

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



BÁO CÁO ĐỀ TÀI
WEBSITE PHÂN TÍCH CẢM XÚC PHẢN HỒI
SINH VIÊN
MÔN: LẬP TRÌNH PYTHON

Giảng viên: Nguyễn Trọng Khánh

Lớp CT01 – Nhóm 5

Mã sinh viên : Họ và tên
B23DCCN879 : Phạm Văn Tư
B23DCCN165 : Bùi Anh Đức
B23DCCN264 : Nguyễn Thị Ngọc Hà
B23DCCN600 : Nguyễn Duy Nghĩa

Hà Nội, 2025

Mục lục

PHẦN I: MÔ HÌNH PHÂN TÍCH CẢM XÚC THEO KHÍA CẠNH	3
1 Giới thiệu đề tài	3
2 Mục tiêu và Phạm vi	3
2.1 Mục tiêu	3
2.2 Phạm vi	4
3 Cơ sở lý thuyết và công nghệ sử dụng	4
3.1 Phân loại cảm xúc	4
3.2 Phân tích cảm xúc theo khía cạnh (ABSA)	4
3.3 PhoBERT	5
3.4 Công nghệ xây dựng ứng dụng	5
4 Dữ liệu (Dataset)	5
4.1 Nguồn dữ liệu	5
4.2 Quy mô và nhãn	5
4.3 Phân phối topic × sentiment	6
4.4 Chiến lược chia tập và tiền xử lý	6
4.5 Ví dụ minh họa	7
5 Phương pháp và Mô hình	7
5.1 Kiến trúc mô hình	7
5.2 Cấu hình huấn luyện	7
5.3 Tác dụng của Pair – ABSA trong hệ thống	8
6 Thực nghiệm	9
7 Kết quả	9
8 Thảo luận và Nhận xét	10
8.1 Điểm mạnh	10
8.2 Hạn chế	10
8.3 Lỗi thường gặp	10
9 Kết luận	10
PHẦN II: ỨNG DỤNG WEBSITE PHÂN TÍCH PHẢN HỒI SINH VIÊN	11
1 Tổng quan hệ thống	11
1.1 Mục tiêu sử dụng	11
1.2 Đối tượng sử dụng	11
2 Kiến trúc ứng dụng	11
2.1 Kiến trúc tổng thể	11
2.2 Các thành phần chính	12
2.3 Luồng hoạt động cơ bản	13

3	Thiết kế cơ sở dữ liệu	13
3.1	Mô hình dữ liệu tổng quan	13
3.2	Bảng users	13
3.3	Bảng feedbacks	13
4	Cài đặt giao diện và các chức năng	14
4.1	Chức năng đăng ký, đăng nhập	14
4.2	Phân tích feedback đơn lẻ	15
4.3	Phân tích file CSV	15
4.4	Xem lịch sử feedback	16
5	Dánh giá sản phẩm	17
5.1	Ưu điểm	17
5.2	Hạn chế	17

PHẦN I: MÔ HÌNH PHÂN TÍCH CẢM XÚC THEO KHÍA CẠNH

1 Giới thiệu đề tài

Trong bối cảnh chuyển đổi số giáo dục, các trường đại học tại Việt Nam thu thập hàng nghìn ý kiến phản hồi của sinh viên mỗi học kỳ nhằm hoàn thiện dịch vụ đào tạo, cải thiện trải nghiệm học tập và nâng cao chất lượng giảng dạy. Khối lượng thông tin lớn, đa dạng cách diễn đạt và giàu sắc thái cảm xúc khiến việc đọc thủ công vừa tốn thời gian vừa thiếu tính nhất quán. Do đó, nhóm thực hiện đề tài “Phân tích cảm xúc theo khía cạnh (Aspect-Based Sentiment Analysis – ABSA) cho phản hồi sinh viên tiếng Việt” với mục tiêu xây dựng một hệ thống tự động đánh giá cảm xúc gắn với từng nhóm vấn đề được sinh viên đề cập.

ABSA vượt ra ngoài nhiệm vụ phân loại cảm xúc thuần túy. Thay vì chỉ xác định văn bản mang thái độ tích cực hay tiêu cực, mô hình phải trả lời hai câu hỏi liên tiếp:

1. Văn bản đang đề cập tới khía cạnh nào (giảng viên, chương trình đào tạo, cơ sở vật chất hay nhóm vấn đề khác)
2. Thái độ của người viết đối với khía cạnh đó là tiêu cực, trung lập hay tích cực.

