

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH



ĐỒ ÁN CUỐI KỲ:

Phân Tích Cảm Xúc Phản Hồi Của Sinh Viên Việt Nam

Lớp: CS221.Q13 – Xử Lý Ngôn Ngữ Tự Nhiên

Giảng viên: TS. Nguyễn Trọng Chính

Thành viên:

1. Nguyễn Hữu Khánh Duy - 23520375
2. Võ Hoàng Minh - 23520961
3. Hồ Hoàng Quân - 23521252

Mục Lục

CHƯƠNG I. GIỚI THIỆU BÀI TOÁN	4
1.1. <i>Bối cảnh bài toán.....</i>	4
1.2. <i>Mục tiêu của đề tài</i>	4
1.2.1. Đối với Giảng viên:.....	4
1.2.2. Đối với Nhà trường:.....	4
1.3. <i>Mô tả bài toán bằng ký hiệu toán học</i>	5
1.4. <i>Các thách thức chính</i>	5
CHƯƠNG II. BỘ NGỮ LIỆU.....	6
2.1. <i>Nguồn dữ liệu.....</i>	6
2.2. <i>Cấu trúc dữ liệu</i>	6
2.3. <i>Tổng quan dữ liệu.....</i>	7
2.3.1. <i>Tổng quan số lượng và phân phối nhãn.....</i>	7
2.3.2. <i>Đặc điểm độ dài văn bản</i>	8
2.3.3. <i>Thống kê tần suất xuất hiện của các từ.....</i>	10
2.4. <i>Phân tích một số mẫu trong bộ ngữ liệu</i>	16
2.4.1. <i>Nhãn Tiêu cực (Negative – Nhãn 0).....</i>	17
2.4.2. <i>Nhãn Trung tính (Neutral – Nhãn 1)</i>	20
2.4.3. <i>Nhãn Tích cực (Positive – Nhãn 2)</i>	22
CHƯƠNG III: PHƯƠNG PHÁP	25
3.1. <i>Bối cảnh ra đời</i>	25
3.2. <i>Kiến trúc chung</i>	26
3.2.1. <i>Lớp RobertaEmbeddings (embeddings)</i>	28
3.2.2. <i>Khối RobertaEncoder (encoder)</i>	32
3.2.3. <i>Khối RobertaClassificationHead (classifier)</i>	36
3.3. <i>Ví dụ cụ thể</i>	37
3.3.1. <i>Tiền xử lý văn bản.....</i>	38
3.3.2. <i>Tokenization với subword (BPE – Byte Pair Encoding)</i>	38

3.3.3. Encoding sang token IDs	39
3.3.4. Embedding layer	39
3.3.5. RoBERTa Encoder và Self-Attention.....	39
3.3.6. Biểu diễn cấp độ câu và Classification Head	40
3.3.7. Dự đoán và giải mã nhãn.....	40
CHƯƠNG IV: CÀI ĐẶT VÀ THỬ NGHIỆM	41
4.1. <i>Chia tập train và xử lý mất cân bằng dữ liệu.....</i>	41
4.2. <i>Tham số huấn luyện</i>	41
CHƯƠNG V: KẾT QUẢ ĐẠT ĐƯỢC	45
5.1. <i>Kết quả</i>	45
5.2. <i>Phân tích các trường hợp đúng/sai</i>	49
CHƯƠNG VI: CÁC HƯỚNG PHÁT TRIỂN TIẾP THEO.....	55
REFERENCE.....	55

CHƯƠNG I. GIỚI THIỆU BÀI TOÁN

1.1. Bối cảnh bài toán

Trong môi trường giáo dục hiện đại, phản hồi từ sinh viên (Student Feedback) đóng vai trò quan trọng giúp nhà trường và giảng viên đánh giá khách quan chất lượng giảng dạy. Những ý kiến này cung cấp cái nhìn thực tế về phương pháp sư phạm, nội dung chương trình và tình trạng cơ sở vật chất. Tuy nhiên, với số lượng phản hồi khổng lồ phát sinh sau mỗi học kỳ, việc xử lý thủ công là bất khả thi, đòi hỏi một giải pháp tự động hóa bằng kỹ thuật Xử lý ngôn ngữ tự nhiên (NLP).

