linh: "linh ứng", "báo mộng", "vong nhập", "xui xẻo", "hóa giải", "bùa chú".
  - Phong tục: "kiêng kỵ", "ngày xấu", "giờ hoàng đạo", "cúng tế", "tam tai", "tháng cõi hồn".
  - Mê tín: "điếc báo", "số phận", "thầy pháp", "nghiệp quả".
- Cấu trúc câu:
  - Khẳng định chắc chắn: "100% người làm điều này sẽ gặp họa".
  - Đe dọa gián tiếp: "Nhà nào không làm theo sẽ mất lộc".
  - Ví dụ thực tế
- Ví dụ 1:
  - Tin giả: "Phụ nữ mang thai không được cắt tóc sau 6h tối, kéo con sinh ra bị dị tật."
  - Sự thật: Không có cơ sở y khoa hay tín ngưỡng nào chứng minh.
- Ví dụ 2:
  - Tin giả: "Đặt 3 quả trứng gà ở góc nhà ngày Rằm sẽ hút tài lộc."
  - Mục đích: Câu view, dẫn đến trang bán hàng "trứng phong thủy".
  - Cách phát hiện

- Kiểm tra nguồn gốc:
  - Thông tin có được dẫn từ sách cổ, chuyên gia văn hóa dân gian uy tín?
  - Đối chiếu với trang của Viện Nghiên cứu Văn hóa Việt Nam hoặc Giáo hội Phật giáo.
- Phân tích ngữ cảnh:
  - Nếu nội dung đe dọa hoặc yêu cầu "chia sẻ để tránh hạn" thì khả năng cao là tin giả.
  - + Giật gân : Sử dụng ngôn từ gây sốc để thu hút sự chú ý, như "kinh hoàng", "chấn động", "không thể tin được". (Boolean/Số thực - Đếm số lần xuất hiện từ khóa giật gân)
  - Cách thức hoạt động
- Khuếch đại sự kiện thông thường: Biến một vụ việc nhỏ thành "thảm họa" hoặc "bí mật động trời".
- Ví dụ:
 

*"Kinh hoàng: Người phụ nữ tử vong sau 1 ngày ăn mì tôm"* (thực tế do ngộ độc thực phẩm, không liên quan mì tôm).

*"Chấn động: Cả làng mắc ung thư vì ăn rau nhiễm độc"* (không có bằng chứng khoa học).
- Tạo cảm giác khẩn cấp: Dùng từ ngữ ám chỉ "nguy cơ tức thì" để ép người đọc hành động (chia sẻ, bấm vào link).
- Ví dụ:
 

*"Cảnh báo: 90% người Việt đang dùng loại nước này mà không biết nó gây ung thư!"*
- Lợi dụng hình ảnh/video shock: Kèm theo ảnh máu me, tai nạn ghê rợn dù không liên quan đến nội dung.
- Đặc điểm ngôn ngữ/nội dung
- Từ khóa giật gân (Boolean/Số thực để đếm tần suất):

- Mức độ kinh hoàng: "*rùng rợn*", "*kinh dị*", "*sưng sốt*", "*hãi hùng*", "*té ngửa*".
- Tính chất khẩn cấp: "*cảnh báo*", "*nguy cơ*", "*khẩn cấp*", "*ngay lập tức*".
- Khẳng định tuyệt đối: "*100%*", "*tất cả*", "*không ai ngờ*", "*bí mật bị giấu kín*".
- Cấu trúc tiêu đề điển hình:
  - "[TÙ GIÁT GÂN] + [SỰ KIỆN] + [HẬU QUẢ KHỦNG KHIẾP]"  
"CHÂN ĐỘNG: Bé trai 3 tuổi chết sau khi tiêm vaccine COVID-19" (thực tế do bệnh lý khác).
  - Ví dụ thực tế
- Ví dụ 1:
  - Tin giả: "*KHÔNG THỂ TIN NỐI: Ăn 1 quả trứng gà luộc buổi sáng tương đương hút 5 điếu thuốc!*"
  - Sự thật: Không có nghiên cứu nào chứng minh, thông tin bịa đặt từ trang web "súrc khỏe bí ẩn".
- Ví dụ 2:
  - Tin giả: "*HÀNG NGÀN NGƯỜI SẼ CHẾT NẾU KHÔNG BIẾT điều này về điện thoại Samsung!*"
  - Mục đích: Dẫn dụ click vào link quảng cáo phần mềm độc hại.
  - Cách phát hiện
- Kiểm tra nguồn tin:
  - Tin từ báo chí chính thống (VNNExpress, Tuổi Trẻ) thường ít dùng từ ngữ giật gân cực đoan.
  - Nếu tiêu đề có dấu chấm than (!) hoặc viết hoa toàn bộ thì nên cảnh giác.
- Phân tích nội dung:
  - So sánh với thông tin từ Bộ Y tế, WHO, hoặc trang fact-check (Tinh Tế, Vietnam Fact Check).

- Kiểm tra xem có dẫn nguồn nghiên cứu cụ thể không (ví dụ: "theo Đại học Harvard" nhưng không có link).
- + Lừa đảo : Tin giả về đầu tư, trúng thưởng hoặc kinh doanh đa cấp nhằm lừa đảo người đọc, thường xuất hiện trong quảng cáo giả mạo. (Boolean/Số thực
  - Phát hiện cụm từ liên quan đến tài chính đáng ngờ)
  - Cách thức hoạt động
- Mồi nhử lợi nhuận cao:
  - Hứa hẹn lãi suất "khủng", lợi nhuận "siêu tốc" với rủi ro bằng 0.
  - Ví dụ:  
 "ĐẦU TƯ 10 TRIỆU - NHẬN NGAY 50 TRIỆU SAU 1 THÁNG! Cam kết hoàn tiền 100%!"  
*"Trúng thưởng iPhone 15 Pro Max khi chia sẻ link sau 3 phút!"*
- Giả danh tổ chức uy tín:
  - Mạo danh ngân hàng, tập đoàn lớn (Vietcombank, VinGroup) hoặc người nổi tiếng để tăng độ tin cậy.
  - Ví dụ:  
 "Chương trình ưu đãi từ Ngân hàng ABC: Nhận 5 triệu VND khi mở thẻ online!" (link giả mạo).
- Lừa đảo đa cấp biến tướng:
  - Dụ dỗ tham gia mô hình "kinh doanh online", "bán hàng tích điểm" với yêu cầu đóng tiền ban đầu.
  - Đặc điểm ngôn ngữ/nội dung
- Từ khóa lừa đảo (Boolean/Số thực để phát hiện):
  - Lợi nhuận phi lý: "*siêu lãi*", "*lãi 300%*", "*kiếm tiền thụ động*", "*không cần vốn*".
  - Trúng thưởng giả: "*bạn đã trúng*", "*nhận quà miễn phí*", "*phản thưởng chờ đợi*".
  - Áp lực thời gian: "*ưu đãi chỉ trong 24h*", "*nhanh tay kéo hết*".

- Giả mạo tổ chức: "hợp tác với Ngân hàng Nhà nước", "được chứng nhận bởi Bộ Tài chính".
- Cấu trúc thông điệp điển hình:
  - "[LỢI NHUẬN/KHỎI SỞ] + [CAM KẾT KHÔNG TƯỞNG] + [HÀNH ĐỘNG GẤP]"  
 "BÍ QUYẾT KIỂM I TỶ TRONG 3 THÁNG: CLICK NGAY ĐỂ NHẬN TÀI LIỆU MIỄN PHÍ!"
  - Ví dụ thực tế
- Ví dụ 1 (Đầu tư ảo):
  - Tin giả: "CƠ HỘI ĐẦU TƯ TIỀN ẢO X3 LÃI SAU 7 NGÀY! Đăng ký ngay để được tư vấn!"
  - Sự thật: Trang web lừa đảo, sau khi nạp tiền sẽ biến mất.
- Ví dụ 2 (Trúng thưởng giả):
  - Tin giả: "Chúc mừng số điện thoại 09xx đã trúng thưởng 1 tỷ từ VinaPhone. Bấm vào đây để nhận tiền!"
  - Mục đích: Lừa đảo thông tin cá nhân hoặc cài mã độc.
  - Cách phát hiện và phòng tránh
- Kiểm tra nguồn gốc:
  - Website có địa chỉ lạ (không phải domain chính thức của ngân hàng/tập đoàn)?
  - Tra cứu tên công ty trên Công thông tin quốc gia
- Cảnh giác với yêu cầu tiền mặt:
  - Bất kỳ khoản "phí đặt cọc", "thuê trúng thưởng" nào đều là dấu hiệu lừa đảo.
- + Thông tin chưa kiểm chứng: Tin tức không có nguồn rõ ràng, chưa được kiểm chứng từ báo chí chính thống hoặc tổ chức uy tín. So sánh các tin chính thống
- Cách thức hoạt động

- Lan truyền qua MXH trước khi kiểm chứng:
  - Thông tin được đăng tải bởi cá nhân/tài khoản ẩn danh, không trích dẫn nguồn cụ thể.
  - Ví dụ:  
"Cô gái Hà Nội nhiễm HIV do tiêm vaccine COVID-19" (chỉ là tin đồn từ một bài đăng Facebook không tên tuổi).
- Dựa vào tin đồn, nghe nói:
  - Sử dụng cụm từ như "*người ta nói rằng...*", "*theo một số nguồn tin...*" để né tránh trách nhiệm.
- Lợi dụng sự kiện "nóng":
  - Bám vào các sự kiện gây tranh cãi (thiên tai, dịch bệnh) để phát tán thông tin sai lệch.
  - Ví dụ:  
"*Bão số 9 sẽ đổ bộ vào TP.HCM vào đêm nay*" (trong khi cơ quan khí tượng chưa công bố).
- Đặc điểm ngôn ngữ/nội dung
- Từ khóa đáng ngờ (Boolean/Số thực để nhận diện):
  - Nguồn tin mơ hồ: "*theo nguồn tin riêng*", "*một số người cho biết*", "*cộng đồng mạng đang xôn xao*".
  - Ngôn ngữ né tránh: "*được cho là*", "*có thể*", "*đang nghi vấn*".
  - Thiếu chi tiết cụ thể: Không ghi rõ địa điểm, thời gian, nhân vật liên quan.
- Cấu trúc phổ biến:
  - "[*THÔNG TIN GÂY SỐC*] + [*NGUỒN KHÔNG RÕ RÀNG*] + [*KÊU GỌI CHIA SẺ*]"  
"*Bé 5 tuổi bị bắt cóc tại công viên Thủ Lệ, nghi phạm là người nước ngoài (theo người dân địa phương)*".
- Ví dụ thực tế
- Ví dụ 1(Dịch bệnh):

- Tin chưa kiểm chứng: "Hà Nội xuất hiện ổ dịch COVID-19 mới tại chợ X, hàng chục ca nhiễm".
- Sự thật: Bộ Y tế khẳng định không có ổ dịch nào tại địa điểm đó.
- Ví dụ 2 (Tai nạn):
  - Tin đồn: "*Xe container tông loạt xe máy trên cao tốc, 10 người chết* (nguồn: Facebook user A)".
  - Kiểm chứng: Cơ quan chức năng xác nhận chỉ có 2 người bị thương nhẹ.
  - Cách phát hiện và kiểm chứng
- Đồi chiêu với báo chí chính thống:
  - Kiểm tra xem VnExpress, Tuổi Trẻ, VietnamNet có đưa tin tương tự không.
  - Sử dụng công cụ Google News (gõ sự kiện + "site:vnexpress.net").
- Kiểm tra nguồn gốc hình ảnh/video:
  - Dùng Google Reverse Image Search hoặc InVID để xem ảnh có bị lấy từ sự kiện cũ không.
- Cảnh giác với tài khoản đăng tin:
  - Tài khoản Facebook không có dấu tích xanh, website không có giấy phép MXH

#### ❖ Tin giả về chính trị

*Tin giả chính trị thường bị bóp méo để thao túng dư luận, gây mất niềm tin vào chính phủ, tổ chức hoặc cá nhân có ảnh hưởng.* Các đặc trưng nhận diện gồm:

- + Gây hoang mang: Nội dung kích động lo ngại về tình hình chính trị, an ninh hoặc kinh tế, thường dùng các cụm từ như "khủng hoảng", "âm mưu", "sụp đổ". (Boolean/Số thực - Kiểm tra từ khóa tiêu cực)
- Cách thức hoạt động
- Khuếch đại mối đe dọa: Biến sự kiện thông thường thành "khủng hoảng nghiêm trọng" hoặc "thảm họa".

- Ví dụ:
    - "*KINH HOÀNG*: Chính phủ sắp phá giá đồng tiền 50% vào tuần tới!"
    - "*ÂM MUỐU*: Tổ chức X đang kiểm soát toàn bộ hệ thống bầu cử!"
  - Tạo cảm giác cấp bách: Ép người đọc hành động ngay (chia sẻ, biểu tình) bằng cách dùng từ ngữ ám chỉ "nguy cơ tức thì".
  - Ví dụ:
    - "*CẢNH BÁO*: Nếu không phản đối ngay, luật này sẽ tước quyền tự do của bạn!"
  - Lợi dụng sự kiện nhạy cảm: Bám vào các vấn đề dễ gây tranh cãi (bầu cử, biểu tình, khủng hoảng kinh tế).
  - Đặc điểm ngôn ngữ/nội dung
- Từ khóa nhạy cảm (Boolean/Số thực)
- Mức độ nghiêm trọng:
  - "khủng hoảng", "sụp đổ", "thảm họa", "diệt vong", "tận thế".
  - Tính chất âm mưu:
  - "âm mưu", "thế lực ngầm", "che giấu", "bí mật động trời".
  - Ngôn ngữ kích động:
  - "phải hành động ngay", "trước khi quá muộn", "nguy cơ không thể đảo ngược".

Cấu trúc tiêu đề điển hình:

- "[TÙ GIẬT GÂN] + [SỰ KIỆN] + [HẬU QUẢ KHỦNG KHIẾP]"
- "CHẨN ĐỘNG: Nền kinh tế Việt Nam sắp sụp đổ vì chính sách mới!"
- Ví dụ thực tế
- Ví dụ 1 (Kinh tế):
  - Tin giả: "*ĐỒNG TIỀN MẤT GIÁ* 80%: Ngân hàng Nhà nước phá sản?"
  - Sự thật: Tỷ giá ổn định, thông tin bị đặt từ trang web giả mạo.
- Ví dụ 2 (An ninh):

- Tin giả: "BẮT CÓC TRẺ EM HÀNG LOẠT: Âm mưu của tổ chức buôn người quốc tế!"

- Kiểm chứng: Công an khẳng định không có vụ bắt cóc nào.

- Cách phát hiện

#### Kiểm tra nguồn tin

- Bước 1: Tìm kiếm thông tin trên báo chính thống (VNEtExpress, Tuổi Trẻ, VietnamPlus).
- Bước 2: Kiểm tra tuyên bố từ cơ quan liên quan (Chính phủ, Ngân hàng Nhà nước, Bộ Công an).

#### Phân tích ngữ cảnh

- Cảnh giác nếu:
  - Không có trích dẫn cụ thể (ví dụ: "theo nguồn tin nội bộ").
  - Hình ảnh/video không rõ nguồn gốc hoặc bị chỉnh sửa.
- + Bôi nhọ cá nhân/tổ chức: Tấn công danh dự của một cá nhân hoặc tổ chức bằng thông tin sai lệch hoặc suy diễn. (Boolean/Số thực - Xác định ngôn ngữ công kích, từ ngữ mang tính xúc phạm)
- Cách thức hoạt động
  - Tấn công trực tiếp vào uy tín của cá nhân/tổ chức bằng thông tin sai sự thật
  - Sử dụng chiến thuật:
    - *Đổ lỗi*: Gán ghép tội lỗi/hành vi xấu không có bằng chứng
    - *Nhân dán*: Dán mác tiêu cực ("tham nhũng", "phản bội")
    - *Suy diễn ác ý*: Từ sự kiện nhỏ suy ra kết luận cực đoan
  - Đặc điểm ngôn ngữ/nội dung
- Từ khóa nhận diện (Boolean)
  - Công kích cá nhân:
    - "tham nhũng", "hối lộ", "biến thủ", "lừa đảo"

- "vô trách nhiệm", "bất tài", "thối nát"
- Xúc phạm tổ chức:
- "bù nhìn", "tay sai", "bán nước"
- "mafia", "lợi ích nhóm", "độc tài"

Cấu trúc điển hình

- "[CÁ NHÂN/TỔ CHỨC] + [HÀNH VI XẤU] + [KẾT LUẬN CỰC ĐOAN]"

"Bộ trưởng X THAM NHŨNG 100 tỷ - Bằng chứng cho thấy cả chính phủ đều thối nát!"