Nhằm thực hiện đồng thời hai yêu cầu này, nhóm lựa chọn tiếp cận “pair classification” – mỗi lần dự đoán mô hình nhận vào một cặp gồm câu phản hồi và khía cạnh truy vấn, sau đó trả lời xem câu có đề cập khía cạnh đó hay không; nếu có thì cảm xúc tương ứng thuộc lớp nào.

Với hệ thống như vậy, nhà trường có thể nhanh chóng tổng hợp điểm mạnh – điểm yếu theo từng khía cạnh, giám sát các vấn đề nổi bật (ví dụ thủ tục hành chính chậm, phòng học nóng, lịch học dày đặc) và tạo nền tảng cho dashboard trực quan phục vụ quyết định quản trị.

2 Mục tiêu và Phạm vi

2.1 Mục tiêu

1. Xây dựng pipeline huấn luyện mô hình ABSA tiếng Việt dựa trên PhoBERT và thiết kế Pair-ABSA

2. Tối ưu đồng thời hai chỉ số đánh giá

- **Acc@4:** tỷ lệ dự đoán chính xác trên toàn bộ bốn lớp (none, negative, neutral, positive). Chỉ số này phản ánh khả năng của mô hình khi xử lý cả trường hợp câu không nhắm tới khía cạnh đang xét.
- **Macro-F1@pos3:** trung bình F1 (macro average) trên ba lớp cảm xúc thực sự (negative, neutral, positive) sau khi loại bỏ lớp “none”. Giá trị cao cho thấy mô hình phân biệt tốt giữa các sắc thái cảm xúc khi câu thực sự đề cập tới khía cạnh.

2.2 Phạm vi

- **Mô hình:** Fine-tuning PhoBERT-base với head phân loại nhiều tầng.
- **Dữ liệu:** kết hợp bộ UIT_VSFC và dữ liệu synthetic do nhóm tự xây dựng để mở rộng ngữ cảnh và cân bằng nhãn.
- **Chia tập:** Train 70%, Validation 30% theo chiến lược phân tầng (stratified) dựa trên tổ hợp khía cạnh × cảm xúc.

3 Cơ sở lý thuyết và công nghệ sử dụng

3.1 Phân loại cảm xúc

Phân loại cảm xúc (sentiment classification) là bài toán gán nhãn thái độ cho văn bản. Phương pháp cổ điển thường sử dụng biểu diễn TF-IDF hoặc n-gram kết hợp với các mô hình tuyến tính như SVM, Logistic Regression. Sự xuất hiện của các mô hình ngôn ngữ tiền huấn luyện (Pre-trained Language Models – PLM) điển hình là BERT đã thay đổi cục diện: nhờ học biểu diễn theo ngữ cảnh sâu, PLM cho phép mô hình phân biệt tinh tế giữa các sắc thái ngôn ngữ, đặc biệt hữu ích với tiếng Việt vốn giàu cấu trúc khẩu ngữ.

3.2 Phân tích cảm xúc theo khía cạnh (ABSA)

ABSA mở rộng bài toán phân loại cảm xúc truyền thống bằng cách gắn cảm xúc với từng khía cạnh cụ thể. Ví dụ, cùng một câu “Giảng viên nhiệt tình nhưng phòng học nóng” sẽ được phân tích thành hai nhận định: (giảng viên, positive) và (cơ sở vật chất, negative). Phương pháp Pair-ABSA (hay QA-style ABSA) mô hình hóa nhiệm vụ này bằng cách đặt câu hỏi cho từng khía cạnh, giúp mô hình tập trung vào phần văn bản liên quan và giảm nhầm lẫn trong trường hợp một câu chứa nhiều chủ đề.

3.3 PhoBERT

PhoBERT là phiên bản RoBERTa được huấn luyện trên kho ngữ liệu tiếng Việt quy mô lớn, hỗ trợ tokenizer phù hợp với đặc thù tiếng Việt (sử dụng đơn vị syllable). Nhờ nền tảng này, PhoBERT cho khả năng biểu diễn ngữ nghĩa giàu thông tin và là lựa chọn tự nhiên khi cần fine-tune cho các tác vụ NLP tiếng Việt như phân loại văn bản, nhận diện thực thể hay – trong đề tài này – phân tích cảm xúc theo khía cạnh.