1.2. Mục tiêu của đề tài

Mục tiêu chính của đề tài là xây dựng một hệ thống phân tích cảm xúc tự động có khả năng xử lý các phản hồi đa dạng của sinh viên, từ đó chuyển hóa dữ liệu thô thành các thông tin có giá trị quản lý. Cụ thể, đề tài hướng tới các mục tiêu chi tiết sau:

1.2.1. Đối với Giảng viên:

- **Nhận diện "điểm nghẽn" kiến thức:** Hệ thống giúp giảng viên xác định chính xác những phần nội dung hoặc khái niệm mà sinh viên đang gặp khó khăn (ví dụ: "giảng quá nhanh", "nội dung lý thuyết quá nặng").
- **Điều chỉnh phương pháp sư phạm kịp thời:** Thay vì đợi đến cuối kỳ, giảng viên có thể nhận được phản hồi định kỳ để điều chỉnh tốc độ, cập nhật slide hoặc bổ sung bài tập thực hành ngay trong quá trình giảng dạy.
- **Thấu hiểu tâm lý người học:** Phân tích cảm xúc giúp giảng viên nắm bắt được thái độ và nguyện vọng của sinh viên, từ đó thu hẹp khoảng cách giao tiếp và tạo môi trường học tập tích cực hơn.

1.2.2. Đối với Nhà trường:

- **Hệ thống đánh giá khách quan:** Cung cấp cho Ban giám hiệu và Phòng Đào tạo một bức tranh tổng thể về mức độ hài lòng của sinh viên đối với từng khoa, từng bộ môn hoặc chương trình đào tạo cụ thể.
- **Phát hiện và giải quyết vấn đề cơ sở vật chất:** Tự động sàng lọc các phản hồi tiêu cực về hạ tầng kỹ thuật như phòng học nóng, máy chiếu mờ, hoặc thiết bị thực hành yếu để có kế hoạch duy tu, nâng cấp trọng điểm.

- **Hỗ trợ ra quyết định dựa trên dữ liệu:** Kết quả phân tích là căn cứ khoa học để nhà trường xây dựng chính sách khen thưởng giảng viên, hoặc thiết kế các khóa bồi dưỡng kỹ năng sư phạm cho những đơn vị có phản hồi chưa tốt.

1.3. Mô tả bài toán bằng ký hiệu toán học

Để chuẩn hóa về mặt kỹ thuật, bài toán được phát biểu dưới dạng hàm mục tiêu như sau:

- **Định nghĩa không gian dữ liệu:** Gọi $D = \{(x_i, y_i)\}_{i=1}^N$ là tập dữ liệu tổng quát gồm N phản hồi đã gán nhãn
- **Thành phần:**
 - x_i : Nội dung văn bản phản hồi thứ i
 - $y_i \in L$: Nhãn cảm xúc tương ứng, với tập nhãn $L = \{\text{Negative}, \text{Neutral}, \text{Positive}\}$
- **Mô hình hóa:** Xây dựng hàm ánh xạ f từ không gian văn bản sang không gian nhãn cảm xúc:
 - **Đầu vào:** Một câu phản hồi x_i
 - **Đầu ra:** Nhãn dự đoán $\hat{y}_i = f(x_i)$ thể hiện thái độ tổng thể của sinh viên trong phản hồi đó.

1.4. Các thách thức chính

Quá trình giải quyết bài toán đối mặt với các khó khăn đặc thù sau:

- **Mất cân bằng dữ liệu (Data Imbalance):** Phản hồi Tích cực và Tiêu cực chiếm đa số, trong khi Trung tính chiếm rất ít, khiến mô hình dễ bị thiên vị (bias).
- **Cấu trúc ngôn ngữ phức tạp:**
 - **Cảm xúc hỗn hợp (Mixed sentiment):** Một câu chứa cả ý khen lẫn ý chê (ví dụ: "Thầy dạy hay nhưng phòng học nóng").
 - **Sự mỉa mai (Sarcasm):** Cảm xúc thực tế trái ngược với nghĩa đen của từ ngữ.
 - **Độ dài phản hồi:** Nhiều phản hồi quá ngắn (không đủ thông tin) hoặc quá dài (nhiều mệnh đề gây nhiễu).