- Ví dụ thực tế

Ví dụ 1 (Cá nhân):

- Tin giả: "DÂN CHỦNG: Lãnh đạo tỉnh Y sở hữu 10 biệt thự ở nước ngoài"
- Thực tế: Ảnh biệt thự được lấy từ Google Images, không liên quan

Ví dụ 2 (Tổ chức):

- Tin giả: "Tập đoàn Z RỦA TIỀN qua dự án từ thiện"
- Kiểm chứng: Không có đơn tố cáo nào từ cơ quan điều tra
- Cách phát hiện

Kiểm tra bằng chứng:

- Yêu cầu xem tài liệu/xác minh của cơ quan có thẩm quyền
- Tra cứu hồ sơ tài sản công khai (nếu có)

Phân tích ngôn ngữ:

- Cảnh giác với các từ:  
"được cho là", "nghi ngờ rằng", "có khả năng"  
"ai cũng biết", "sự thật không thể chối cãi"
- + Thuyết âm mưu: Truyền bá thông tin không có căn cứ, cho rằng có một thế lực ngầm đang thao túng sự kiện. (Boolean/Số thực - Phát hiện các cụm từ như "sự thật bị che giấu", "thế lực ngầm")

- Cách thức hoạt động
  - Xây dựng kịch bản phức tạp: Dựng lên các câu chuyện có tính liên kết giả tạo giữa các sự kiện rờ rạc
  - Lợi dụng tâm lý nghi ngờ: Khai thác xu hướng không tin tưởng vào thông tin chính thống của một bộ phận công chúng
  - Tạo cảm giác "biết sự thật": Nuôi dưỡng niềm tin rằng chỉ có họ mới dám nói lên sự thật
- Đặc điểm ngôn ngữ/nội dung

Từ khóa nhận diện (Boolean/Số thực)

- Ngôn ngữ bí mật:
  - "sự thật bị che giấu", "họ không muốn bạn biết"
  - "tài liệu mật", "bí mật động trời"
- Chỉ định thế lực:
  - "nhóm lợi ích", "tầng lớp tinh hoa"
  - "chính phủ ngầm", "hội kín"
- Tính chất phi lý:
  - "kế hoạch 100 năm", "âm mưu toàn cầu"
  - "không phải ngẫu nhiên", "tất cả đều có liên hệ"

Cấu trúc điển hình

- "[SỰ KIỆN THỰC] + [GIẢI THÍCH ÂM MƯU] + [KÊU GỌI HÀNH ĐỘNG]"

"Dịch COVID-19 thực ra là vũ khí sinh học - Đây là kế hoạch của các tập đoàn được để bán vaccine - Hãy chia sẻ để mọi người biết sự thật!"

- Ví dụ thực tế

Ví dụ 1 (Y té):

- Tin giả: "Vaccine COVID chứa chip theo dõi - Bill Gates thừa nhận trong cuộc phỏng vấn bị rò rỉ"
- Thực tế: Video bị cắt ghép từ nhiều nguồn khác nhau

Ví dụ 2 (Chính trị):

- Tin giả: "Biểu tình ở Mỹ thực chất do Trung Quốc tài trợ để làm suy yếu phuong Tây"
- Kiểm chứng: Không có bằng chứng tài chính nào được cung cấp

Ví dụ 3 (Khoa học):

- Tin giả: "NASA đã giấu kín bằng chứng về người ngoài hành tinh từ những năm 1960"
- Sự thật: Các tài liệu được trích dẫn đã bị dịch sai hoàn toàn
  - Cách phát hiện

Kiểm tra tính logic:

- Liệu câu chuyện có quá hoàn hảo khi gán ghép mọi thứ thành âm mưu?
- Có bằng chứng trực tiếp hay chỉ là suy diễn?

Phân tích nguồn tin:

- Nguồn tin có phải là chuyên gia trong lĩnh vực?
- Thông tin có được báo chí chính thống đưa tin?

Nhận diện đặc điểm:

- Luôn cho rằng mọi sự trùng hợp đều có chủ ý
- Bác bỏ mọi bằng chứng phản bác như là "một phần của âm mưu"
- + Xuyên tạc sự thật: Bóp méo nội dung sự kiện để gây hiểu lầm, như thay đổi ngữ cảnh hoặc chỉnh sửa hình ảnh. (Boolean/Số thực - So sánh với nguồn tin gốc, kiểm tra mức độ sai lệch)
- Cách thức hoạt động
  - Thủ thuật cắt ghép ngữ cảnh: Lấy sự kiện có thật nhưng bỏ qua bối cảnh quan trọng, chỉ giữ lại phần gây hiểu lầm.
  - Kỹ thuật hình ảnh đánh lừa: Sử dụng ảnh cũ hoặc từ sự kiện khác để minh họa cho tin mới.
  - Phương pháp thông kê gian dối: Trích dẫn số liệu thật nhưng diễn giải sai mục đích hoặc bối cảnh.

- Đặc điểm ngôn ngữ/nội dung

Dấu hiệu nhận biết

- Ngôn ngữ mơ hồ: Sử dụng cụm từ như "nhiều người cho rằng", "theo một số nguồn tin" mà không dẫn nguồn cụ thể.
- Từ ngữ đánh tráo: Thay đổi từ khóa quan trọng trong trích dẫn gốc để làm sai lệch ý nghĩa.
- Công cụ đo lường: Có thể đo tỷ lệ phần trăm sai lệch khi so sánh văn bản gốc với bản đăng lại.

- Ví dụ điển hình

Ví dụ 1 (video):

- Tin giả: "Lãnh đạo X nói 'không cần quan tâm đến dân nghèo'"
- Thực tế: Câu nói bị cắt từ đoạn đầy đủ có ngữ cảnh khác

Ví dụ 2 (hình ảnh):

- Tin giả sử dụng ảnh biếu tình cũ để minh họa cho sự kiện mới
- Kiểm chứng bằng công cụ tìm kiếm hình ảnh ngược

Ví dụ 3 (văn bản):

- Trích dẫn luật bị cắt xén làm sai lệch nội dung gốc
- Cần đối chiếu với văn bản pháp luật đầy đủ
- Cách phát hiện

Kiểm Tra Văn Bản

- So sánh với bản gốc:
  - Tìm kiếm đoạn trích dẫn + tên tác giả/tài liệu gốc
  - Ví dụ: "Lãnh đạo nói X" → tra cứu phát biểu đầy đủ
- Phân tích ngữ cảnh:
  - Đọc toàn bộ văn bản thay vì đoạn trích
  - Kiểm tra xem có bị bỏ qua các từ phủ định như "không", "chưa"

- + Gây bất ổn xã hội: Nội dung kích động biểu tình, bạo loạn, chia rẽ dân tộc hoặc kêu gọi chống đối chính quyền. (Boolean/Số thực - Phát hiện các ngôn ngữ kích động bạo lực)
- Cách thức hoạt động
  - Kích động trực tiếp:
    - Sử dụng lời lẽ kêu gọi hành động bạo lực hoặc bất hợp pháp
    - Ví dụ: "Hãy xuống đường đập phá để thể hiện sự phẫn nộ!"
  - Tạo chia rẽ cộng đồng:
    - Phân loại xã hội thành các nhóm đối lập
    - Ví dụ: "Người dân tộc X đang bị chính quyền đàn áp!"
  - Lợi dụng sự kiện nhạy cảm:
    - Bám vào các vấn đề dễ gây tranh cãi (tôn giáo, dân tộc, bất công xã hội)
- Đặc điểm ngôn ngữ/nội dung

Từ khóa nhận diện

- Kêu gọi hành động:
  - "xuống đường", "đấu tranh", "lật đổ", "nổi dậy"
  - "hãy chung tay", "cần hành động ngay"
- Ngôn ngữ chia rẽ:
  - "chúng ta với họ", "kẻ thù của nhân dân"
  - "sự khác biệt không thể hòa giải"
- Từ ngữ bạo lực:
  - "tiêu diệt", "trùng trị", "trả thù"
  - "máu phai đổ", "hy sinh vì lý tưởng"

Cấu trúc điển hình

- "[Vấn đề] + [Kẻ thù] + [Lời kêu gọi hành động]"  
 "Chính sách thuế mới là bất công! Chính phủ đang bóc lột dân nghèo!  
 Hãy biểu tình vào ngày mai!"
- Ví dụ điển hình

Ví dụ 1 (Biểu tình):

- Tin giả: "Cảnh sát đánh chết người biểu tình ôn hòa - Hãy xuống đường trả thù!"
- Thực tế: Không có trường hợp tử vong nào xảy ra

Ví dụ 2 (Chia rẽ dân tộc):

- Tin giả: "Người dân tộc X bị cấm nói tiếng mẹ đẻ ở trường học"
- Kiểm chứng: Hoàn toàn không có quy định này

Ví dụ 3 (Bạo loạn):

- Tin giả kêu gọi tấn công trụ sở chính quyền địa phương
- Nội dung được đăng trên các nhóm kín Facebook
- Cách phát hiện

Phân tích ngôn ngữ

- Nhận diện từ khóa kích động:
  - Tìm các từ như "xuống đường", "đấu tranh", "lật đổ", "trả thù"
  - Cảnh giác với cụm từ mang tính đối đầu: "chúng ta với họ", "kẻ thù nhân dân"
- Đánh giá giọng điệu:
  - Phát hiện văn phong kích động, kêu gọi hành động tức thì
  - Chú ý các câu mệnh lệnh: "Hãy làm ngay!", "Phải hành động trước khi quá muộn!"

Kiểm tra nguồn phát tán

- Xác minh tài khoản đăng tải:
  - Kiểm tra lịch sử hoạt động, thời gian tạo tài khoản
  - Phát hiện tài khoản giả mạo, clone hoặc bot
- Theo dõi đường lan truyền:
  - Xác định các nhóm/trang chia sẻ nội dung
  - Phát hiện các tài khoản đăng tải cùng lúc nhiều nội dung tương tự

### 3.1.1.4. Bộ dữ liệu đã được xây dựng

| id   | text | title               | content | source | isinstance           | label | normal_text | tin_gia_ve_o_song_xa_hoi | tin_gia_ve_chinh_tr | gay_ho | phonetic | giat_gar | luu_ao | thong_b | boi_rho | thuyet_t | xuyen_t | gay_bat | tu_ngu | text |
|------|------|---------------------|---------|--------|----------------------|-------|-------------|--------------------------|---------------------|--------|----------|----------|--------|---------|---------|----------|---------|---------|--------|------|
|      |      |                     |         |        |                      |       |             | text                     | text                | text   | text     | text     | text   | text    | text    | text     | text    | text    | text   | text |
| 7096 | 7090 | Quốc... My đà ...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7097 | 7091 | Viet... Nga đà ...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7098 | 7092 | Hà N... Hà Nội ...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7099 | 7093 | TPHCM... TPHCM...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7100 | 7094 | Viet... Trung Q...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7101 | 7095 | Chinh... Chính p... | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7102 | 7096 | Hà N... Mỹ đà c...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7103 | 7097 | TPH... Chính p...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7104 | 7098 | Viet... Nga ép ...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7105 | 7099 | Quốc... Trung Q...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7106 | 7100 | Viet... Mỹ đan...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7107 | 7101 | Hà N... Chính p...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7108 | 7102 | TPHCM... TPHCM...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7109 | 7103 | Viet... Trung Q...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7110 | 7104 | Chin... Nga tha...  | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |
| 7111 | 7105 | Hà N... Mỹ âm ...   | 0       | 1      | Tin giả về chính trị | false | 0           | 9                        | 0.0                 | 0.0    | 0.0      | 0        | 0      | 0.0     | 0.0     | 0.0      | 0.0     | 0.0     | 0.0    | 0    |

Hình 3.1: Bộ dữ liệu đã được xây dựng trên PostgreSQL

Dưới đây là một số ví dụ cụ thể về các nguồn báo chí và trang mạng xã hội có thể được sử dụng để xây dựng bộ dữ liệu nghiên cứu tin giả, phân loại theo nhóm tin thật (reliable sources) và tin giả/đáng ngờ Báo chí chính thống tại Việt Nam (có giấy phép, thuộc cơ quan nhà nước hoặc tòa soạn uy tín):

- Báo điện tử: Tuổi Trẻ Online , Thanh Niên ,VnExpress ,VietnamPlus

Dân Trí ,Zing News

- Báo in/báo truyền hình: Báo Nhân Dân (cơ quan ngôn luận của Đảng Cộng sản Việt Nam) ,Đài Truyền hình Việt Nam (VTV) , Đài Tiếng nói Việt Nam (VOV)

Fanpage Facebook chính thống (đã được xác minh – Verified):

- Fanpage chính thức của các báo: Tuổi Trẻ Online , VnExpress , Thanh Niên
- Fanpage cơ quan nhà nước: Bộ Y tế Việt Nam ,Bộ Công an

Nhóm trang giả mạo chính trị (thường đăng tin xuyên tạc, không kiểm chứng):

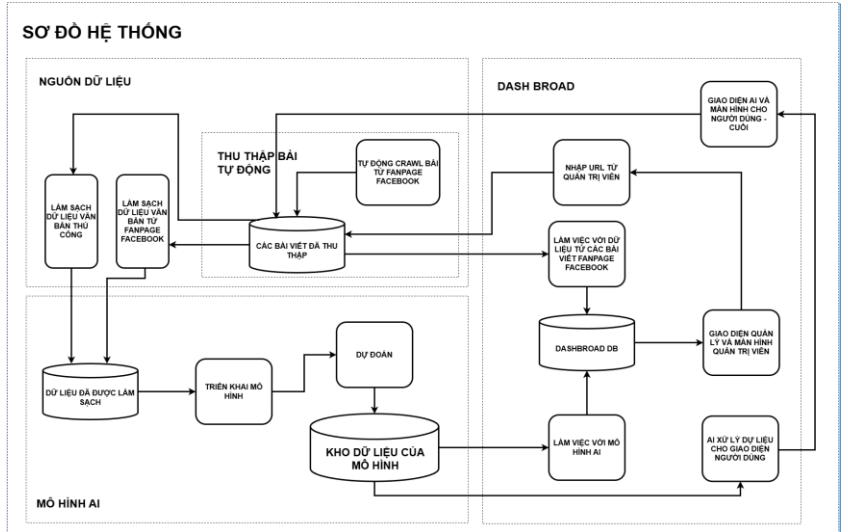
- Các trang liên quan đến Việt Tân (tổ chức được Chính phủ Việt Nam xếp vào nhóm "phản động"):
  - Một số trang Facebook ẩn danh như "Việt Nam Thời Báo" (không phải báo chính thống), "Tiếng Dân", "Báo Mới" (giả mạo tên báo thật).

- Các kênh YouTube/Twitter không rõ nguồn gốc, thường đăng tải thông tin chống phá Nhà nước.

Nhóm trang giả mạo đồi sóng xã hội (báo lá cải, giật tít câu view):

- Trang Facebook đăng tin giật gân, không nguồn:
  - "Tin Nóng 24h", "Đồi Sóng Sóc", "Chuyện Lạ Việt Nam" (thường đăng tin vịt như "người phụ nữ sinh ra rắn", "bí mật động trời về UFO").
  - Các trang clone (nhái) báo chính thống, ví dụ: "VnExpress News" (viết sai chính tả để đánh lừa).
- Trang báo lá cải nước ngoài (nhưng phát tán tại Việt Nam):
  - "Vietnam Daily News" (trang giả mạo, không liên quan đến báo chí Việt Nam).
  - Các trang đăng tin thất thiệt về người nổi tiếng, ví dụ: "Sao Việt Bí Ân".

### 3.1.2. Sơ đồ hệ thống



**Commented [2]:** chúng em đã chỉnh sửa cho phù hợp với bài toán code thực tế của chúng em

Hình 3.2: Sơ đồ hệ thống

## Các module chức năng của hệ thống

### **Module thu thập dữ liệu (Crawler Service):**

- + Thu thập bài viết: Truy cập fanpage, cuộn trang, và trích xuất nội dung bài viết từ các thẻ HTML.
- + Làm sạch dữ liệu: Loại bỏ ký tự đặc biệt, chuẩn hóa khoảng trắng, và đảm bảo dữ liệu phù hợp cho phân loại.
- + Lưu trữ dữ liệu: Lưu bài viết thô và đã làm sạch vào cơ sở dữ liệu.