3.4 Công nghệ xây dựng ứng dụng

- **Ngôn ngữ:** Python.
- **Web framework:** Flask, tích hợp Flask-Login (xác thực người dùng), Flask-WTF/WTForms (xử lý form), Flask-SQLAlchemy (ORM làm việc với SQLite).
- **Cơ sở dữ liệu:** SQLite, file `feedback_analysis.db` lưu trong thư mục `instance/`.
- **Thư viện học máy:** PyTorch, Transformers, tokenizers, `huggingface-hub` để tải tokenizer và trọng số mô hình từ Hugging Face.
- **Giao diện:** HTML template với Jinja2, CSS/Bootstrap, JavaScript (file `static/js/app.js`) để gửi request, hiển thị kết quả và lịch sử feedback.

4 Dữ liệu (Dataset)

4.1 Nguồn dữ liệu

- **UIT_VSFC:** bộ dữ liệu phản hồi sinh viên tiếng Việt gắn nhãn khía cạnh (topic) và cảm xúc (sentiment). Đây là nguồn dữ liệu chính, phản ánh tình huống thực tế với nhiều câu ngắn gọn, khẩu ngữ, xen lẫn lỗi chính tả.
- **Dữ liệu synthetic:** nhóm tự tạo thêm các câu phản hồi dựa trên các mẫu câu thực tế nhằm mở rộng phạm vi khía cạnh và giảm hiện tượng mất cân bằng (đặc biệt ở lớp neutral). Sau khi kết hợp, toàn bộ dữ liệu đều được chuẩn hóa và khử trùng lặp theo khoá sentence .

4.2 Quy mô và nhãn

- Tổng số câu sau chuẩn hóa: **20.221**.

- Khía cạnh sử dụng trong dự án (4 lớp): `giang_vien`, `chuong_trinh`, `co_so_vat_chat`, `khac`.
- Cảm xúc gốc (3 lớp): `negative`, `neutral`, `positive`. Khi huấn luyện Pair – ABSA, hệ thống bổ sung thêm lớp `none` để biểu diễn trường hợp câu không đề cập đến khía cạnh đang xét.

4.3 Phân phối topic × sentiment

Bảng 1: Thống kê phân phối sentiment theo topic

topic_id	total	class_0 (neg)	class_1 (neu)	class_2 (pos)	0_pct (%)	1_pct (%)	2_pct (%)
0	12604	4492	487	7625	35.6	3.9	60.5
1	4036	2727	362	947	67.6	9.0	23.5
3	1803	721	427	655	40.0	23.7	36.3
2	1778	1105	225	448	62.1	12.7	25.2

Nhận xét:

- Topic 0 (tương ứng nhóm “giảng viên”) chiếm số lượng lớn nhất, đồng thời tỉ lệ phản hồi tích cực (class_2) vượt trội. Điều này gợi ý rằng sinh viên thường dành lời khen cho giảng viên nhiều hơn chê trách.
- Topic 1 (“chương trình”) thiên về phản hồi tiêu cực (class_0 chiếm 67.6%), phản ánh mối quan tâm của sinh viên đến lịch học, nội dung học phần, chính sách tín chỉ.
- Topic 3 và topic 2 (tương ứng các nhóm nhỏ hơn như “khác” hoặc “cơ sở vật chất”) có quy mô dữ liệu khiêm tốn, dẫn đến nguy cơ mô hình thiên vị những khía cạnh lớn nếu không xử lý cân bằng.
- Ở tất cả các topic, lớp neutral chiếm tỷ lệ thấp, là nguyên nhân khiến việc cân bằng bằng class weight và negative sampling trở nên quan trọng.

4.4 Chiến lược chia tập và tiền xử lý

- **Chia tập:** 70% Train (14.154 câu) – 30% Validation (6.067 câu) theo phân tầng tổ hợp topic × sentiment để giữ phân phối nhãn ổn định.
- **Chuẩn hóa văn bản:** chuẩn Unicode NFC, loại bỏ khoảng trắng thừa; tạo thêm bản không dấu để dò từ khóa liên quan đến khía cạnh khi lấy mẫu âm (“none”).

- **Khử trùng lặp:** loại bỏ các câu trùng nhau nhằm tránh việc mô hình học lặp thông tin.
- **Tokenization:** sử dụng tokenizer của PhoBERT, mã hóa cặp (prompt, sentence) với `max_length = 256`, padding cố định và cắt bớt phần câu nếu vượt quá độ dài.