CHƯƠNG II. BỘ NGỮ LIỆU

2.1. Nguồn dữ liệu

Trong đề tài này, nhóm sử dụng bộ ngữ liệu **UIT-VSFC (Vietnamese Students' Feedback Corpus)**. Đây là bộ dữ liệu tiêu chuẩn cho các nghiên cứu liên quan đến phân tích cảm xúc trong lĩnh vực giáo dục tại Việt Nam.

Bộ ngữ liệu bao gồm hơn **16.000 câu** phản hồi của sinh viên, được thu thập và gán nhãn thủ công bởi con người cho bài toán phân loại cảm xúc. Theo các nghiên cứu công bố, chất lượng của bộ dữ liệu đạt độ đồng thuận giữa những người gán nhãn (inter-annotator agreement) lên tới hơn **91%** đối với nhãn cảm xúc.

Chất lượng của bộ ngữ liệu UIT-VSFC đã được kiểm chứng qua các mô hình cơ sở (Baseline). Với bộ phân loại Maximum Entropy, nhãn cảm xúc đạt được chỉ số **F1-score xấp xỉ 88%**. Đây là một con số ấn tượng, khẳng định độ tin cậy của nhãn gán và tính khả thi khi triển khai các mô hình học máy chuyên sâu hơn trên bộ dữ liệu này.

2.2. Cấu trúc dữ liệu

Mỗi mẫu dữ liệu trong bộ ngữ liệu (Data Instance) bao gồm hai thành phần chính:

- Sentence (str): Nội dung câu phản hồi của sinh viên.
- Sentiment (int): Nhãn cảm xúc được mã hóa dưới dạng số nguyên với 3 giá trị:
 - 0: Tiêu cực (Negative)
 - 1: Trung tính (Neutral)
 - 2: Tích cực (Positive)

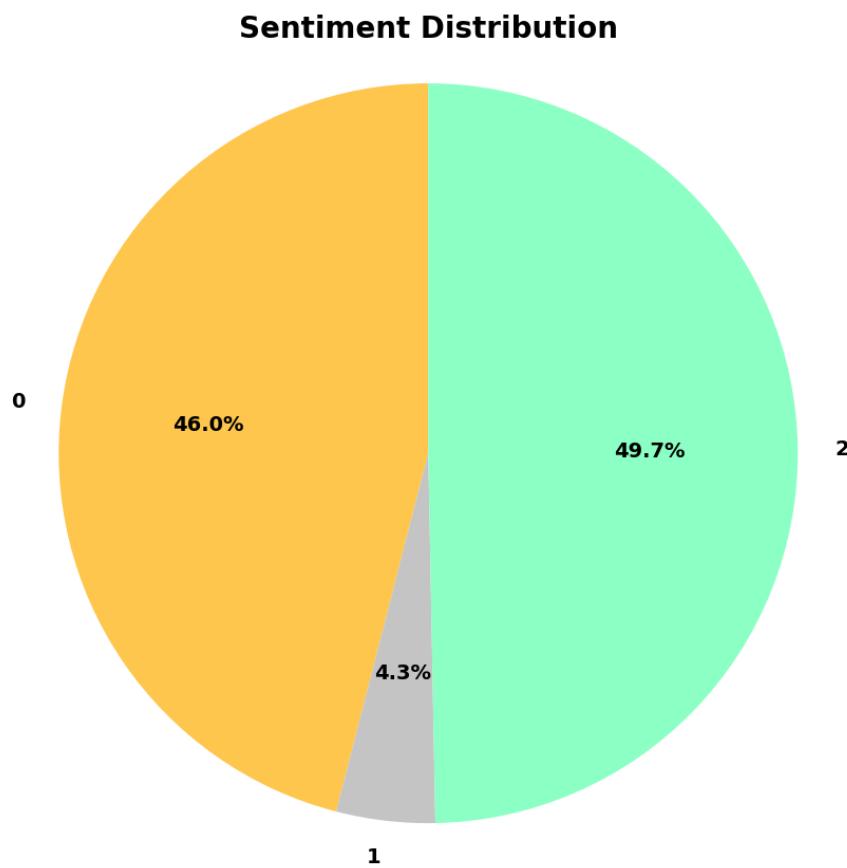
		sentence	sentiment
0		slide giáo trình đầy đủ .	2
1		nhiệt tình giảng dạy , gần gũi với sinh viên .	2
2		đi học đầy đủ full điểm chuyên cần .	0
3		chưa áp dụng công nghệ thông tin và các thiết ...	0
4		thầy giảng bài hay , có nhiều bài tập ví dụ ng...	2
5		giảng viên đảm bảo thời gian lên lớp , tích cự...	2
6		em sẽ nợ môn này , nhưng em sẽ học lại ở các h...	1
7		thời lượng học quá dài , không đảm bảo tiếp th...	0
8		nội dung môn học có phần thiếu trọng tâm , hầu...	0
9		cần nói rõ hơn bằng cách trình bày lên bảng th...	0