### **Module phân loại tự động (Training Service)**

- + Phân loại batch: Phân loại đồng thời nhiều bài viết thành ba nhãn: Tin thường, Tin giả về đời sống xã hội, Tin giả về chính trị.
- + Lưu trữ kết quả: Lưu kết quả phân loại vào các bảng trong cơ sở dữ liệu.
- + Quản lý trạng thái: Đánh dấu bài viết đã xử lý để tránh trùng lặp.

### **Module phân loại thủ công (Crawl Service và Dashboard Service)**

- + Nhập bài viết: Nhận nội dung bài viết từ người dùng.
- + Phân loại: Gửi yêu cầu đến Training Service để phân loại bài viết.
- + Lưu trữ kết quả: Lưu bài viết và kết quả phân loại vào cơ sở dữ liệu riêng.

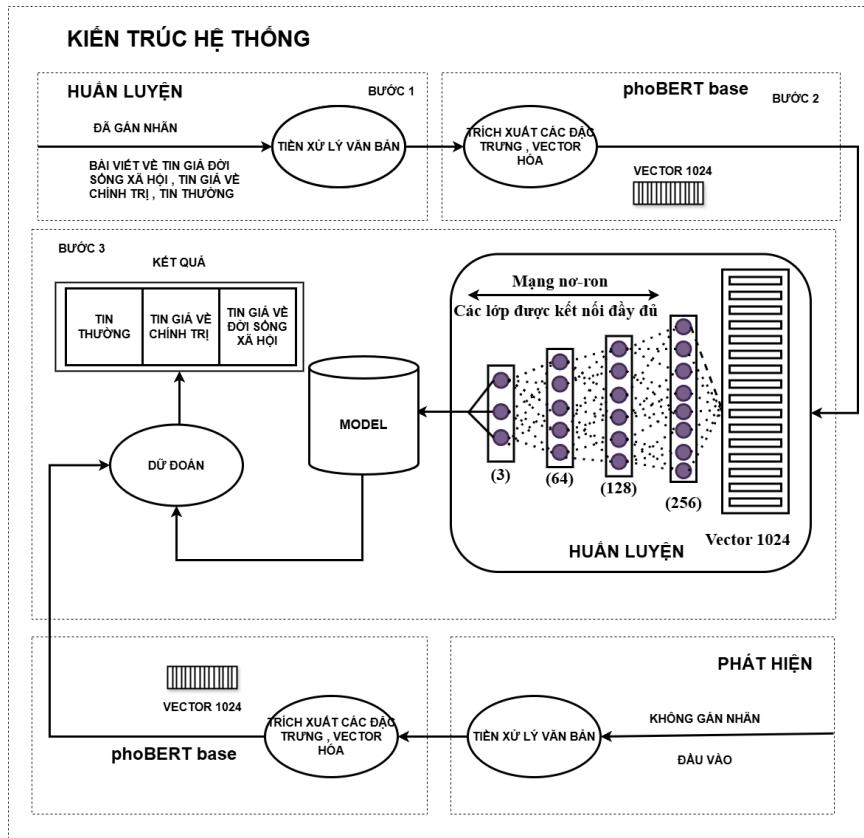
### **Module hiển thị kết quả (Dashboard Service)**

- + Hiển thị bài viết đã phân loại: Truy vấn dữ liệu từ display\_data và các bảng danh mục.
- + Hiển thị dữ liệu thô: Truy vấn dữ liệu từ raw\_fanpage\_facebook.
- + Quản lý kết quả thủ công: Truy vấn kết quả từ classified\_results.

### **Module thống kê và phân tích (Dashboard Service)**

- + Thống kê dữ liệu: Tính toán số lượng bài viết theo ngày và tỷ lệ các loại tin tức.
- + Tìm kiếm bài viết: Tìm kiếm bài viết dựa trên từ khóa.
- + So sánh bài viết: So sánh hai bài viết cụ thể hoặc thống kê tổng quát theo nhãn.

### 3.1.3. Kiến trúc mô hình



Hình 3.3: Sơ đồ kiến trúc hệ thống

Sơ đồ này mô tả kiến trúc hệ thống xử lý ngôn ngữ tự nhiên (NLP) tiếng Việt, với các bước chính từ đầu vào đến đầu ra. Dưới đây là giải thích từng phần:

Đầu vào (Đầu vào):

- Hệ thống nhận các bài viết tiếng Việt từ các nguồn như báo chí, hội thoại, tin tức chính trị, tin thường. Đây là dữ liệu thô ban đầu.

Huấn luyện (Bước 1 - Bước 2):

- Bước 1: Dữ liệu đầu vào được tiền xử lý văn bản (tiền xử lý văn bản) để chuẩn hóa và làm sạch dữ liệu.

- Sau đó, dữ liệu được đưa vào mô hình phoBERT base (một mô hình BERT được huấn luyện trước cho tiếng Việt) để trích xuất các đặc trưng trung (vector hóa), tạo ra các vector 1024 chiều (vector 1024).
- Bước 2: Tiếp tục huấn luyện mô hình phoBERT base trên dữ liệu đã được tiền xử lý để tinh chỉnh (fine-tuning), nhằm cải thiện hiệu suất trên các tác vụ cụ thể.

Kết quả (Bước 3):

- Sau huấn luyện, hệ thống tạo ra mô hình (model) có khả năng xử lý các tác vụ như phân loại tin tức (tin thường), tin giả (tin giả về đối song khai), hoặc tin chính trị (tin giả về chính trị).
- Mô hình này được sử dụng để dự đoán (dự đoán) trên dữ liệu mới.

Mạng nơ-ron (Mạng nơ-ron):

- Dữ liệu được biểu diễn dưới dạng các lớp nơ-ron (3, 64, 128, 256) trong mạng nơ-ron. Các lớp này đại diện cho các tầng xử lý của mô hình, với số nơ-ron tăng dần để học các đặc trưng phức tạp hơn.
- Kết quả cuối cùng sau huấn luyện là các vector 1024 chiều, được sử dụng cho các tác vụ tiếp theo.
- Phát hiện (Bước 4):
  - Dữ liệu đầu vào mới (đầu vào) được tiền xử lý văn bản và đưa qua mô hình phoBERT base để trích xuất đặc trưng (vector hóa).
  - Các vector 1024 chiều này được đưa vào mô hình đã huấn luyện để dự đoán (không gán nhãn), ví dụ: xác định xem một đoạn văn bản có phải là tin giả hay không.

Tóm lại: Hệ thống này sử dụng phoBERT base để xử lý và phân loại văn bản tiếng Việt. Quy trình gồm tiền xử lý, huấn luyện mô hình, và sử dụng mô hình để phát hiện tin giả hoặc phân loại tin tức, với các vector 1024 chiều là đầu ra chính của các bước xử lý.

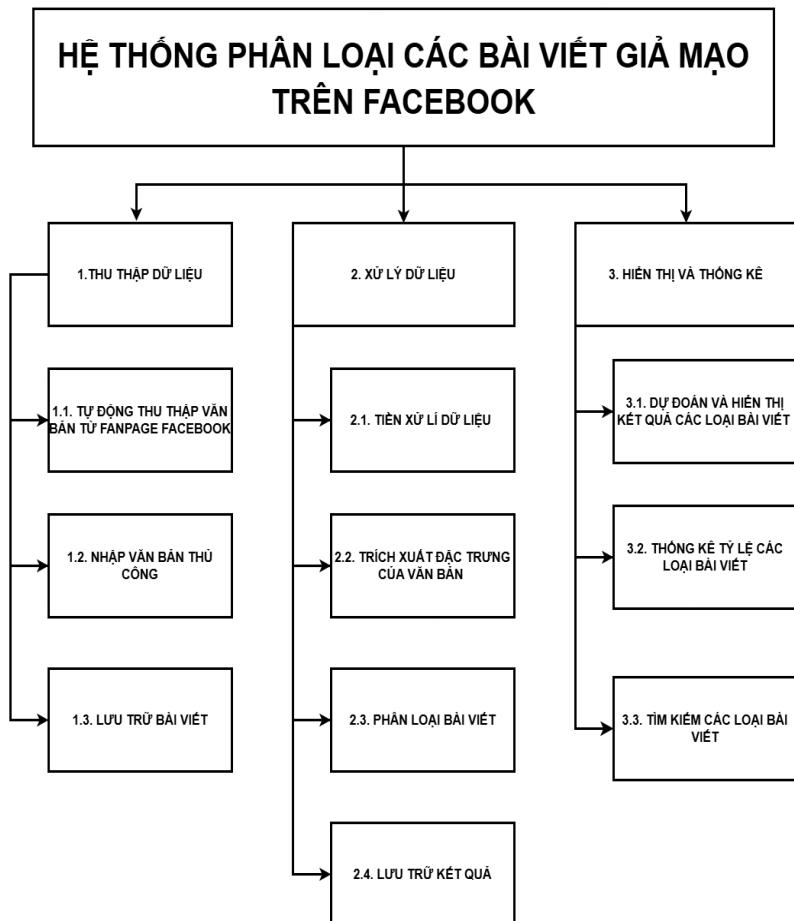
### **3.1.4. Biểu đồ phân cấp chức năng - BFD và sơ đồ luồng dữ liệu - DFD**

#### **3.1.4.1. Gom nhóm chức năng**

*Bảng 3.1: Bảng gom nhóm chức năng*

| STT | Gom nhóm chức năng   | Tên chức năng                                 |
|-----|----------------------|---|
| 1.1 | Thu thập dữ liệu     | Tự động thu thập dữ liệu từ Fanpage Facebook  |
| 1.2 |                      | Nhập văn bản thủ công                         |
| 1.3 |                      | Lưu trữ bài viết                              |
| 2.1 | Xử lý dữ liệu        | Tiền xử lý dữ liệu                            |
| 2.2 |                      | Trích xuất đặc trưng của văn bản              |
| 2.3 |                      | Phân loại bài viết                            |
| 2.4 |                      | Lưu trữ kết quả                               |
| 3.1 | Hiển thị và thống kê | Dự đoán và hiển thị kết quả các loại bài viết |
| 3.2 |                      | Thống kê tỉ lệ các bài viết                   |
| 3.3 |                      | Tìm kiếm các bài viết                         |

*3.1.4.2 Biểu đồ phân rã chức năng*



*Hình 3.4: Biểu đồ phân cấp chức năng - BFD*

### 3.1.4.3. Sơ đồ luồng dữ liệu - DFD

| Khái niệm    | Ký hiệu<br>(DeMarco<br>& Youdon) | Ký hiệu<br>(Gane &<br>Sarson) | Ý nghĩa                                     |
|--------------|----------------------------------|-------------------------------|---|
| Ô xử lý      | ○                                | □                             | Một trong các hoạt động bên trong HTTT      |
| Dòng dữ liệu | →                                | →                             | Sự chuyển đổi thông tin giữa các thành phần |
| Kho dữ liệu  | ↔                                | ↔                             | Vùng chứa dữ liệu, thông tin trong HTTT     |
| Đầu cuối     | □                                | □                             | Một tác nhân bên ngoài HTTT                 |

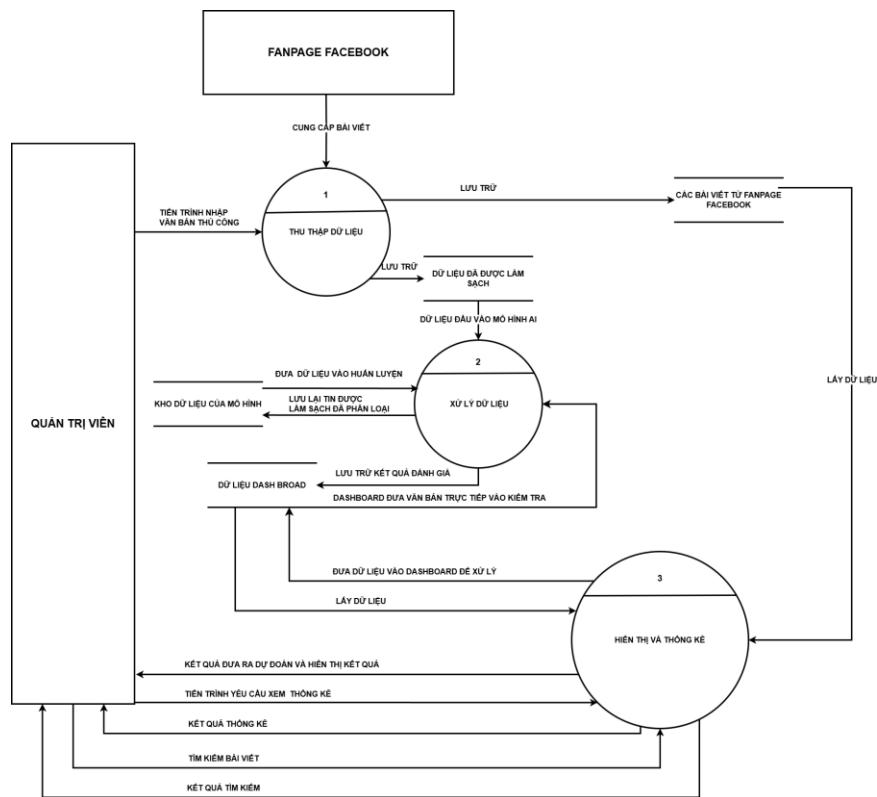
Hình 3.5: Ký hiệu sử dụng

### 3.1.4.4. Sơ đồ mức ngữ cảnh (DFD LEVEL 0)



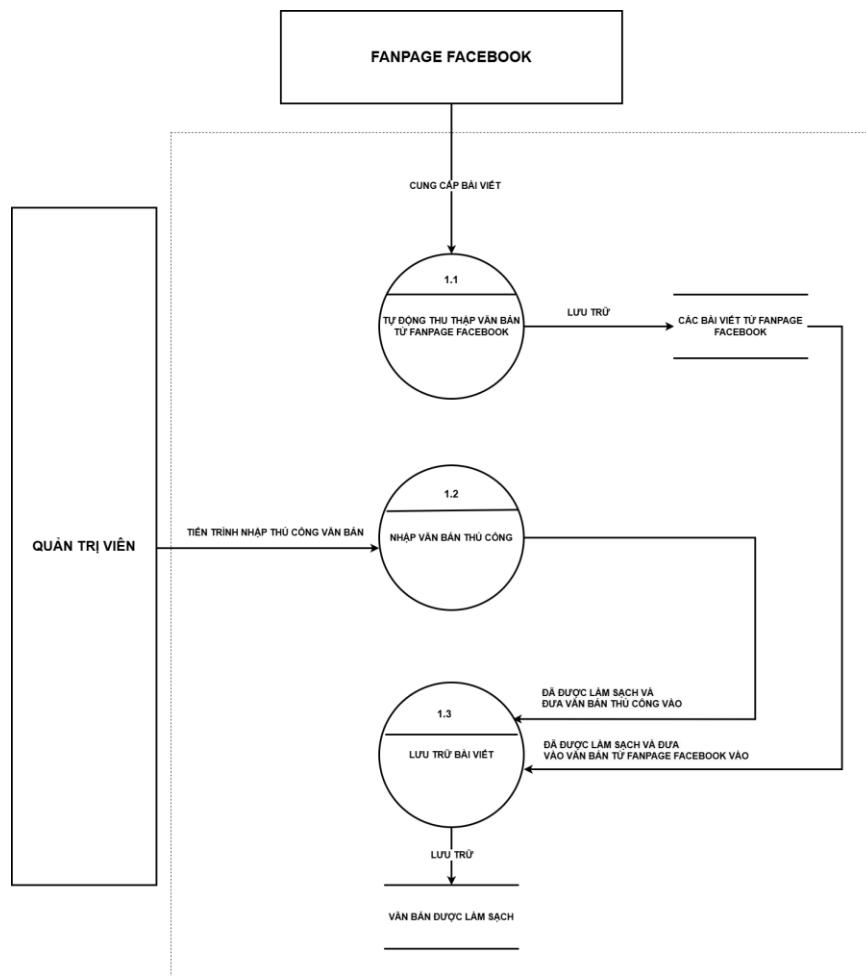
Hình 3.6: Sơ đồ mức ngữ cảnh - DFD level 0