4.5 Ví dụ minh họa

- “Giảng viên giảng dễ hiểu và hỗ trợ nhiệt tình.” → Khía cạnh `giang_vien`, cảm xúc `positive`.
- “Phòng học quá nóng, không bật điều hòa.” → Khía cạnh `co_so_vat_chat`, cảm xúc `negative`.
- “Thủ tục đăng ký học phần hơi rườm rà.” → Khía cạnh `chuong_trinh`, cảm xúc `negative`.
- “Thủ tục văn phòng một cửa phản hồi chậm.” → Khía cạnh `khac`, cảm xúc `negative`.

5 Phương pháp và Mô hình

5.1 Kiến trúc mô hình

- **Backbone:** PhoBERT –base (tất cả bộ tham số được fine-tune).
- **Head phân loại:** MLP hai tầng gồm Dropout → Linear → GELU → LayerNorm → Dropout → Linear, đầu ra 4 lớp (none/negative/neutral/positive) với hệ số dropout 0.3.
- Đầu vào Pair – ABSA: ghép prompt mô tả khía cạnh (ví dụ “Dánh giá giảng viên về giảng dạy, thái độ...”) với câu phản hồi thông qua tokenizer; lấy embedding [CLS] để phân loại

5.2 Cấu hình huấn luyện

- **Loss:** CrossEntropy kết hợp label smoothing = 0.1 nhằm giảm hiện tượng mô hình quá tự tin vào một lớp.
- **Class weights:** tính dựa trên phân phối nhãn, sau đó điều chỉnh sàn cho lớp none bằng hệ số alpha ($0.18 \rightarrow 0.15 \rightarrow 0.12$) để tránh mô hình bỏ qua lớp này.

- **Negative sampling:**
 - + Epoch 1 – 4: mỗi câu sinh 1 mẫu âm (khía cạnh khác).
 - + Epoch ≥ 5 : tăng lên 2 mẫu âm để cải thiện khả năng phân biệt **none**.
- **Optimizer:** AdamW với hai nhóm tham số, LR backbone = 1e-5, LR head = 1e-4, weight_decay = 3e-4.
- **Scheduler:** CosineAnnealingLR ($T_{\text{max}} = 8$, $\text{eta}_{\text{min}} = \text{LR}_{\text{backbone}}/10$) giúp giảm dần learning rate theo dạng cos.
- **Kỹ thuật ổn định:** Mixed Precision (AMP) và gradient clipping (1.0) để tránh exploding gradient.
- **Batch size:** 64; **Epoch:** 8.

5.3 Tác dụng của Pair – ABSA trong hệ thống

- **Tập trung theo khía cạnh:** prompt đóng vai trò “truy vấn”, giúp mô hình chú ý đúng phần văn bản liên quan đến khía cạnh cần đánh giá, giảm nhiễu từ các chi tiết không liên quan.
- **Xử lý câu đa khía cạnh:** một câu có thể được dự đoán nhiều lần với các prompt khác nhau, từ đó thu được đánh giá riêng cho từng khía cạnh, hạn chế nhầm lẫn giữa các nhóm.
- **Học nhân “none” rõ ràng:** việc chủ động tạo cặp âm (khía cạnh không xuất hiện) khiến mô hình học tốt cách nhận biết câu không đề cập khía cạnh; điều này phản ánh trực tiếp qua F1 cao của lớp **none**.
- **Mở rộng linh hoạt:** muốn thêm khía cạnh mới chỉ cần thiết kế prompt và cung cấp ví dụ, không phải thay đổi cấu trúc đầu ra.
- **Thuận tiện trong báo cáo:** dự đoán đã gắn sẵn khía cạnh nên dễ dàng tổng hợp số liệu theo từng nhóm (giảng viên, chương trình, cơ sở vật chất, khác).
- **Tái sử dụng backbone:** PhoBERT được dùng chung cho mọi khía cạnh, giảm chi phí huấn luyện so với việc xây dựng mô hình riêng rẽ.