2.3. Tổng quan dữ liệu

2.3.1. Tổng quan số lượng và phân phối nhãn

Dựa trên kết quả thống kê từ tập dữ liệu thực tế:

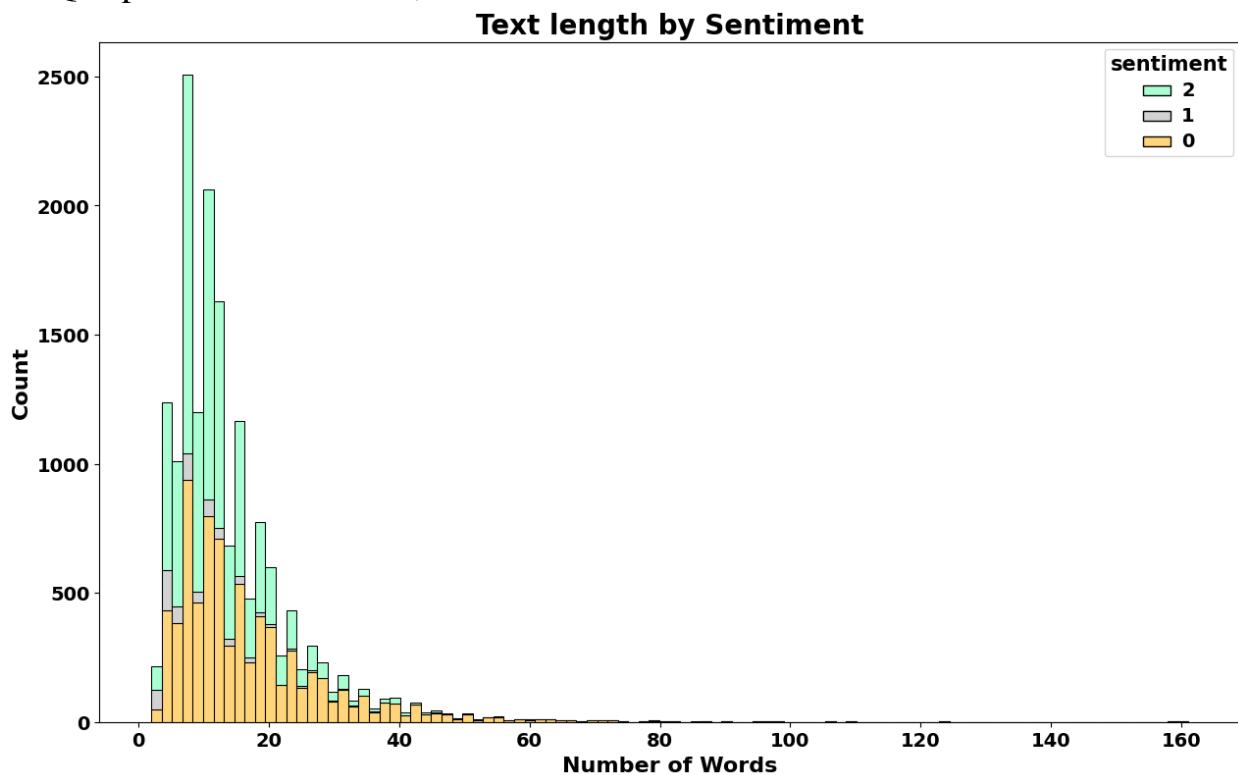
- Tổng số mẫu: 16,175 dòng dữ liệu
- Tỷ lệ phân phối nhãn:
 - Nhãn Tiêu cực (0): 7439 mẫu
 - Nhãn Trung tính (1): 698 mẫu
 - Nhãn Tích cực (2): 8038 mẫu