### 3.1.4.5 Sơ đồ luồng dữ liệu mức đỉnh (DFD LEVEL 1)



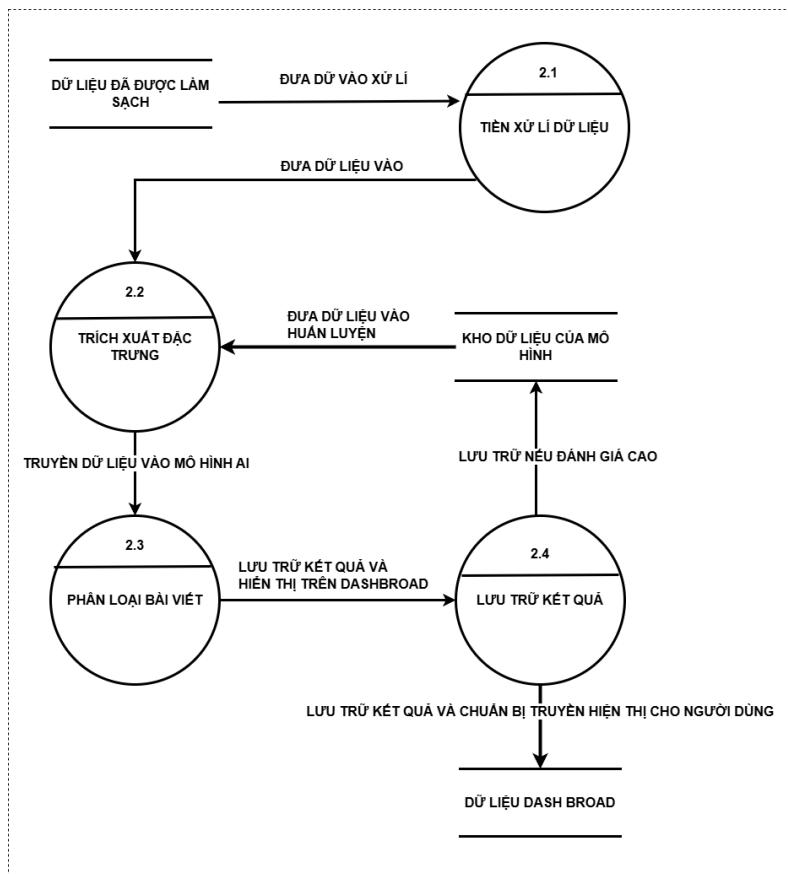
Hình 3.7: Sơ đồ luồng dữ liệu mức đỉnh - DFD level 1 phân rã từ DFD level 0

### 3.1.4.6 Sơ đồ luồng dữ liệu mức dưới đỉnh thu thập dữ liệu



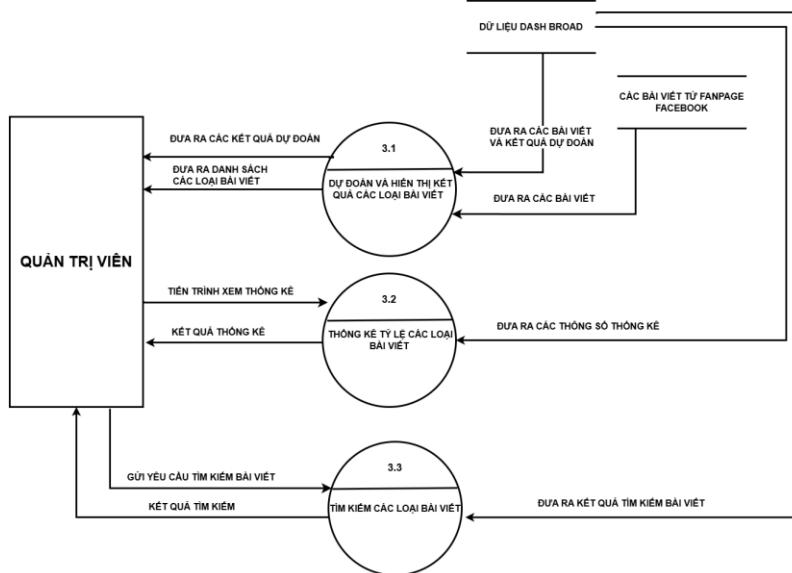
Hình 3.8: Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phân rã từ (1)  
thu thập dữ liệu

### 3.1.4.7 Sơ đồ luồng dữ liệu mức dưới đỉnh xử lý dữ liệu (DFD LEVEL 2)



Hình 3.9 : Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phân rã từ (2)  
xử lý dữ liệu

### 3.1.4.8 Sơ đồ luồng dữ liệu mức dưới đỉnh Hiển thị và thống kê

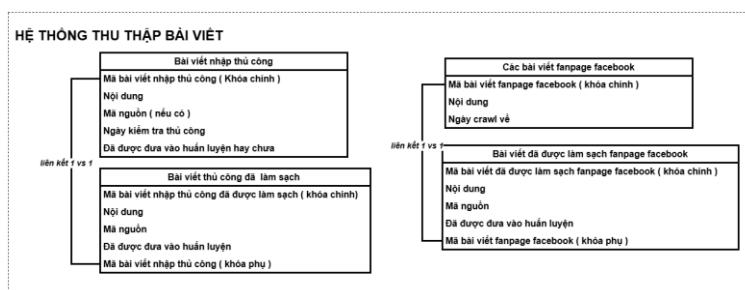


Hình 3.10: Sơ đồ luồng dữ liệu mức dưới đỉnh - DFD Level 2 phan rã từ (3)

Hiển thị và thống kê

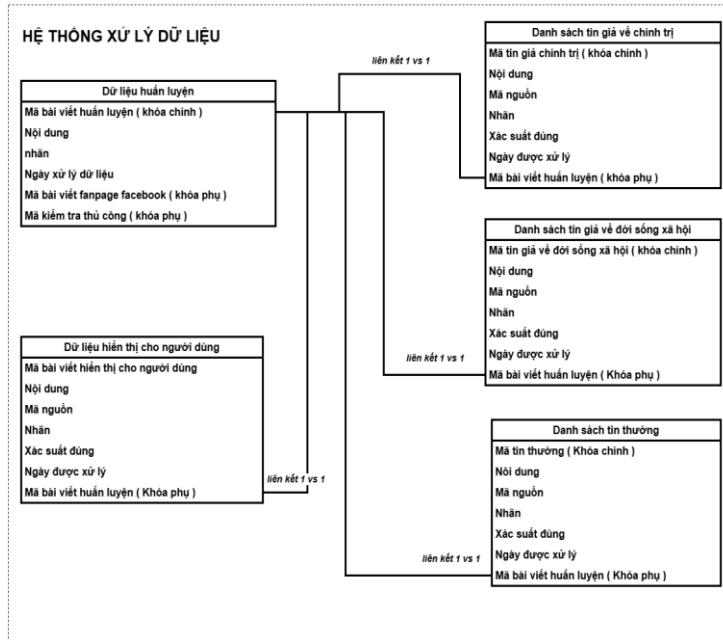
### 3.1.5. Sơ đồ quan hệ dữ liệu - ERD

#### 3.1.5.1 Cơ sở dữ liệu - hệ thống lấy bài viết tự động từ trên fanpage facebook



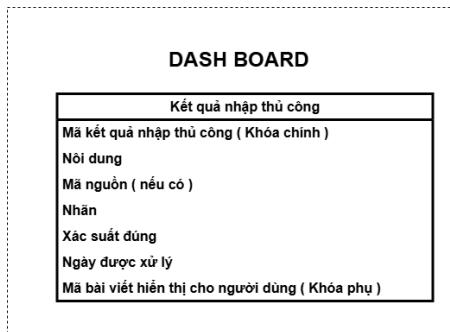
Hình 3.11: Cơ sở dữ liệu - hệ thống crawl bài tự động

### 3.1.5.2 Cơ sở dữ liệu - hệ thống huấn luyện model AI



Hình 3.12: Cơ sở dữ liệu - hệ thống huấn luyện model AI

### 3.1.5.3 Cơ sở dữ liệu - hệ thống dashboard hiển thị cho người dùng



Hình 3.13: Cơ sở dữ liệu - hệ thống dashboard hiển thị cho người dùng

### 3.1.6 CÁC BẢN THIẾT KẾ GIAO DIỆN

#### 3.1.6.1 Giao diện AI & Màn hình (Người dùng - Cuối)

The screenshot shows a user interface for searching news articles. At the top, there are three buttons: 'Kiểm tra tin giả' (Check fake news), 'Admin', and 'Crawl & lọc' (Crawl & filter). Below these are two tabs: 'Kiểm tra và xử lý tin giả' (Check and handle fake news) and 'Hiển thị câu hỏi' (Display questions). A large text input field labeled 'Hiển thị trả lời' (Display answer) is present. At the bottom, there is a search bar with the placeholder 'Nhập tin cần kiểm tra...' (Enter news to check) and a yellow 'Gửi' (Send) button.

Hình 3.14: Giao diện AI và Màn hình (Người dùng - Cuối)

#### 3.1.6.2 Giao diện Admin panel

##### Thống kê



Hình 3.15: Giao diện Admin panel - thống kê

**Danh sách các bài viết ( bài thường , bài giả về đời sống xã hội , bài giả về chính trị , các bài đã cào từ fanpage facebook )**



Hình 3.16: Giao diện Admin panel - danh sách các bài viết

**3.1.6.3 Giao diện Crawl - Lọc**



Hình 3.17: Giao diện Admin panel - crawl và lọc

### 3.2. Triển khai hệ thống

#### 3.2.1 Chức năng Phân loại bài viết

##### Tiền xử lý và chia dữ liệu

- Ánh xạ nhãn văn bản thành số để mô hình học máy có thể xử lý. Điều này cần thiết vì các mô hình như neural network yêu cầu đầu vào số
- Ta lấy 60% tin đã được lấy ngẫu nhiên để train
- Và 20% tin trong 40% còn lại để kiểm tra và 20% còn lại để đánh giá

##### Trích xuất đặc trưng

Chuẩn hóa văn bản, chuyển thành chuỗi và cắt ngắn nếu vượt quá max\_length  
Đầu vào : texts = ["Chính phủ ban hành chính sách", "Tin giả về chính trị"],  
max\_length = 10

đầu ra : texts = ["Chính phủ ba", "Tin giả về "]

-Lấy một batch văn bản từ danh sách texts.

Đầu vào : texts = ["Chính phủ ba", "Tin giả về "], batch\_size = 2, i = 0.

Đầu ra : batch\_texts = ["Chính phủ ba", "Tin giả về "]

-Mã hóa batch văn bản thành tensor cho PhoBERT.

Đầu vào: batch\_texts = ["Chính phủ ba", "Tin giả về "]

Đầu ra : encoded\_input = {

```
'input_ids': tensor([[101, 1234, 5678, 9012, 102, 0],
[101, 2345, 6789, 3456, 102, 0]]),
'attention_mask': tensor([[1, 1, 1, 1, 1, 0],
[1, 1, 1, 1, 1, 0]])
```

}

- Chạy PhoBERT để tạo biểu diễn ngữ nghĩa.

- đầu vào : endcoded\_input

đầu ra :model\_output.last\_hidden\_state = tensor([

```
[[0.1, -0.2, ..., 0.3], # [CLS]
```

```
[0.2, -0.1, ..., 0.4], # Chính
```

```

...,           # phủ, ba, [SEP] , pad
[0.0, 0.0, ..., 0.0]], # pad
[[0.3, -0.4, ..., 0.5], # [CLS]
[0.4, -0.3, ..., 0.6], # Tin
...,           # giả, vè, [SEP] , pad
[0.0, 0.0, ..., 0.0]] # pad
)

```

Trích xuất vector [CLS] và chuyển thành mảng NumPy

Đầu vào : last\_hidden\_state (shape: (2, 6, 768))

Đầu ra batch\_features = array([

```
[0.1, -0.2, ..., 0.3], # [CLS] cho "Chính phủ ba"
```

```
[0.3, -0.4, ..., 0.5] # [CLS] cho "Tin giả vè "
```

], dtype=float32)

-Gộp các đặc trưng từ tất cả batch thành ma trận.

Đầu vào: features = [array([[0.1, -0.2, ..., 0.3], [0.3, -0.4, ..., 0.5]])]

Đầu ra array([

```
[0.1, -0.2, ..., 0.3], # Đặc trưng cho "Chính phủ ba"
```

```
[0.3, -0.4, ..., 0.5] # Đặc trưng cho "Tin giả vè "], dtype=float32)
```

| Tầng                              | Thuật toán          | Công thức    | Ví dụ minh họa   | Vai trò  |
|-----------------------------------|---------------------|--------------|--|--|
| FC1,<br>FC2,<br>FC3, FC4<br>, FC5 | Biến đổi tuyến tính | $y = Wx + b$ | $x = [0.1, 0.2] \rightarrow y = [0.51, 1.12]$<br>(fc1) | Tạo biểu diễn đặc trưng, là đầu vào cho tầng tiếp theo; tối ưu hóa trong huấn luyện. |

|         |                   |   |   |  |
|---------|-------------------|---|---|--|
| ReLU    | ReLU              | $f(x) = \max(0, x)$                                 | $x = [0.51, -1.12] \rightarrow y = [0.51, 0]$                               | Thêm tính phi tuyến, cải thiện khả năng học mẫu phức tạp.            |
| Dropout | Bernoulli Dropout | $y = x * \text{mask} / (1 - \text{dropout\_rate})$  | $x = [0.51, -1.12] \rightarrow y = [0.6375, 0]$ (mask = [1, 0])             | Dự đoán nhẫn, hiển thị xác suất trong suy luận và đánh giá.          |
| Softmax | Softmax           | $\text{softmax}(x_i) = \exp(x_i) / \sum(\exp(x_j))$ | $\text{logits} = [19.09, 9.56, 1.94] \rightarrow \text{probs} = [1.0, ...]$ | Tính toán đầu ra (logits, probs) cho huấn luyện, suy luận, đánh giá. |

Bảng 3.2: Bảng thuật toán trong mạng nơ ron của PhoBERT

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} & \text{F1} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} & \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \end{aligned}$$

Hình 3.18: Các chỉ số đánh giá

- Precision: Đo tỷ lệ dự đoán dương tính đúng (ít nhầm tin thật thành giả).
- Recall: Đo khả năng tìm ra tất cả tin giả (ít bỏ sót tin giả).
- Accuracy: Tỷ lệ dự đoán đúng tổng thể.
- F1 Score: Trung bình hài hòa của Precision và Recall, cân bằng cả hai

Để đánh giá hiệu quả của mô hình phân loại tin tích cực, tin tiêu cực, tin trung lập, các độ đo quan trọng đã được sử dụng bao gồm Accuracy, Precision, Recall và F1 Score. Các độ đo này giúp xác định mức độ chính xác và hiệu suất của mô hình trong việc phân loại nội dung tích cực, tiêu cực, hoặc trung lập. Dưới đây là định nghĩa và ý nghĩa của từng độ đo:

- **Accuracy (Độ chính xác)**

Accuracy thể hiện tỷ lệ các dự đoán đúng (bao gồm cả tin tích cực và tin tiêu cực) trên tổng số mẫu dự đoán. Công thức tính như sau:

$$\text{Accuracy} = \frac{TP+TN}{\text{TotalSamples}}$$

*Hình 3.19: Công thức Accuracy*

- Trong đó:

- (TP) (True Positive): Số lượng tin tích cực được dự đoán đúng.
- (TN) (True Negative): Số lượng tin tiêu cực được dự đoán đúng.

Độ đo này cho biết mức độ chính xác tổng thể của mô hình trên toàn bộ tập dữ liệu

- **Precision (Độ chính xác của dự đoán tích cực)**

Precision đo lường tỷ lệ các dự đoán tích cực thực sự đúng trong tổng số các dự đoán được mô hình nhận định là tích cực. Công thức tính:

$$\text{Precision} = \frac{TP}{TP+FP}$$

*Hình 3.20 : Công thức Precision*

- Trong đó:

- (TP): Số lượng tin tích cực được dự đoán đúng (như trên).
- (FP) (False Positive): Số lượng tin tiêu cực bị dự đoán sai thành tích cực.
- Precision giúp đánh giá độ tin cậy của mô hình khi dự đoán một nội dung là tích cực.