6 Thực nghiệm

1. Thực hiện huấn luyện 8 epoch với lịch negative sampling thay đổi ($1 \rightarrow 2$).
2. Đánh giá sau mỗi epoch bằng các thước đo:
 - **Acc@4**: số dự đoán chính xác trên tổng số cặp ở bốn lớp.
 - **Acc@pos3**: độ chính xác trên ba lớp cảm xúc thực (bỏ lớp none). Chỉ số này cho biết tỷ lệ dự đoán đúng khi câu thật sự có nháy khía cạnh.
 - **Macro-F1@pos3**: trung bình cộng F1 của ba lớp negative, neutral, positive; nhấn mạnh tính cân bằng giữa các lớp cảm xúc.
 - **Precision/Recall/F1 lớp none**: đánh giá riêng khả năng nhận biết “không dễ cặp khía cạnh”.
3. Ghi lịch sử loss/accuracy theo epoch, trực quan hóa đồ thị và confusion matrix để phân tích lỗi.

7 Kết quả

Bảng 2: Kết quả huấn luyện theo epochs

Epoch	ValLoss	ValAcc4	ValAcc_pos3	ValMacroF1_pos3	ValNone (P/R/F1)
1	0.9060	0.8479	0.7584	0.7645	0.839 / 0.937 / 0.886
2	0.8316	0.8790	0.8055	0.8105	0.868 / 0.953 / 0.908
3	0.8263	0.8821	0.8492	0.8352	0.914 / 0.915 / 0.914
4	0.8255	0.8931	0.8367	0.8296	...

Nhận xét:

- Acc@4 đạt trên 0.89 từ epoch 4, chứng tỏ mô hình phân loại ổn định trên cả bốn lớp.
- F1 lớp none duy trì quanh 0.91, cho thấy chiến lược negative sampling và điều chỉnh class weight hiệu quả.
- Macro-F1@pos3 tăng từ ~ 0.76 (epoch 1) lên ~ 0.83 (epoch 3–4), phản ánh khả năng phân biệt cảm xúc thực sự ngày càng tốt.
- Loss validation giảm dần, không xuất hiện dấu hiệu overfitting trong 8 epoch.

8 Thảo luận và Nhận xét

8.1 Điểm mạnh

- Mô hình nhận diện tốt lớp none nhờ thiết kế Pair-ABSA và negative sampling có kiểm soát.
- Acc@4 và Macro-F1@pos3 cao, chứng minh khả năng phân loại cảm xúc chính xác ở bối cảnh dữ liệu tiếng Việt.
- Kiến trúc Pair-ABSA cho phép mở rộng số khía cạnh linh hoạt, phù hợp khi nhà trường muốn theo dõi thêm các nhóm dịch vụ khác.

8.2 Hạn chế

- Mô hình chưa thực sự hoạt động tốt đối với những câu có đa khía cạnh mà trong đó mỗi khía cạnh lại mang cảm xúc trái ngược nhau, dễ bị lẫn nhau sentiment cho câu feedback.
- Lớp neutral ít mẫu, dẫn tới F1 của lớp này vẫn thấp hơn so với hai lớp còn lại.
- Tập dữ liệu đang chưa được cân bằng giữa các topic và giữa các nhãn sentiment trong từng topic.

8.3 Lỗi thường gặp

- Câu chứa nhận xét lẫn lộn (ví dụ “giảng viên nhiệt tình nhưng bài giảng hơi nhanh”) dễ bị nhầm giữa negative và positive.
- Các câu đề cập khía cạnh mờ nhạt hoặc dùng ẩn dụ có thể bị dự đoán nhầm sang none.
- Câu có 2 khía cạnh thường bị nhầm nhãn sentiment.

9 Kết luận

Dề tài đã xây dựng thành công pipeline ABSA tiếng Việt dựa trên PhoBERT và phương pháp Pair – ABSA hoạt động tốt trong tác vụ phân loại feedback sinh viên khi câu chỉ bao gồm 1 khía cạnh duy nhất, đạt Acc@4 ~0.89 và Macro-F1@pos3 ~0.83 trên tập validation. Mô hình thể hiện khả năng nhận diện tốt lớp none nghĩa là biết được khi

nào thì câu đi với khía cạnh này và khi nào không, cung cấp nền tảng để có thể tiếp tục phát triển sau này.

PHẦN II: ỨNG DỤNG WEBSITE PHÂN TÍCH PHẢN HỒI SINH VIÊN

1 Tổng quan hệ thống

1.1 Mục tiêu sử dụng

Mục tiêu của website là đưa mô hình PhoBERT Pair-ABSA vào một sản phẩm có thể sử dụng được trong thực tế. Người dùng cuối (giảng viên, cán bộ quản lý, trợ giảng, ...) không cần biết tới mô hình hay code Python phía dưới mà vẫn có thể:

- Đăng nhập hệ thống bằng tài khoản cá nhân.
- Nhập nhanh các câu phản hồi của sinh viên hoặc tải lên file CSV chứa nhiều phản hồi.
- Nhận kết quả phân tích cảm xúc theo từng khía cạnh, lưu lại vào cơ sở dữ liệu để tra cứu về sau.