- **Recall (Độ bao phủ của tin tích cực)**

Recall đo lường tỷ lệ các tin tích cực được mô hình phát hiện đúng trên tổng số tin thực sự là tích cực. Công thức tính:

$$\text{Recall} = \frac{TP}{TP+FN}$$

*Hình 3.21: Công thức Recall*

- Trong đó:
    - (TP): Số lượng tin tích cực được dự đoán đúng (như trên).
    - (FN) (False Negative): Số lượng tin tích cực bị dự đoán sai thành tiêu cực.
- Recall cho biết khả năng mô hình nhận diện đầy đủ các nội dung tích cực.

● **F1 Score (Độ đo trung bình hài hòa)**

F1 Score là trung bình hài hòa giữa Precision và Recall, được sử dụng để đánh giá hiệu suất tổng thể của mô hình khi cần cân bằng giữa Precision và Recall. Công thức tính:

$$F1 = \frac{2 \cdot (\text{precision} \cdot \text{recall})}{\text{precision} + \text{recall}}$$

*Hình 3.22: Công thức F1 Score*

F1 Score đặc biệt hữu ích khi tập dữ liệu không cân bằng, giúp đánh giá toàn diện hiệu quả phân loại của mô hình.

```
(venv) PS D:\Project\DA\DATNV2\creatPTH> py v4.py
Using device: cpu
Đang tải PhoBERT...
Đã tải PhoBERT thành công!
Đang tải ConvNet model...
Đã tải ConvNet thành công!
Đánh giá mô hình...
Accuracy: 0.8775, Precision: 0.8864, Recall: 0.8775, F1 Score: 0.8789
Confusion Matrix:
[[682  62  31]
 [ 13 390  16]
 [ 19  55 332]]
```

*Hình 3.23: Kết quả đánh giá Dataset của hệ thống*

### 3.2.2 So sánh hiệu suất với các mô hình liên quan

| Tiêu chí               | Hệ thống của tôi   | Bài báo 1   | Bài báo 2                                     | Bài báo 3   |
|------------------------|--|---|---|---|
| Mô hình                | PhoBERT + ConvNet  | PhoBERT + CNN/LSTM                                | PhoBERT + SVM/Logistic Regression , phân cụm  | Naive Bayes , SVM , CNN, RNN , LSTM                               |
| Độ chính xác (độ đoán) | ~85-95% (dựa trên Accuracy 0.8777, F1-score 0.8789)        | Vượt trội hơn Naive Bayes, SVM (ước lượng 85-90%) | Cao hơn SVM, Random Forest (ước lượng 80-85%) | CNN/LSTM cải thiện so với học máy truyền thống (ước lượng 85-90%) |
| Tiền xử lý dữ liệu     | Tương tự, sử dụng preprocess_for_phobert                   | Loại bỏ ký tự đặc biệt, chuẩn hóa, tokenization   | Tương tự, có phân cụm dữ liệu                 | Không đề cập chi tiết   |
| Lưu trữ dữ liệu        | Lưu vào display_data với cột probabilities JSONB           | Không đề cập chi tiết                             | Không đề cập chi tiết                         | Không đề cập chi tiết   |
| Ứng dụng thực tế       | Giám sát tin tức, phân tích xu hướng, kiểm chứng thông tin | Ứng dụng trên web, mạng xã hội                    | Hỗ trợ phát hiện tin giả trên mạng xã hội     | Đề xuất cải thiện phương pháp tự động                             |

Bảng 3.3 : Bảng so sánh hiệu các với các đề tài liên quan

**Commented [8]:** thêm bảng so sánh hiệu suất với các đề tài liên quan đã nghiên cứu

Kết luận

- Hệ thống của tôi thể hiện hiệu suất cao (Accuracy 0.8777, F1-score 0.8789, dự đoán 85-95%) và cạnh tranh với các nghiên cứu trong Bài báo 1, 2, 3.
- Hệ thống của tôi tương đồng với Bài báo 1 về mô hình và tiền xử lý, vượt trội hơn Bài báo 2 và Bài báo 3 về lưu trữ dữ liệu (JSONB) và ứng dụng thực tế.
- Để tối ưu hơn, Hệ thống của tôi có thể áp dụng kỹ thuật tăng cường dữ liệu (như Bài báo 2 đề xuất) hoặc tinh chỉnh PhoBERT (như Bài báo 1 gợi ý).

### **3.2.3 Chức năng crawl dữ liệu từ bài viết**

Sử dụng thư viện Selenium

```
from selenium import webdriver
```

tự động truy cập vào google chrome đã được đăng nhập facebook và profile google

file: config.py ( crawl\_services)

```
def crawl_facebook_page(url: str, target_posts: int) -> list:
    options = webdriver.ChromeOptions()
    options.add_argument(f"user-data-dir={PROFILE_PATH}")
    options.add_argument("--profile-directory=Default")
    options.add_argument("--disable-blink-features=AutomationControlled")
    options.add_argument("--start-maximized")

    service = Service(CHROMEDRIVER_PATH)
    driver = webdriver.Chrome(service=service, options=options)
```

tự động scroll trang để tải dữ liệu trang website để có được nhiều thẻ hơn trong xpath lấy được số tin bằng với số max\_attempts quy định trong biến môi trường file : env ( crawl\_service)

```
PROFILE_PATH=C:/Users/proca/AppData/Local/Google/Chrome/User
Data
TARGET_POSTS=10
CRAWL_SERVICE_PORT=8001
```

file crawler.py ( crawl\_services)

```
for attempt in range(max_attempts):
    driver.execute_script("window.scrollTo(0,
document.body.scrollHeight);")
    time.sleep(3)
```

### **3.3. Kết quả hệ thống**

#### **3.3.1. Giao diện AI & Màn hình (Người dùng - Cuối)**

##### **3.3.1.1. Mô tả**

Endpoint /manual-classify trong Crawl Service cho phép người dùng nhập bài viết để phân loại thành Tin thường, Tin giả về đời sống xã hội, hoặc Tin giả về chính trị

Mã nguồn : File: crawl\_service/routes.py

### **Luồng xử lý**

Nhận yêu cầu:

- Người dùng gửi POST đến /manual-classify với TextInput (content, url tùy chọn).
- Dữ liệu được xác thực qua Pydantic (models.py).

Lưu bài viết:

- Lưu vào bảng manual\_entries (content, url, created\_at) trong crawl\_db, trả về manual\_entry\_id.
- Lưu vào bảng filtered\_data (content, url, is\_approved = TRUE).

Gửi yêu cầu phân loại:

- Gửi POST đến training\_service: /classify-manual (URL: http://localhost:8002/classify-manual) với JSON chứa content và url.
- Training Service sử dụng PhoBERT và ConvNet để phân loại, trả về label, probability, và probabilities.

Lưu kết quả:

- Lưu kết quả (content, url, label, probability, probabilities) vào bảng classified\_results trong crawl\_db.

Trả về kết quả:

- Trả về JSON chứa message, label, probability, và probabilities cho người dùng.

### 3.3.1.2. Kết quả

Hình 3.24: Kết quả đánh giá Dataset của hệ thống

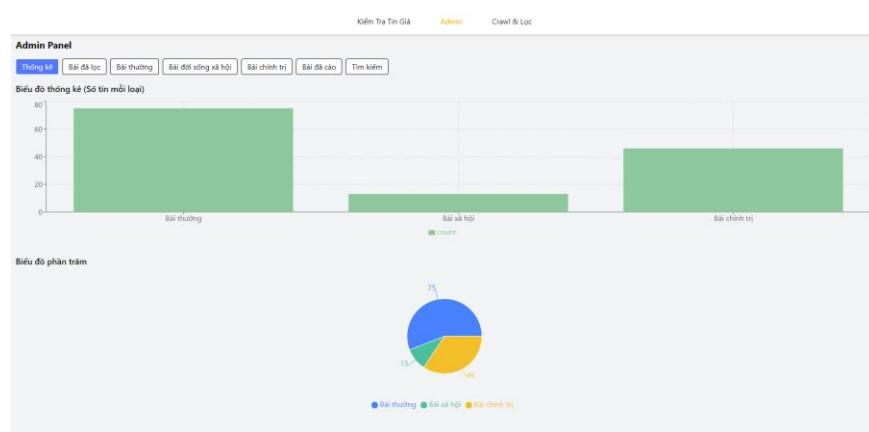
### 3.3.2. Giao diện Admin panel - thống kê

#### 3.3.2.1. Mô tả

Endpoint /stats trong Dashboard Service cung cấp chức năng thống kê dữ liệu bài viết đã phân loại, bao gồm số lượng bài viết theo ngày (time\_stats), tổng số bài viết (total\_articles), và tỷ lệ phần trăm các loại tin tức (ratios). Endpoint truy vấn bảng display\_data trong dashboard\_db, hỗ trợ lọc theo khoảng thời gian, và trả về kết quả dạng JSON cho frontend hiển thị.

Mã nguồn : File: dashboard\_service/routes.py

#### 3.3.2.2. Kết quả



Hình 3.25: Kết quả thống kê của Admin Panel

### 3.3.3. Giao diện Admin panel - danh sách các bài viết

#### 3.3.3.1. Mô tả

Các chức năng hiển thị danh sách bài viết theo danh mục cho phép người dùng xem các bài viết đã phân loại từ dashboard\_db, bao gồm bài viết thường (normal\_news), bài viết giả về đời sống xã hội (fake\_social\_news), và bài viết giả về chính trị (fake\_political\_news). Các endpoint /normal-news, /fake-social-news, và /fake-political-news trong dashboard\_service truy vấn dữ liệu từ các bảng tương ứng, trả về danh sách bài viết để hiển thị trên giao diện frontend.

File: dashboard\_service/routes.py

## Luồng xử lý

Nhận yêu cầu:

- Người dùng gửi GET đến một trong các endpoint /normal-news, /fake-social-news, hoặc /fake-political-news với tham số limit (mặc định 100).

Kết nối cơ sở dữ liệu:

- Kết nối đến dashboard\_db bằng connect\_dashboard\_db() (sử dụng psycopg2 với cấu hình từ DASHBOARD\_DB\_CONFIG).

Truy vấn dữ liệu:

- Thực hiện truy vấn SQL trên bảng tương ứng (normal\_news, fake\_social\_news, hoặc fake\_political\_news).
- Lấy các cột content, url, probability, processed\_at, sắp xếp theo processed\_at giảm dần, giới hạn bởi limit.

Xử lý kết quả:

- Chuyển dữ liệu từ truy vấn thành danh sách JSON với các trường content, url, probability, processed\_at.
- Trả về kết quả:
- Trả về JSON chứa danh sách bài viết cho frontend hiển thị.

### 3.3.3.2. Kết quả

| Admin Panel  |                                   |                  |
|--|-----------------------------------|------------------|
| Thống kê   | Bài đã lọc                        | Bài thường       |
| Bài đổi sống xã hội  | Bài chính trị                     | Bài đã cáo       |
|  |                                   | Admin            |
|  |                                   | Crawl & Lọc      |
|  |                                   | Kiểm Tra Tin Giả |
|  |                                   | Tìm kiếm         |
| Nội dung   | URL                               | Xác suất         |
| Sau một thời gian ngắn xác minh Công an phường Hoàng Liệt đã xác định được chủ tài sản là anh Trương Quang Nhật SN 1980 và tên hành trao trả toàn bộ số tài sản trên cho anh Trương Quang Nhật   | https://www.facebook.com/Theanh28 | 0.9999           |
| Sau hơn một thế kỷ mòn mỏi chờ đợi của cả co thể đơn hài cốt con trai trở về vào ngày 34/2025  | https://www.facebook.com/Theanh28 | 0.9980           |
| Ở cái tuổi xưa nay hiếm cụ Phạm Thị Lài 104 tuổi mẹ liệt sĩ Nguyễn Công Hòa đã quên dù thứ nhưng chưa bao giờ cụ quên người con đang nằm lai đâu đó ở chiến trường   | https://www.facebook.com/Theanh28 | 0.9998           |
| Đôi tay đã nhăn nhó của mẹ liên tục sờ lên là cổ phủ trên hõi đất liệt sĩ Nguyễn Công Hòa SN 1951 quê xã Đại Đồng huyện Thanh Chương tỉnh Nghệ An khiến ai cũng xúc động   | https://www.facebook.com/Theanh28 | 1.0000           |
| Hơn nửa thế kỷ trôi qua người mẹ đã chờ đợi được giây phút đón con trai trở về bên bầy giờ cụ đã thực hiện được ước nguyện lớn nhất đời mình   | https://www.facebook.com/Theanh28 | 1.0000           |
| Ở cái tuổi xưa nay hiếm cụ Phạm Thị Lài 104 tuổi mẹ liệt sĩ Nguyễn Công Hòa đã quên dù thứ nhưng chưa bao giờ cụ quên người con đang nằm lai đâu đó ở chiến trường   | https://www.facebook.com/Theanh28 | 0.9998           |
| Trước đó CTH đã công bố điểm thi vào lớp 10 Trường THPT Lê Hồng Phong là Toán 8 điểm Ngữ văn 8.5 điểm Tiếng Anh 6.4 điểm theo công thức môn Toán và Văn nhân đổi công với điểm tiếng Anh thi sinh này đạt tổng 39/50 điểm trở thành thủ khoa đầu vào năm học 2024-2025 của Trường THPT Lê Hồng Phong | https://www.facebook.com/Theanh28 | 0.9976           |
| Tuy nhiên chỉ sau khi vào học được thời gian ngắn có điểm tố cáo được ký tên là học sinh Trường THPT Lê Hồng Phong gửi đến Sở  | https://www.facebook.com/Theanh28 | 0.9958           |

Hình 3.26: Kết quả danh sách bài viết Admin Panel

### **3.3.4. Giao diện Admin panel - crawl và lọc**

#### **3.3.4.1. Mô tả**

Chức năng thu thập dữ liệu (crawl) trong crawl\_service sử dụng endpoint /crawl để thu thập bài viết từ fanpage Facebook, làm sạch và lưu trữ vào crawl\_db (bảng raw\_fanpage\_facebook, clean\_fanpage\_facebook, filtered\_data). Chức năng lọc dữ liệu áp dụng thuật toán tiền xử lý để loại bỏ ký tự đặc biệt, chuẩn hóa văn bản, và đánh dấu trạng thái phê duyệt, đảm bảo dữ liệu sẵn sàng cho phân loại.

Crawl Service: Endpoint /crawl

File: crawl\_service/routes.py

Hàm thu thập và lọc dữ liệu

File: crawl\_service/crawler.py

File: crawl\_service/database.py

#### **Luồng xử lý**

Nhận yêu cầu:

- Người dùng gửi POST đến /crawl với CrawlRequest (url, target\_posts).
- Dữ liệu được xác thực qua Pydantic (models.py).

Thu thập dữ liệu:

- Hàm crawl\_facebook\_page sử dụng Selenium để truy cập fanpage, cuộn trang, và trích xuất nội dung từ thẻ HTML (userContentWrapper).
- Lấy tối đa target\_posts bài viết, lưu vào danh sách posts.

Lưu dữ liệu thô:

- Lưu bài viết vào bảng raw\_fanpage\_facebook (content, created\_at).

Lọc và lưu dữ liệu sạch:

- Áp dụng preprocess\_for\_phobert để làm sạch văn bản.
- Lưu vào clean\_fanpage\_facebook (content, url, process = FALSE).

Lưu dữ liệu lọc:

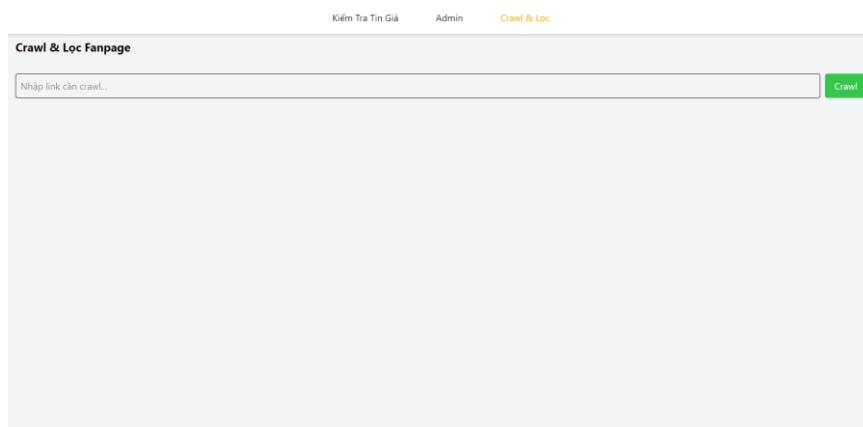
- Lưu vào filtered\_data (content, url, is\_approved = FALSE).

Trả về kết quả:

- Trả về JSON chứa url, posts, và thông báo số bài viết đã lưu.

### 3.3.4.2. Kết quả

Bước 1 Nhập một đường URL fanpage facebook rồi ấn Crawl



Hình 3.27: Kết quả giao diện nhập URL của người dùng

Bước 2 Án lọc

| Nội dung  |
|---|
| Trong thời gian này, các cơ quan, công sở trong phạm vi cả nước và các cơ quan đại diện của Việt Nam tại nước ngoài treo cờ rủ, có dài băng tang như quy định; không tổ chức các hoạt động vui chơi, giải trí công cộng.  |
| Canada chính thức áp thuế "trả đũa" Mỹ.   |
| "Mọi quốc gia đều gọi cho tôi. Đó là vẻ đẹp của những gì chúng ta làm. Chúng ta ở vị trí làm chủ tình hình", ông Trump nói.   |
| Nóng Tổng thống Hàn Quốc Yoon Suk-yeol chính thức bị phê trật   |
| Greenland có dân số gần 57.000 người và diện tích khoảng 2.16 triệu km <sup>2</sup> , nằm trên tuyến đường ngắn nhất từ Bắc Mỹ tới châu Âu. Hòn đảo có nguồn khoáng sản dồi dào, phần lớn chưa được khai thác.            |
| Để tỏ lòng thương tiếc, đồng chí Khamtay Siphandone, Chính phủ nước CHXHCN Việt Nam và Ủy ban Trung ương Mặt trận Tổ quốc Việt Nam quyết định để tang đồng chí Khamtay Siphandone theo nghi thức Quốc tang từ 4-5/4/2023. |
| TREO CỜ RỦ QUỐC TANG ĐỒNG CHÍ KHAMTAY SIPHANDONE - NGUYỄN CHỦ TỊCH NƯỚC CỘNG HÒA DÂN CHỦ NHÂN DÂN LÀO   |
| Tổng thống Trump từng nhiều lần tuyên bố muốn sáp nhập Greenland vào Mỹ và không loại trừ phương án sử dụng vũ lực để làm điều này.   |
| #tuancy #babbling #2unique #2funquepodcast  |
| Hiện chưa rõ "thứ tốt đẹp" mà ông Trump đề cập trong các cuộc thương lượng là gì, nhưng ông chủ Nhà Trắng khẳng định đơn thuế mang lại cho Mỹ sức mạnh to lớn trên tầm phản.  |

Hình 3.28: Kết quả crawl của Admin Panel

### Bước 3 Sẽ nhảy sang phần admin để hiển thị kết quả

Kiểm Tra Tin Giả Admin Crawl & Lọc

**Admin Panel**

Thống kê Bài đã lọc Bài thường Bài đời sống xã hội Bài chính trị Bài đã cáo Tim kiếm

| Nội dung  | URL   | Loại                 | Xác suất |
|---|---|----------------------|----------|
| Trong thời gian này các cơ quan công sở trong phạm vi cả nước và các cơ quan đại diện của Việt Nam tại nước ngoài treo cờ rủ có dài băng tang như quy định không tổ chức các hoạt động vui chơi giải trí công cộng  | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.9998   |
| Hiện chưa rõ thứ tốt đẹp mà ông Trump đề cập trong các cuộc thương lượng là gì nhưng ông chủ Nhà Trắng khẳng định đòn thuế mang lại cho Mỹ sức mạnh to lớn trên tầm phản  | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin giả về chính trị | 0.9999   |
| tuancriy bacding 28unique 28uniquepodcast   | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.9994   |
| Tổng thống Trump từng nhiều lần tuyên bố muốn sáp nhập Greenland vào Mỹ và không loại trừ phương án sử dụng vũ lực để làm điều này  | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin giả về chính trị | 1.0000   |
| TREO CỜ RỦ QUỐC TANG ĐÔNG CHÍ KHAMTAY SIPHANDONE NGUYỄN CHỦ TỊCH NƯỚC CỘNG HÒA DÂN CHỦ NHÂN DÂN LÀO   | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin giả về chính trị | 1.0000   |
| Để tỏ lòng thương tiếc đồng chí Khamtay Siphandone Chính phủ nước CHXHCN Việt Nam và Ủy ban Trung ương Mặt trận Tổ quốc Việt Nam quyết định để tang đồng chí Khamtay Siphandone theo nghi thức Quốc tang từ 4542025 | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 1.0000   |
| Greenland có dân số gần 57000 người và diện tích khoảng 216 triệu km <sup>2</sup> nằm trên tuyến đường ngắn nhất từ Bắc Mỹ tới châu Âu  | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.9997   |
| Hòn đảo có nguồn khoáng sản dồi dào phản ứng chưa được khai thác  | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.9997   |
| Nóng Tổng thống Hàn Quốc Yoon Sukyeol chính thức bị phê truất   | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.9397   |
| Mỗi nước nà đều nói rằng tôi là và đến cửa nhàm nà chung ta làm Chung ta ở vị trí làm chủ tình hình Ann Trump nói   | <a href="https://www.facebook.com/Theanh28">https://www.facebook.com/Theanh28</a> | Tin thường           | 0.5044   |

Hình 3.29: Kết quả lọc các bài viết URL fanpage Facebook của Admin Panel

## TÀI LIỆU THAM KHẢO

**Commented [9]:** Sửa lại tham khảo cho đúng định dạng

- [1] Santosh Kumar Uppada, Parth Patel & Sivaselvan B., An image and text-based multimodal model for detecting fake news in OSN's:

<https://link.springer.com/article/10.1007/s10844-022-00764-y>

- [2] Towards explainable fake news detection and automated content credibility assessment: Polish internet and digital media use-case

<https://tinyurl.com/3rr7trvm>

- [3] Discriminating AI-generated Fake News

<https://tinyurl.com/4mts4m7z>

- [4] “Cẩm nang phòng chống tin giả, tin sai sự thật trên không gian mạng”, Bộ Thông tin và Truyền thông, 2022.

[https://drive.google.com/file/d/1cnVWfLp7G8cHHoSJYuLCJbJ8s87DeH\\_0/vie](https://drive.google.com/file/d/1cnVWfLp7G8cHHoSJYuLCJbJ8s87DeH_0/view)

- [5] Dat Quoc Nguyen, Anh Tuan Nguyen. *PhoBERT: Pre-trained language models for Vietnamese*, 2020.

- [6] TRƯƠNG QUỐC ĐỊNH, PHAN THỊ THÚY KIỀU, Nghiên cứu giải pháp phát hiện tin giả trên mạng xã hội bằng ngôn ngữ tiếng Việt, 2022  
<https://jubes.ueh.edu.vn/Content/ArticleFiles/9918f8e1-fdaa-44eb-8f55-5fc4c759428d/JABES-2021-2-V27.pdf>

- [7] Phân loại tính chất vụ việc bài báo mạng dựa trên mô hình PhoBERT  
 ThS. Lê Ngọc An, Nguyễn Đình Toàn, Lê Trường Thiên, Dương Trần Đức.(PDF) PHÂN LOẠI TÍNH CHẤT VỤ VIỆC BÀI BÁO MẠNG DỰA TRÊN MÔ HÌNH PhoBERT

- [8] Võ Trung Hùng, Ninh Khánh Chi , Trần Anh Kiệt, PHÁT HIỆN TỰ ĐỘNG TIN GIẢ: THÀNH TỰU VÀ THÁCH THỨC, 2022 <https://jst-ud.vn/jst-ud/article/download/7816/5527/9039>

- [9] Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, Mark Johnson.

*VnCoreNLP: A Vietnamese Natural Language Processing Toolkit*, 2018

<https://aclanthology.org/N18-5012/>

[10] Aysu Ezen-Can. *A Comparison of LSTM and BERT for Small Corpus*, 2020.

[https://www.researchgate.net/publication/344234220\\_A\\_Comparison\\_of\\_LSTM\\_and\\_BERT\\_for\\_Small\\_Corpus](https://www.researchgate.net/publication/344234220_A_Comparison_of_LSTM_and_BERT_for_Small_Corpus)

## PHỤ LỤC

### A. Code giao diện AI & Màn hình (Người dùng - Cuối)

#### *Backend*

Mã nguồn: File: crawl\_service/routes.py

```
@router.post("/manual-classify")
async def manual_classify(input: TextInput):
    logger.info("Bắt đầu xử lý /manual-classify")
    try:
        conn = psycopg2.connect(**CRAWL_DB_CONFIG)
        cursor = conn.cursor()
        # Lưu vào manual_entries
        cursor.execute("""
            CREATE TABLE IF NOT EXISTS manual_entries (
                id SERIAL PRIMARY KEY,
                content TEXT NOT NULL,
                url TEXT,
                created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
            );
        """)
        cursor.execute("""
            INSERT INTO manual_entries (content, url)
            VALUES (%s, %s)
            RETURNING id;
        """, (input.content, input.url))
        manual_entry_id = cursor.fetchone()[0]
        conn.commit()
        # Lưu vào filtered_data
        cursor.execute("""
            CREATE TABLE IF NOT EXISTS filtered_data (
                id SERIAL PRIMARY KEY,
                content TEXT NOT NULL,
                url TEXT,
                is_approved BOOLEAN DEFAULT FALSE,
                created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
            );
        """)
        cursor.execute("""
            INSERT INTO filtered_data (content, url, is_approved)
            VALUES (%s, %s, %s)
            RETURNING id;
        """, (input.content, input.url, True))
        filtered_data_id = cursor.fetchone()[0]
        conn.commit()
    except Exception as e:
        logger.error(f"Error occurred during manual classification: {str(e)}")
        return {"error": str(e)}
    finally:
        cursor.close()
        conn.close()
    return {"id": manual_entry_id, "filtered_id": filtered_data_id}
```

**Commented [10]:** Thêm phụ lục, chuyển code về phần phụ lục

```

    );
""")  

cursor.execute("""  

    INSERT INTO filtered_data (content, url, is_approved)  

    VALUES (%s, %s, %s);  

""", (input.content, input.url, True))  

conn.commit()  

# Gửi yêu cầu phân loại đến Training Service  

response = requests.post(  

    f"{TRAINING_SERVICE_URL}/classify-manual",  

    json={"content": input.content, "url": input.url},  

    timeout=20
)  

response.raise_for_status()  

result = response.json()  

# Lưu kết quả vào classified_results  

cursor.execute("""  

    CREATE TABLE IF NOT EXISTS classified_results (  

        id SERIAL PRIMARY KEY,  

        content TEXT NOT NULL,  

        url TEXT,  

        label TEXT NOT NULL,  

        probability FLOAT NOT NULL,  

        probabilities JSONB,  

        classified_at           TIMESTAMP          DEFAULT  

CURRENT_TIMESTAMP  

    );  

""")  

cursor.execute("""  

    INSERT INTO classified_results (content, url, label, probability,  

probabilities)  

    VALUES (%s, %s, %s, %s, %s);  

""", (input.content, input.url, result["label"], result["probability"],  

json.dumps(result["probabilities"])))

```

```

conn.commit()
cursor.close()
conn.close()
return result

except Exception as e:
    logger.error(f'Lỗi khi xử lý: {str(e)}')
    raise HTTPException(status_code=500, detail=f'Lỗi: {str(e)}')

```

### ***Frontend***

```

///////////
// FakeNewsChecker
///////////

const FakeNewsChecker = () => {
    const [messages, setMessages] = useState([]);
    const [inputText, setInputText] = useState("");

    const handleSendMessage = async () => {
        if (!inputText.trim()) return;
        setMessages((prev) => [...prev, { sender: "user", content: inputText }]);

        try {
            const response = await fetch(`#${API_BASE}/manual-classify`, {
                method: "POST",
                headers: { "Content-Type": "application/json" },
                body: JSON.stringify({
                    content: inputText,
                    url: null
                })
            });
            const data = await response.json();
            setMessages((prev) => [...prev, { sender: "bot", content: data.label }]);
        } catch (error) {

```

```

    setMessages((prev) => [...prev, { sender: "bot", content: "Lỗi khi gửi dữ
liệu." }]);
}

setInputText("");
};

const handleKeyPress = (e) => {
  if (e.key === "Enter") handleSendMessage();
};

return (
  <div className="p-4 bg-gray-100 h-screen w-full pt-16 flex flex-col
overflow-auto">
  <h2 className="text-xl font-bold mb-4">Kiểm Tra & Xử Lý Tin
Giả</h2>
  <div className="flex-1 bg-white p-4 rounded shadow overflow-auto">
    {messages.map((msg, index) => (
      <div
        key={index}
        className={`p-2 my-2 rounded max-w-xs ${msg.sender === "user"
          ? "bg-blue-200 ml-auto text-right"
          : "bg-gray-200 mr-auto text-left"
        }`}
      >
        {msg.content}
      </div>
    )));
  </div>
  <div className="mt-4 flex">
    <input
      className="flex-1 p-2 border rounded"
      type="text"
      placeholder="Nhập tin cần kiểm tra...">
  
```

```

    value={inputText}
    onChange={(e) => setInputText(e.target.value)}
    onKeyDown={handleKeyPress}
  />
  <button
    onClick={handleSendMessage}
    className="ml-2 px-4 py-2 bg-blue-500 text-white rounded"
  >
    Gửi
  </button>
</div>
</div>
);
};

```

## B. Code giao diện Admin panel - thống kê

### *Backend*

Mã nguồn : File: dashboard\_service/routes.py

```

@router.get("/stats")
async def get_stats(start_date: str = None, end_date: str = None):
    try:
        logger.info("Bắt đầu xử lý /stats")
        conn = connect_dashboard_db()
        cursor = conn.cursor()

        # Truy vấn thống kê theo ngày
        query = """
            SELECT DATE(processed_at) as date, COUNT(*) as count
            FROM display_data
            WHERE (%s IS NULL OR processed_at >= %s) AND (%s IS NULL
        OR processed_at <= %s)
            GROUP BY DATE(processed_at)
            ORDER BY date DESC;
        """

```

```

cursor.execute(query, (start_date, start_date, end_date, end_date))
time_stats = [{"date": str(row[0]), "count": row[1]} for row in
cursor.fetchall()]

# Tổng số bài viết
query = """
    SELECT COUNT(*) FROM display_data
    WHERE (%s IS NULL OR processed_at >= %s) AND (%s IS NULL
OR processed_at <= %s);
"""

cursor.execute(query, (start_date, start_date, end_date, end_date))
total_articles = cursor.fetchone()[0]

# Tỷ lệ theo nhãn
query = """
    SELECT label, COUNT(*) as count
    FROM display_data
    WHERE (%s IS NULL OR processed_at >= %s) AND (%s IS NULL
OR processed_at <= %s)
    GROUP BY label;
"""

cursor.execute(query, (start_date, start_date, end_date, end_date))
label_counts = {row[0]: row[1] for row in cursor.fetchall()}
total = sum(label_counts.values())
ratios = {label: (count / total * 100) if total > 0 else 0 for label, count in
label_counts.items()}

cursor.close()
conn.close()
logger.info(f"Trả về thông kê: {len(time_stats)} ngày, tổng
{total_articles} bài")
return {
    "time_stats": time_stats,
    "total_articles": total_articles,
}

```

```

    "ratios": ratios
}
except Exception as e:
    logger.error(f'Lỗi khi lấy thông kê: {str(e)}')
    raise HTTPException(status_code=500, detail=f'Lỗi khi lấy thông kê:
{str(e)}')