1.2 Đối tượng sử dụng

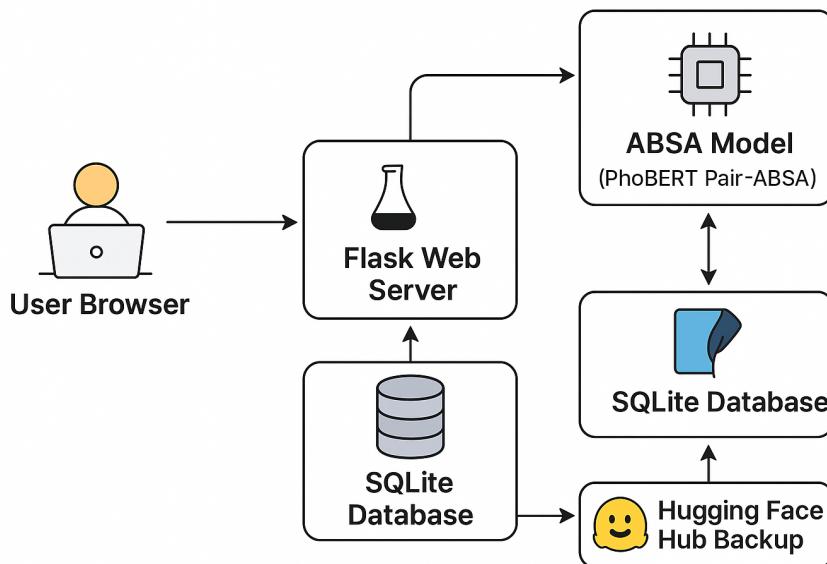
- **Giảng viên:** xem lại phản hồi của sinh viên về học phần mình phụ trách.
- **Cán bộ quản lý:** theo dõi bức tranh cảm xúc chung về chương trình đào tạo, cơ sở vật chất.
- **Nhóm nghiên cứu / phát triển:** thử nghiệm mô hình ABSA trên dữ liệu mới và ghi nhận lỗi thực tế.

2 Kiến trúc ứng dụng

2.1 Kiến trúc tổng thể

Ứng dụng được xây dựng theo mô hình client-server gồm ba lớp chính:

- **Giao diện người dùng (UI)**: các trang HTML được render bởi Flask, sử dụng Bootstrap để bố cục và CSS, JavaScript để xử lý sự kiện phía trình duyệt.
- **Lớp xử lý nghiệp vụ (Flask backend)**: nhận yêu cầu từ người dùng, gọi mô hình PhoBERT Pair-ABSA để suy luận, thao tác với cơ sở dữ liệu, kiểm tra phân quyền.
- **Lớp dữ liệu**: cơ sở dữ liệu SQLite lưu người dùng, phản hồi và kết quả phân tích; ngoài ra có cơ chế sao lưu/khôi phục lên Hugging Face Hub.



Hình 1: Kiến trúc tổng thể của hệ thống website

2.2 Các thành phần chính

- **Flask app (app.py)**: khai báo ứng dụng, cấu hình kết nối cơ sở dữ liệu, đăng ký các *route* xử lý (trang chủ, đăng nhập, dự đoán, lịch sử, ...), khởi tạo mô hình PhoBERT và scheduler backup.
- **Mô hình ABSA**: được đóng gói trong các file Python riêng (model, tokenizer, hàm `analyze_feedback()`) và được *import* để dùng trong route `/predict` và route xử lý CSV.
- **Lớp dữ liệu**: sử dụng SQLAlchemy để định nghĩa các bảng User, Feedback; lớp DatabaseManager phụ trách backup/restore dưới dạng JSON.
- **Giao diện**: các template trong thư mục `templates/` và file JavaScript trong `static/js/` để hiện kết quả trực quan.

2.3 Luồng hoạt động cơ bản

1. Người dùng truy cập website, đăng nhập bằng tài khoản đã đăng ký.
2. Khi nhập một câu phản hồi và bấm “Phân tích”, trình duyệt gửi yêu cầu AJAX (JSON) tới route `/predict`.
3. Backend kiểm tra dữ liệu đầu vào, gọi mô hình ABSA để phân tích từng khía cạnh.
4. Kết quả được lưu vào bảng `feedbacks` và trả ngược về cho trình duyệt ở dạng JSON.
5. JavaScript trên trình duyệt hiển thị kết quả theo từng thẻ (card) cảm xúc và cập nhật bảng lịch sử.