```

### ***Frontend***

```

// Phần fetch dữ liệu thông kê
useEffect(() => {
    if (activeTab === "stats") {
        const fetchStats = async () => {
            try {
                const [normalRes, socialRes, politicalRes] = await Promise.all([
                    fetch(`${API_BASE}/normal-news`),
                    fetch(`${API_BASE}/fake-social-news`),
                    fetch(`${API_BASE}/fake-political-news`)
                ]);
                const normalData = await normalRes.json();
                const socialData = await socialRes.json();
                const politicalData = await politicalRes.json();
                const normalCount = Array.isArray(normalData.normal_news)
                    ? normalData.normal_news.length
                    : 0;
                const socialCount = Array.isArray(socialData.fake_social_news)
                    ? socialData.fake_social_news.length
                    : 0;
                const politicalCount = Array.isArray(politicalData.fake_political_news)
                    ? politicalData.fake_political_news.length
                    : 0;
                setData({ normalCount, socialCount, politicalCount });
            } catch (error) {
                logger.error(`Error fetching stats: ${error}`);
            }
        };
        fetchStats();
    }
});

```

```

} catch (error) {
  console.error(error);
  setData({ error: "Lỗi khi lấy dữ liệu thống kê" });
}
};

fetchStats();
}

[, [activeTab]];

// Phần hiển thị biểu đồ thống kê
{activeTab === "stats" && (
<div>
  <h3 className="text-lg font-semibold mb-2">Biểu đồ thống kê (Số tin
mỗi loại)</h3>
  <ResponsiveContainer width="100%" height={300}>
    <BarChart
      data={[
        { category: "Bài thường", count: data?.normalCount || 0 },
        { category: "Bài xã hội", count: data?.socialCount || 0 },
        { category: "Bài chính trị", count: data?.politicalCount || 0 }
      ]}
    >
      <CartesianGrid strokeDasharray="3 3" />
      <XAxis dataKey="category" />
      <YAxis />
      <Tooltip />
      <Legend />
      <Bar dataKey="count" fill="#82ca9d" />
    </BarChart>
  </ResponsiveContainer>

  <h3 className="text-lg font-semibold mt-8 mb-2">Biểu đồ phần
trăm</h3>
  <ResponsiveContainer width="100%" height={300}>

```

```

<PieChart>
  <Pie
    data={[
      { name: "Bài thường", value: data?.normalCount || 0 },
      { name: "Bài xã hội", value: data?.socialCount || 0 },
      { name: "Bài chính trị", value: data?.politicalCount || 0 }
    ]}
    cx="50%"
    cy="50%"
    label
    outerRadius={80}
    dataKey="value"
  >
  {[{
    { name: "Bài thường", value: data?.normalCount || 0 },
    { name: "Bài xã hội", value: data?.socialCount || 0 },
    { name: "Bài chính trị", value: data?.politicalCount || 0 }
  ].map((entry, index) => (
    <Cell key={`cell-${index}`} fill={COLORS[index % COLOR
    COLORS.length]} />
  ))}
  </Pie>
  <Tooltip />
  <Legend />
</PieChart>
</ResponsiveContainer>
</div>
)

```

### C. Code giao diện Admin panel - danh sách các bài viết

#### *Backend*

File: dashboard\_service/routes.py

```

@router.get("/normal-news")
async def get_normal_news(limit: int = 100):
    try:
        conn = connect_dashboard_db()
        cursor = conn.cursor()
        cursor.execute("""
            SELECT content, url, probability, processed_at
            FROM normal_news
            ORDER BY processed_at DESC
            LIMIT %s;
        """, (limit,))
        rows = cursor.fetchall()
        logger.info(f"Trả về {len(rows)} bài viết từ normal_news")
        cursor.close()
        conn.close()
        return {
            "normal_news": [
                {"content": row[0], "url": row[1], "probability": row[2],
                 "processed_at": row[3]}
                for row in rows
            ]
        }
    except Exception as e:
        logger.error(f'Lỗi khi lấy normal_news: {str(e)}')
        raise HTTPException(status_code=500, detail=f'Lỗi khi lấy
normal_news: {str(e)}')

@router.get("/fake-social-news")
async def get_fake_social_news(limit: int = 100):
    try:
        conn = connect_dashboard_db()
        cursor = conn.cursor()
        cursor.execute("""
            SELECT content, url, probability, processed_at
        """)

```

```

FROM fake_social_news
ORDER BY processed_at DESC
LIMIT %s;
"""", (limit,))
rows = cursor.fetchall()
logger.info(f"Trả về {len(rows)} bài viết từ fake_social_news")
cursor.close()
conn.close()
return {
    "fake_social_news": [
        {"content": row[0], "url": row[1], "probability": row[2],
         "processed_at": row[3]}
        for row in rows
    ]
}
except Exception as e:
    logger.error(f'Lỗi khi lấy fake_social_news: {str(e)}')
    raise HTTPException(status_code=500, detail=f'Lỗi khi lấy
fake_social_news: {str(e)}')

@router.get("/fake-political-news")
async def get_fake_political_news(limit: int = 100):
    try:
        conn = connect_dashboard_db()
        cursor = conn.cursor()
        cursor.execute("""
            SELECT content, url, probability, processed_at
            FROM fake_political_news
            ORDER BY processed_at DESC
            LIMIT %s;
        """", (limit,))
        rows = cursor.fetchall()
        logger.info(f"Trả về {len(rows)} bài viết từ fake_political_news")
        cursor.close()
    
```

```

conn.close()
return {
    "fake_political_news": [
        {"content": row[0], "url": row[1], "probability": row[2],
    "processed_at": row[3]}
        for row in rows
    ]
}
except Exception as e:
    logger.error(f'Lỗi khi lấy fake_political_news: {str(e)}')
    raise HTTPException(status_code=500, detail=f'Lỗi khi lấy
fake_political_news: {str(e)}')

```

### **Frontend**

```

/ Phần fetch dữ liệu cho các loại bài viết
useEffect(() => {
    if (activeTab === "filtered") {
        fetchData(`${API_BASE}/display-data`);
    } else if (activeTab === "normal") {
        fetchData(`${API_BASE}/normal-news`);
    } else if (activeTab === "social") {
        fetchData(`${API_BASE}/fake-social-news`);
    } else if (activeTab === "political") {
        fetchData(`${API_BASE}/fake-political-news`);
    } else if (activeTab === "crawled") {
        fetchData(`${API_BASE}/raw-fanpage-facebook`);
    }
}, [activeTab]);

// Hàm render bảng dữ liệu
const renderTable = (dataArray) => {
    if (!Array.isArray(dataArray) || dataArray.length === 0) return
<div>Không có dữ liệu bảng</div>;

```

```

return (
  <table className="w-full border-collapse border border-gray-300">
    <thead>
      <tr className="bg-gray-200">
        <th className="border p-2">Nội dung</th>
        <th className="border p-2">URL</th>
        <th className="border p-2">Loại</th>
        <th className="border p-2">Xác suất</th>
      </tr>
    </thead>
    <tbody>
      {dataArray.map((item, index) => (
        <tr key={index} className="hover:bg-gray-100">
          <td className="border p-2">{item.content}</td>
          <td className="border p-2">{item.url}</td>
          <td className="border p-2">{item.label}</td>
          <td className="border p-2">{item.probability}</td>
        </tr>
      )))
    </tbody>
  </table>
);
};

// Phần hiển thị các tab
<div className="mb-4 flex gap-2 flex-wrap">
  {[

    { tab: "stats", label: "Thông kê" },
    { tab: "filtered", label: "Bài đã lọc" },
    { tab: "normal", label: "Bài thường" },
    { tab: "social", label: "Bài đòi sống xã hội" },
    { tab: "political", label: "Bài chính trị" },
    { tab: "crawled", label: "Bài đã càو" }
  ]
}

```

```

].map(({ tab, label }) => (
  <button
    key={tab}
    className={`px-3 py-1 border rounded cursor-pointer ${activeTab ===
tab ? "bg-blue-500 text-white" : "bg-white text-black"
  }`}
    onClick={() => setActiveTab(tab)}
  >
    {label}
  </button>
))
</div>

// Phần hiển thị nội dung các tab bài viết
{activeTab !== "stats" && (
  <div className="bg-white p-4 rounded shadow">
    {data && !data.error
      ? renderTable(getTableData())
      : data?.error || "Chưa có dữ liệu"
    }
  </div>
)
}

```

#### D. Code giao diện Admin panel - crawl và lọc

##### *Backend*

Crawl Service: Endpoint /crawl

File: crawl\_service/routes.py

```

@router.post("/crawl", response_model=CrawlResponse)
async def crawl_endpoint(request: CrawlRequest):
  logger.info(f"Bắt đầu xử lý /crawl với url={request.url},
target_posts={request.target_posts}")
  try:
    create_tables()

```

```

posts = crawl_facebook_page(request.url, request.target_posts)
save_raw_to_db(posts)
save_clean_to_db(posts, request.url)
save_filtered_data(posts, request.url)
return {
    "url": request.url,
    "posts": posts,
    "message": f'Dã crawl, lưu {len(posts)} bài viết vào raw_fanpage_facebook, clean_fanpage_facebook và filtered_data!'
}
except Exception as e:
    logger.error(f'Lỗi khi crawl: {str(e)}')
    raise HTTPException(status_code=500, detail=f'Lỗi khi crawl: {str(e)}')

```

Hàm thu thập và lọc dữ liệu

File: crawl\_service/crawler.py

```

def crawl_facebook_page(url: str, target_posts: int) -> list:
    driver = init_driver()
    driver.get(url)
    scroll_page(driver)
    posts = []
    post_elements = driver.find_elements(By.XPATH,
        '//div[@class="userContentWrapper"]')
    for element in post_elements[:target_posts]:
        content = element.text.strip()
        if content:
            posts.append({"content": content})
    driver.quit()
    return posts

def preprocess_for_phobert(text: str) -> str:
    text = re.sub(r'^\w\s]', " ", text) # Loại ký tự đặc biệt
    text = re.sub(r'\s+', ' ', text).strip() # Chuẩn hóa khoảng trắng

```

```
return text
```

File: crawl\_service/database.py

```
def save_raw_to_db(posts: list):
    conn = psycopg2.connect(**CRAWL_DB_CONFIG)
    cursor = conn.cursor()
    cursor.execute("""
        CREATE TABLE IF NOT EXISTS raw_fanpage_facebook (
            id SERIAL PRIMARY KEY,
            content TEXT NOT NULL,
            created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
        );
    """)
    cursor.executemany("""
        INSERT INTO raw_fanpage_facebook (content)
        VALUES (%s);
    """, [(post["content"],) for post in posts])
    conn.commit()
    cursor.close()
    conn.close()

def save_clean_to_db(posts: list, url: str):
    conn = psycopg2.connect(**CRAWL_DB_CONFIG)
    cursor = conn.cursor()
    cursor.execute("""
        CREATE TABLE IF NOT EXISTS clean_fanpage_facebook (
            id SERIAL PRIMARY KEY,
            content TEXT NOT NULL,
            url TEXT,
            process BOOLEAN DEFAULT FALSE,
            created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
        );
    """)
```

```

cleaned_posts = [(preprocess_for_phobert(post["content"]), url) for post in
posts]
cursor.executemany("""
    INSERT INTO clean_fanpage_facebook (content, url)
    VALUES (%s, %s);
""", cleaned_posts)
conn.commit()
cursor.close()
conn.close()

def save_filtered_data(posts: list, url: str):
    conn = psycopg2.connect(**CRAWL_DB_CONFIG)
    cursor = conn.cursor()
    cursor.execute("""
        CREATE TABLE IF NOT EXISTS filtered_data (
            id SERIAL PRIMARY KEY,
            content TEXT NOT NULL,
            url TEXT,
            is_approved BOOLEAN DEFAULT FALSE,
            created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
        );
    """)
    cleaned_posts = [(preprocess_for_phobert(post["content"]), url, False) for
post in posts]
    cursor.executemany("""
        INSERT INTO filtered_data (content, url, is_approved)
        VALUES (%s, %s, %s);
    """, cleaned_posts)
    conn.commit()
    cursor.close()
    conn.close()

```

### ***Frontend***

```
//////////
```

```
// CrawlFilter
///////////
const CrawlFilter = () => {
  const [inputUrl, setInputUrl] = useState("");
  const [crawledData, setCrawledData] = useState([]);
  const navigate = useNavigate();

  const handleCrawl = async () => {
    if (!inputUrl.trim()) return;
    try {
      const response = await fetch("http://127.0.0.1:8001/crawl", {
        method: "POST",
        headers: { "Content-Type": "application/json" },
        body: JSON.stringify({ url: inputUrl })
      });
      const data = await response.json();
      setCrawledData(data.posts || []);
    } catch (error) {
      console.error(error);
    }
  };

  const handleFilter = async () => {
    try {
      await fetch("http://127.0.0.1:8002/process", { method: "POST" });
      navigate("/admin", { state: { tab: "filtered" } });
    } catch (error) {
      console.error("Lỗi khi xử lý lọc:", error);
    }
  };
}

const renderTable = (dataArray) => {
  if (!Array.isArray(dataArray) || dataArray.length === 0) return
<div>Không có dữ liệu bảng</div>;
```

```

return (
  <table className="w-full border-collapse border border-gray-300">
    <thead>
      <tr className="bg-gray-200">
        <th className="border p-2">Nội dung</th>
        <th className="border p-2">URL</th>
        <th className="border p-2">Loại</th>
        <th className="border p-2">Xác suất</th>
      </tr>
    </thead>
    <tbody>
      {dataArray.map((item, index) => (
        <tr key={index} className="hover:bg-gray-100">
          <td className="border p-2">{item.content}</td>
          <td className="border p-2">{item.url}</td>
          <td className="border p-2">{item.label}</td>
          <td className="border p-2">{item.probability}</td>
        </tr>
      )))
    </tbody>
  </table>
);
};

return (
  <div className="p-4 bg-gray-100 h-screen w-full pt-16 flex flex-col overflow-auto">
    <h2 className="text-xl font-bold mb-4">Crawl & Lọc Fanpage</h2>
    <div className="mt-4 flex">
      <input
        className="flex-1 p-2 border rounded"
        type="text"
        placeholder="Nhập link cần crawl..."
        value={inputUrl}
      >
    </div>
  </div>
);

```

```
        onChange={(e) => setInputUrl(e.target.value)}
    />
    <button
        onClick={handleCrawl}
        className="ml-2 px-4 py-2 bg-green-500 text-white rounded"
    >
        Crawl
    </button>
</div>
{crawledData.length > 0 && (
    <div className="mt-4">
        <h3 className="text-lg font-semibold mb-2">Danh sách bài
        crawl</h3>
        {renderTable(
            crawledData.map((post) => ({
                content: post,
                url: "",
                label: "",
                probability: ""
            }))
        )}
    <button
        onClick={handleFilter}
        className="mt-2 px-4 py-2 bg-blue-500 text-white rounded"
    >
        Lọc
    </button>
</div>
)}
</div>
);
};
```

### Mẫu số 20.1g: Biên bản họp Hội đồng đánh giá khóa luận tốt nghiệp

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 10 tháng 05 năm 2025

**BIÊN BẢN HỌP HỘI ĐỒNG  
ĐÁNH GIÁ KHÓA LUẬN TỐT NGHIỆP (HĐ: 1)**

Thời gian: ..... 8<sup>h</sup> 20' .....

Địa điểm: P. 11, Khu giảng đường Khoa Công nghệ thông tin

Thành viên Hội đồng (Tổng số: ....03.., có mặt: 03....., vắng mặt: .....00/....):

1. Chủ tịch: GS. TS. Thái Thành Sơn
  2. Uỷ viên: TS. Vũ Xuân Hạnh
  3. UV thư ký: ThS. Mai Thị Thúy Hà

Họ và tên sinh viên: Trần Trọng Trường

Mã sinh viên: 20A10010189

Lop: 2010A05

## Tên đề tài: Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy

Ho và tên người hướng dẫn: ThS. Lê Ngọc An

## NỘI DUNG NHẬN XÉT, ĐÁNH GIÁ

### I. Nhận xét, đánh giá:

## II. Kiến nghị:

### III. Kết luận của Hội đồng

- #### 1. Điểm kết luận của Hội đồng:

Bằng số: ..... 6,5

Bằng chữ: ...Sau ngày... năm.....

2. Kết quả điểm đã công bố công khai cho sinh viên.
  3. Các hồ sơ kèm theo gồm: Khóa luận tốt nghiệp (bản dùng cho thư ký); bản nhận xét của người hướng dẫn; bản nhận xét của phản biện; phiếu chấm điểm của các thành viên Hội đồng; phiếu tổng hợp điểm.
- Buổi chấm khóa luận tốt nghiệp của sinh viên Trần Trọng Trường kết thúc hồi 24/8 cùng ngày, biên bản đã được đọc lại cho mọi người có mặt cùng nghe và không ai có ý kiến khác.

THƯ KÝ  
(Ký ghi rõ họ tên)

ThS. Mai Thị Thúy Hà

CHỦ TỊCH  
(Ký ghi rõ họ tên)

GS. TS. Thái Thanh Sơn

### Mẫu số 20.1g: Biên bản họp Hội đồng đánh giá khóa luận tốt nghiệp

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 40 tháng 05 năm 2025

## BIÊN BẢN HỌP HỘI ĐỒNG

Thời gian: ..... 8<sup>h</sup> 20.....

Địa điểm: P. 21, Khu giảng đường Khoa Công nghệ thông tin

Thành viên Hội đồng (Tổng số: ..05..., có mặt: ..05..., vắng mặt: ..0.....):

1. Chủ tịch: GS. TS. Thái Thanh Sơn
  2. Uỷ viên: TS. Vũ Xuân Hạnh
  3. UV thư ký: ThS. Mai Thị Thúy Hà

Họ và tên sinh viên: Nguyễn Phương Nam

Mã sinh viên: 20A10010115

Llop: 2010A05

Tên đề tài: Nghiên cứu phương pháp tư đồng phân loại bài viết mang xã hội bằng

hoc máy

Ho và tên người hướng dẫn: ThS Lê Ngọc An

## NỘI DUNG NHÂN XÉT, ĐÁNH GIÁ

### I. Nhận xét, đánh giá:

1. Nghe xem và làm già:  
Tiêu chí: là ai biết viết tin tức là gì? Câu hỏi: Võ  
tôn đế biết bao nhiêu viết tin tức? Đáp: Võ Tông là một  
lão thành, là một trong những nhà soạn thảo văn kiện nổi  
tiếng, là một chuyên viên luyện chữ ngày.

## II. Kiến nghị:

bíp cáo chia tay nhau. Thiên phu là' Cuman. Nết  
tết mèo lèn còn k' tài b' h'ng cái y'c' t'c' d'c' m'c' a'c'  
t'ng' b'c' t'c' t'c' th'm ch'p d'c' m'c' h'ng'c'.  
T'nh' bay' ch'p t'c' cái v'nh' t'c' Hu'e' h'iu' d'c' t'c'  
h'nh' h'iu' m'c' t'c' h'iu' k'c' s'c' y'c' h'iu' 'u'c' c'k'c' quan  
III. Kết luận của Hội đồng Chay'c' code v'c' ch'p l'c' d'c' d'c' t'

### III. Kết luận của Hội đồng Chuyển code về-

- #### 1. Điểm kết luận của Hội đồng:

Bằng số: ..... 60

Bảng chữ: *Sáu điểm*

2. Kết quả điểm đã công bố công khai cho sinh viên.
3. Các hồ sơ kèm theo gồm: Khóa luận tốt nghiệp (bản dùng cho thư ký); bản nhận xét của người hướng dẫn; bản nhận xét của phản biện; phiếu chấm điểm của các thành viên Hội đồng; phiếu tổng hợp điểm.

Buổi chấm khóa luận tốt nghiệp của sinh viên Nguyễn Phương Nam kết thúc hồi *8/4/2015* cùng ngày, biên bản đã được đọc lại cho mọi người có mặt cùng nghe và không ai có ý kiến khác.

THƯ KÝ  
(Ký ghi rõ họ tên)

ThS. Mai Thị Thúy Hà

CHỦ TỊCH  
(Ký ghi rõ họ tên)

GS. TS. Thái Thanh Sơn

### Mẫu số 20.1g: Biên bản họp Hội đồng đánh giá khóa luận tốt nghiệp

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 10. tháng 05 năm 2025

**BIÊN BẢN HỌP HỘI ĐỒNG**

Thời gian: ..... 8h20

Địa điểm: P. 2A, Khu giảng đường Khoa Công nghệ thông tin

Thành viên Hội đồng (Tổng số: ..0.3..., có mặt: ..0.5..., vắng mặt: ..0.D.....):

1. Chủ tịch: GS. TS. Thái Thành Sơn
  2. Uỷ viên: TS. Vũ Xuân Hạnh
  3. UV thư ký: ThS. Mai Thị Thúy Hà

Họ và tên sinh viên: Bùi Việt Anh

Mã sinh viên: 19A10010261

Lop: 1910A02

Tên đề tài: Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy

Ho và tên người hướng dẫn: Th.S Lê Ngọc An

## NỘI DUNG NHÂN XÉT, ĐÁNH GIÁ

### I. Nhận xét, đánh giá:

Tiêu chí: là văn bản biêt miết tin giả là gì? Cần  
để xác định là: là cái gì? là giả? là thật?  
để xác định là: là có đc để là bao nhiêu?  
để xác định là: là có đc (ay và) đc, điều số 2: là bao nhiêu? là bao nhiêu?  
để xác định là: là có đc, điều số 3: là bao nhiêu?

## II. Kiến nghị:

Bài cáo chua徇天理... Thực phản bối lican.  
 Một số vòi徇天理... Bối bối cát yết trai  
 徇天理?徇天理... Khoa tin, bối chon... là mìn... huy  
 Trong bối chia bối các vòi đã thừa hiếu dùn... và sòng hiếu  
 bối vòi bối vòi... cát... giao... cát... tên quan... chuyển... vòi ve  
 III. Kết luận của Hội đồng phu lær, Sel day lai guyen bok cao

### III. Kết luận của Hội đồng

1. Điểm kết luận của Hội đồng:  
Bằng số: 6,8

Bằng chữ: ...Lê...diệp.....

2. Kết quả điểm đã công bố công khai cho sinh viên.
3. Các hồ sơ kèm theo gồm: Khóa luận tốt nghiệp (bản dùng cho thư ký); bản nhận xét của người hướng dẫn; bản nhận xét của phản biện; phiếu chấm điểm của các thành viên Hội đồng; phiếu tổng hợp điểm.  
Buổi chấm khóa luận tốt nghiệp của sinh viên Bùi Việt Anh kết thúc hồi ..14.12.  
cùng ngày, biên bản đã được đọc lại cho mọi người có mặt cùng nghe và không ai có ý kiến khác.

THƯ KÝ  
(Ký ghi rõ họ tên)

ThS. Mai Thị Thúy Hà

CHỦ TỊCH  
(Ký ghi rõ họ tên)

GS. TS. Thái Thanh Sơn

## NHẬN XÉT HƯỚNG DẪN ĐỒ ÁN/KHOÁ LUẬN TỐT NGHIỆP ĐẠI HỌC

Tên đề tài: Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy

- |                                 |                    |               |
|---------------------------------|--------------------|---------------|
| 1. Sinh viên: Nguyễn Phương Nam | Mã SV: 20A10010115 | Lớp: 2010A005 |
| 2. Sinh viên: Bùi Việt Anh      | Mã SV: 19A10010261 | Lớp: 1910A02  |
| 3. Sinh viên: Trần Trọng Trường | Mã SV: 20A10010189 | Lớp: 2010A05  |

### NỘI DUNG NHẬN XÉT

#### 1. Nhận xét ý nghĩa thực tiễn của đề tài

Đề tài “Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy” tập trung nghiên cứu các phương pháp và kỹ thuật nhằm nâng cao hiệu quả của hệ thống phát hiện tin giả góp phần vào việc cải thiện an toàn thông tin trên môi trường mạng trong tương lai.

#### 2. Đánh giá Kết quả thực hiện đề tài

##### 2.1. Mức độ hoàn thiện so với mục tiêu đặt ra

- Đề án nghiên cứu giải pháp phát hiện tin giả dựa trên PhoBERT. Từ đó đưa ra được các đặc điểm, ưu, nhược điểm của mô hình hiện tại.
- Cài đặt và triển khai hệ thống PhoBERT và chạy kết quả trên tập dữ liệu tự tạo ra dưới các tiêu chí đặc trưng.
- Đã dựng được hệ thống đánh giá tin giả
- Cần mở rộng thêm dữ liệu để có phương án đánh giá tốt hơn.

##### 2.2. Những nội dung đã đạt được

- Khảo sát được nhu cầu thực tiễn đang cần về phát hiện tin giả trên môi trường mạng.
- Tìm hiểu được cơ chế hoạt động của PhoBERT và các đặc điểm của PhoBERT.
- Đưa ra được phương án để đánh giá mô hình.

### 3. Kết luận

3.1. Cho phép sinh viên bảo vệ đề tài trước Hội đồng bảo vệ đồ án tốt nghiệp khoa Công nghệ Thông tin – Trường ĐH Mở HN hay không?

ĐÓNG Ý       KHÔNG ĐÓNG Ý

### 3.2. Điểm đánh giá cho từng sinh viên

| STT  | Sinh viên         | Điểm bằng số | Điểm viết bằng chữ |
|--|-------------------|--------------|--------------------|
| A11 - Tuân thủ kế hoạch, làm việc trách nhiệm (5%) |                   |              |                    |
| 1  | Nguyễn Phương Nam | 9,0          | Chín phẩy không    |
| 2  | Bùi Việt Anh      | 9,0          | Chín phẩy không    |
| 3  | Trần Trọng Trường | 9,0          | Chín phẩy không    |
| A22 - Kết quả báo cáo tổng kết (55%)               |                   |              |                    |
| 1  | Nguyễn Phương Nam | 9,0          | Chín phẩy không    |
| 2  | Bùi Việt Anh      | 9,0          | Chín phẩy không    |
| 3  | Trần Trọng Trường | 9,0          | Chín phẩy không    |

### 4. Thông tin người hướng dẫn

Họ và tên: Lê Ngọc An

Học vị: Thạc sĩ

Học hàm/Chức danh khoa học: .....

Cơ quan công tác: Khoa CNTT, Trường ĐH Mở Hà Nội

Hà Nội, ngày 06/05/2025

Người hướng dẫn



Lê Ngọc An

1.1

Mẫu số 20.1a: Phiếu nhận xét đỗ án tốt nghiệp

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 07 tháng 05 năm 2025

**PHIẾU NHẬN XÉT ĐỖ ÁN TỐT NGHIỆP**

(Dùng cho người phản biện ĐATN)

**I. Thông tin chung**

- Họ và tên người phản biện: Vũ Xuân Hạnh
- Học hàm, học vị: Tiến sĩ
- Đơn vị công tác: Khoa Công nghệ Thông tin – trường Đại học Mở Hà Nội
- Nhóm sinh viên thực hiện:

| Họ tên            | Ngày sinh  | Mã SV       | Lớp     |
|-------------------|------------|-------------|---------|
| Nguyễn Phương Nam | 13/08/2002 | 20A10010115 | 2010A05 |
| Trần Trọng Trường | 08/11/2002 | 20A10010189 | 2010A05 |
| Bùi Việt Anh      | 01/05/2001 | 19A10010261 | 1910A02 |

- Tên đề tài: NC phương pháp tự động phân loại bài viết mạng xã hội bằng học máy.
- Họ và tên người hướng dẫn: Ths. Lê Ngọc An

**II. Nhận xét về khóa luận**

2.1. Nhận xét về hình thức (bố cục, định dạng, hành văn):

- Định dạng quyền bao cáo toàn văn chưa đúng định dạng, tài liệu tham khảo sai quy định, đưa code về phần phụ lục (nếu cần), thiếu phần đánh giá, kết luận.
- Một số tiêu mục trình bày sơ sài

2.2. Mục tiêu và nội dung:

- Xây dựng một hệ thống có khả năng tự động phân loại bài viết trên mạng xã hội theo chủ đề dựa trên các phương pháp học máy.
- Cung cấp một mô hình học máy có thể áp dụng trên các nền tảng mạng xã hội để nhận diện nội dung, phân loại bài viết theo các nhóm chủ đề như chính trị, xã hội, v..v.
- Triển khai mô hình thực nghiệm, đánh giá độ chính xác và hiệu suất của hệ thống trên tập dữ liệu thực tế, đồng thời đề xuất hướng cải thiện để nâng cao chất lượng phân loại bài viết.

2.3 Kết quả đạt được:

- Nghiên cứu tổng quan về PhoBERT
- Thu thập và tiền xử lý dữ liệu
- Xây dựng và phát triển trên PhoBERT
- Triển khai ứng dụng

#### 2.4. Kết luận và đề nghị:

Nội dung và hàm lượng kiến thức phù hợp với yêu cầu của một đồ án tốt nghiệp Đại học chuyên ngành Công nghệ phần mềm.

Tuy nhiên, đồ án cần nghiên cứu, trích chọn các đặc trưng của từng loại tin để phân loại cho hiệu quả cao hơn. Nhóm sinh viên có đưa ra một số từ khóa nhưng trong file huấn luyện không được thể hiện,. Một số mục trong quyền báo cáo trình bày khá sơ sài.

#### 2.5. Các vấn đề cần bổ sung, chỉnh sửa:

- Các yếu tố (đặc trưng) của từng loại tin, tích chọn đưa vào huấn luyện.
- Trình bày chi tiết các vấn đề thực hiện được, so sánh hiệu suất mô hình với các nghiên cứu có liên quan
- Chuyển code về phần phụ lục, định dạng lại quyền báo cáo theo đúng quy định

#### III. Câu hỏi sinh viên phải trả lời trước Hội đồng (ít nhất 02 câu hỏi)

- Câu hỏi 1: Độ dài tin (chuỗi đầu vào) tối đa mô hình có thể phát hiện được?
- Câu hỏi 2: Các đặc trưng (keyword) được sử dụng như thế nào?
- Câu hỏi 3: Có 3 loại tin, nhãn của tập huấn luyện định dạng thế nào?

#### IV. Điểm phân biện

| STT | Họ và tên         | Điểm bằng số | Điểm bằng chữ  | Ghi chú |
|-----|-------------------|--------------|----------------|---------|
| 1   | Nguyễn Phượng Nam | 7.0          | Bảy phẩy không |         |
| 2   | Trần Trọng Trường | 7.0          | Bảy phẩy không |         |
| 3   | Bùi Việt Anh      | 7.0          | Bảy phẩy không |         |

#### V. Đánh giá chung

Đề nghị  Được bảo vệ:

Không được bảo vệ:

Cán bộ phản biện

Ký tên (ghi rõ họ tên)

Vũ Xuân Hạnh

TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI  
KHOA CÔNG NGHỆ THÔNG TIN

**GIẤY XÁC NHẬN CHỈNH SỬA ĐỒ ÁN TỐT NGHIỆP**

(Theo yêu cầu của Hội đồng chấm bảo vệ đồ án)

(Mẫu cập nhật 11/2016)

**1. Thông tin chung**

1.1. Tên đề tài: Nghiên cứu phương pháp tự động phân loại bài viết mạng xã hội bằng học máy

1.2. Sinh viên thực hiện: (Họ và tên, lớp)

Bùi Việt Anh – 1910A02

Nguyễn Phương Nam – 2010A05

Trần Trọng Trường – 2010A05

1.3. Giảng viên hướng dẫn: ThS. Lê Ngọc An

**2. Nội dung đã chỉnh sửa theo yêu cầu trong Biên bản bảo vệ đồ án**

(phản này sinh viên khai đầy đủ)

**2.1 Về giải pháp thực hiện đề tài**

**2.2. Về quyền báo cáo đồ án**

| Số trang | Nội dung đã chỉnh sửa  |
|----------|--|
| 78       | Sửa lại tham khảo cho đúng định dạng                               |
|          | Thêm phụ lục   |
| 69       | Thêm bảng so sánh hiệu suất với các đề tài liên quan đã nghiên cứu |
| 50       | Chỉnh sửa lại sơ đồ hệ thống                                       |
| 33 - 48  | Bổ sung thêm các yếu tố đặc trưng cho từng nhóm tin                |

**3. Xác nhận của Giảng viên hướng dẫn:**

Sinh (chữ ký) Chánh Sửa: Nguyễn Cao Cua, Lời đính: *Mr. Lê Ngọc Anh*

Hà Nội, ngày 15 tháng 5 năm 2025

TM Hội đồng chấm bảo vệ

(Thư ký hội đồng xác nhận  
sau khi đối chiếu biên bản,  
kiểm tra nội dung chỉnh sửa)

*MR. LE NGOC ANH*  
Mai Thị Trung Hân

Sinh viên / nhóm sinh viên

*Q*  
Nguyễn Phương Nam  
Anh  
Bùi Việt Anh  
Trung  
Trần Trọng Trường  
1/1