3 Thiết kế cơ sở dữ liệu

3.1 Mô hình dữ liệu tổng quan

Cơ sở dữ liệu gồm hai bảng chính: `users` và `feedbacks`. Quan hệ giữa chúng là 1-n: mỗi người dùng có thể gửi nhiều phản hồi, nhưng mỗi bản ghi phản hồi chỉ thuộc về một người dùng.

3.2 Bảng users

- `id`: khoá chính, số nguyên tự tăng.
- `username`: tên đăng nhập, duy nhất.
- `password_hash`: mật khẩu đã được băm (không lưu plaintext).
- `is_admin`: cờ đánh dấu tài khoản quản trị.
- `created_at`: thời gian tạo tài khoản.

Bảng này phục vụ xác thực và phân quyền trong hệ thống.

3.3 Bảng feedbacks

- `id`: khoá chính, số nguyên tự tăng.
- `text`: nội dung câu phản hồi gốc của sinh viên.

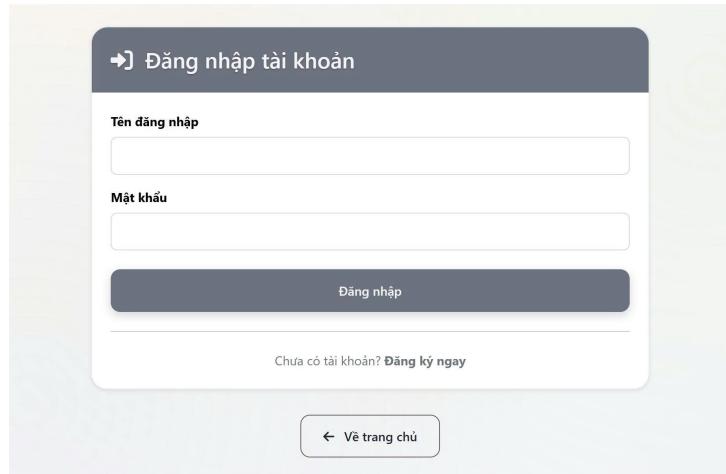
- **topic**: khía cạnh chính mà mô hình phát hiện (giảng viên, chương trình, cơ sở vật chất, khác).
- **sentiment**: cảm xúc đối với khía cạnh đó (negative, neutral, positive).
- **topic_confidence**: độ tin cậy khi gán khía cạnh.
- **sentiment_confidence**: độ tin cậy của dự đoán cảm xúc.
- **user_id**: khoá ngoại tham chiếu tới bảng **users**.
- **created_at**: thời gian hệ thống ghi nhận phản hồi.

Cấu trúc này cho phép truy vấn linh hoạt, ví dụ: lọc feedback theo người dùng, theo khoảng thời gian, hoặc thống kê tần suất/điểm số/tỷ lệ/chi tiêu theo từng khía cạnh.

4 Cài đặt giao diện và các chức năng

4.1 Chức năng đăng ký, đăng nhập

Giao diện đăng nhập/đăng ký được xây dựng trên Bootstrap, gồm form nhập *username*, *password* và nút gửi:



Hình 2: Giao diện đăng nhập hệ thống

Một số điểm chính:

- Sử dụng Flask-Login để quản lý phiên đăng nhập (`login_user`, `logout_user`).
- Mật khẩu được băm bằng thư viện như `bcrypt` hoặc `werkzeug.security`.
- Tài khoản admin có thêm quyền truy cập vào trang thống kê hoặc trang quản lý cơ sở dữ liệu.

4.2 Phân tích feedback đơn lẻ

Trên trang chính, người dùng nhập một câu phản hồi và bấm “Phân tích”:

Hình 3: Giao diện phân tích một feedback đơn lẻ

Sau khi gửi:

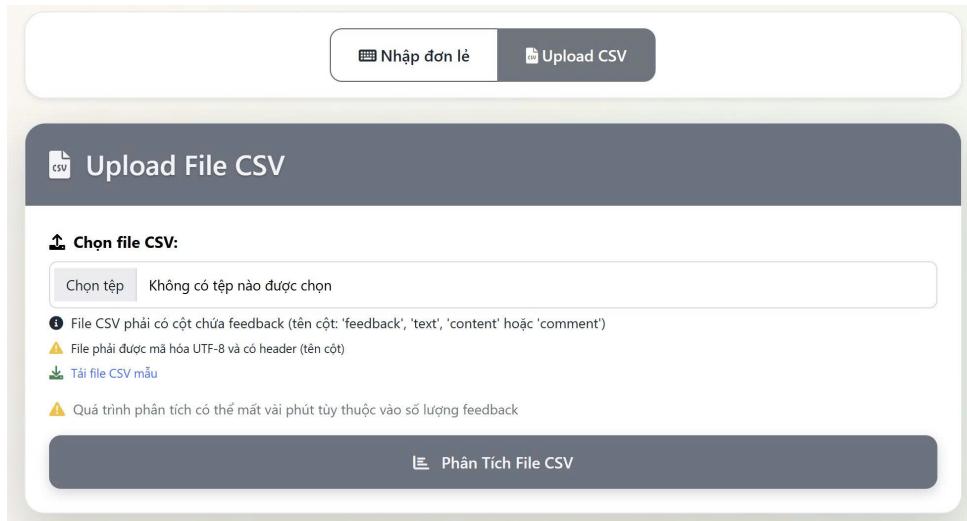
- JavaScript thu nội dung từ `textarea`, gửi AJAX tới `/predict`.
- Backend trả về danh sách các khía cạnh đã phát hiện, mỗi khía cạnh kèm cảm xúc và độ tin cậy.
- Giao diện hiển thị kết quả bằng các thẻ màu hoặc bảng, giúp người dùng dễ quan sát.

4.3 Phân tích file CSV

Hệ thống hỗ trợ tải lên một file CSV để phân tích hàng loạt:

Quy trình:

1. Người dùng chọn file CSV, hệ thống kiểm tra định dạng.
2. Backend đọc file, tự động nhận diện cột chứa nội dung phản hồi (dựa vào tên cột như `feedback`, `text`, ...).
3. Mỗi dòng được gửi qua mô hình phân tích; kết quả được lưu vào cơ sở dữ liệu.
4. Kết quả tổng hợp (số dòng xử lý thành công, số dòng lỗi, vài dòng minh họa) được trả về và hiển thị cho người dùng.



Hình 4: Giao diện phân tích hàng loạt từ file CSV

4.4 Xem lịch sử feedback

Người dùng có thể xem lại lịch sử các phản hồi đã gửi và kết quả phân tích:

Ngày	Tên Sinh Viên	Tín Cậy (%)
16/04/2020	Giảng Viên	100.0%
16/04/2020	Trung Tính	99.0%
16/04/2020	Giảng Viên	100.0%
16/04/2020	Tính Cực	83.1%
16/04/2020	Giảng Viên	100.0%
16/04/2020	Thầy giáo đã chấm điểm rất công bằng và minh bạch cho tất cả sinh viên.	80.2%
16/04/2020	Giảng Viên	100.0%
16/04/2020	Tính Cực	61.4%

Hình 5: Trang lịch sử feedback

Tính năng:

- Bộ lọc theo thời gian (hôm nay, 7 ngày gần nhất, 30 ngày, khoảng thời gian tự chọn).
- Hiển thị cột nội dung phản hồi, khía cạnh, cảm xúc, thời gian phân tích.
- Có thể mở rộng để xuất lịch sử ra CSV phục vụ thống kê ngoài hệ thống.

5 Đánh giá sản phẩm

5.1 Ưu điểm

- Giao diện web thân thiện, dễ sử dụng với cả người không có nền tảng kỹ thuật.
- Tích hợp trực tiếp mô hình ABSA nên người dùng nhận kết quả gần như thời gian thực.
- Lưu trữ dữ liệu tập trung, có lịch sử giúp dễ dàng tra cứu và phân tích xu hướng lâu dài.
- Cơ chế backup cơ sở dữ liệu giúp giảm rủi ro mất dữ liệu khi triển khai.

5.2 Hạn chế

- Mới hỗ trợ ngôn ngữ tiếng Việt, bộ khía cạnh còn đơn giản (4 nhóm chính).
- Hiệu năng phụ thuộc cấu hình máy chủ; khi phân tích file CSV lớn thời gian xử lý có thể tăng.
- Giao diện thống kê trực quan (biểu đồ, dashboard) còn hạn chế.