Nhận xét: Dữ liệu cho thấy sự mất cân bằng nghiêm trọng ở lớp Trung tính. Số lượng phản hồi Trung tính chỉ chiếm một phần rất nhỏ so với hai lớp còn lại, điều này minh chứng cho thách thức "Mất cân bằng dữ liệu" đã đề cập ở Chương 1.

2.3.2. Đặc điểm độ dài văn bản

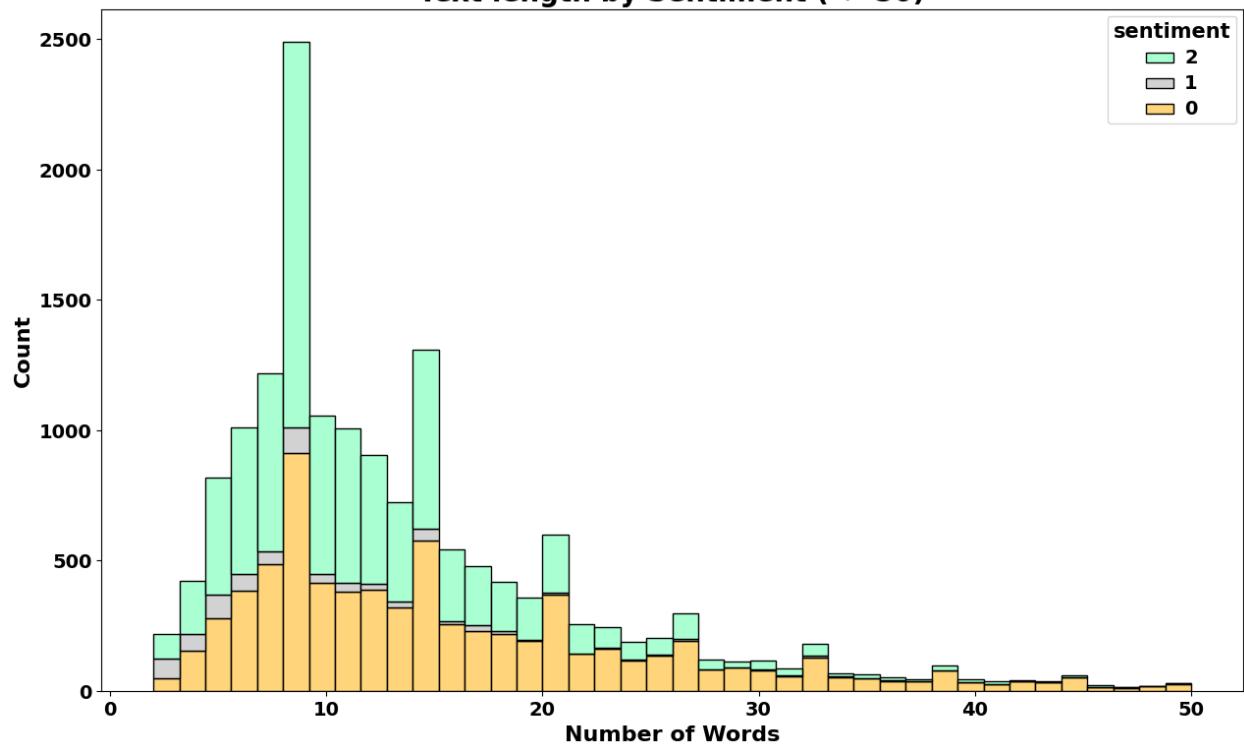
Qua phân tích biểu đồ độ dài câu:



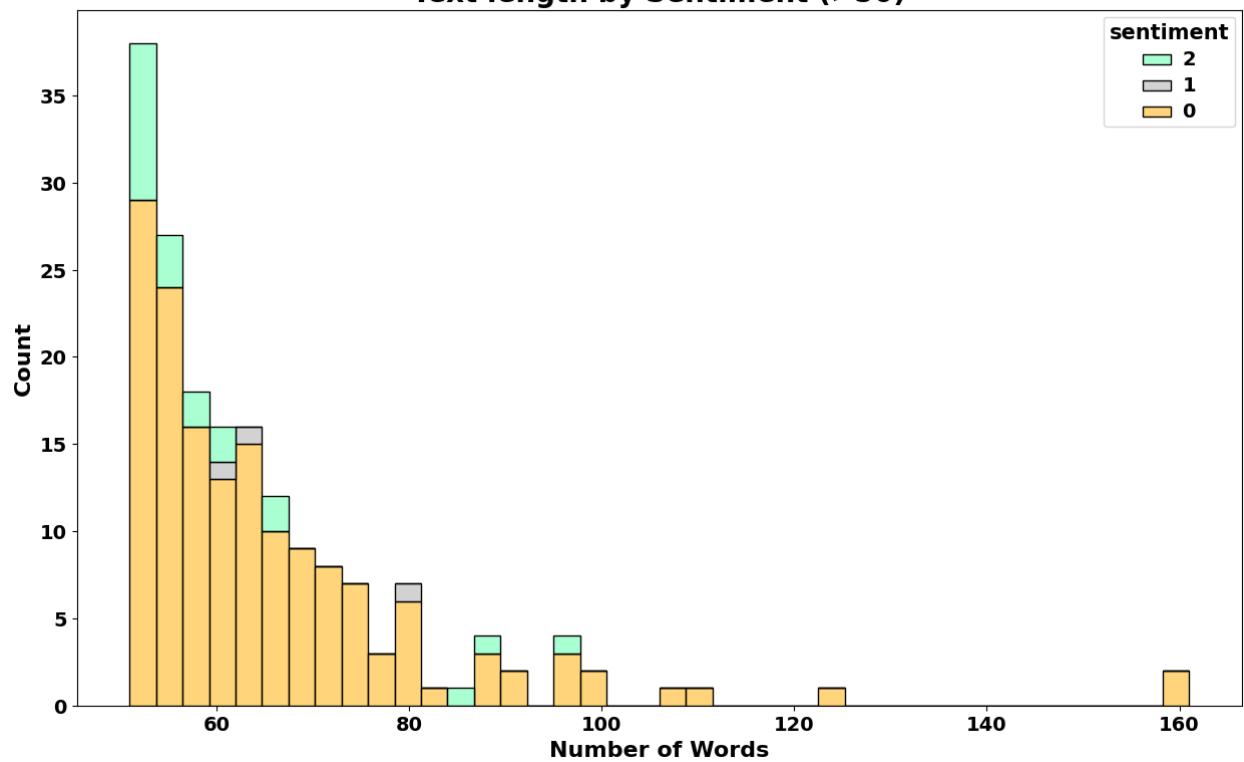
- **Phân phối độ dài:** Đa số các câu phản hồi có độ dài ngắn, tập trung trong khoảng từ **5 đến 20** từ.
- **Độ dài theo cảm xúc:** Các phản hồi Tích cực (màu xanh) và Tiêu cực (màu cam) có phân phối độ dài tương đương nhau, thường là các câu ngắn gọn, súc tích. Một số ít phản hồi có độ dài lớn (trên 40 từ), thường là các câu kể hoặc gộp ý chi tiết chứa nhiều mệnh đề phức tạp.

Ta chia ra 2 phía, phân phối có chiều dài ít hơn hoặc bằng 50 chữ và nhiều hơn 50 chữ:

Text length by Sentiment (<=50)



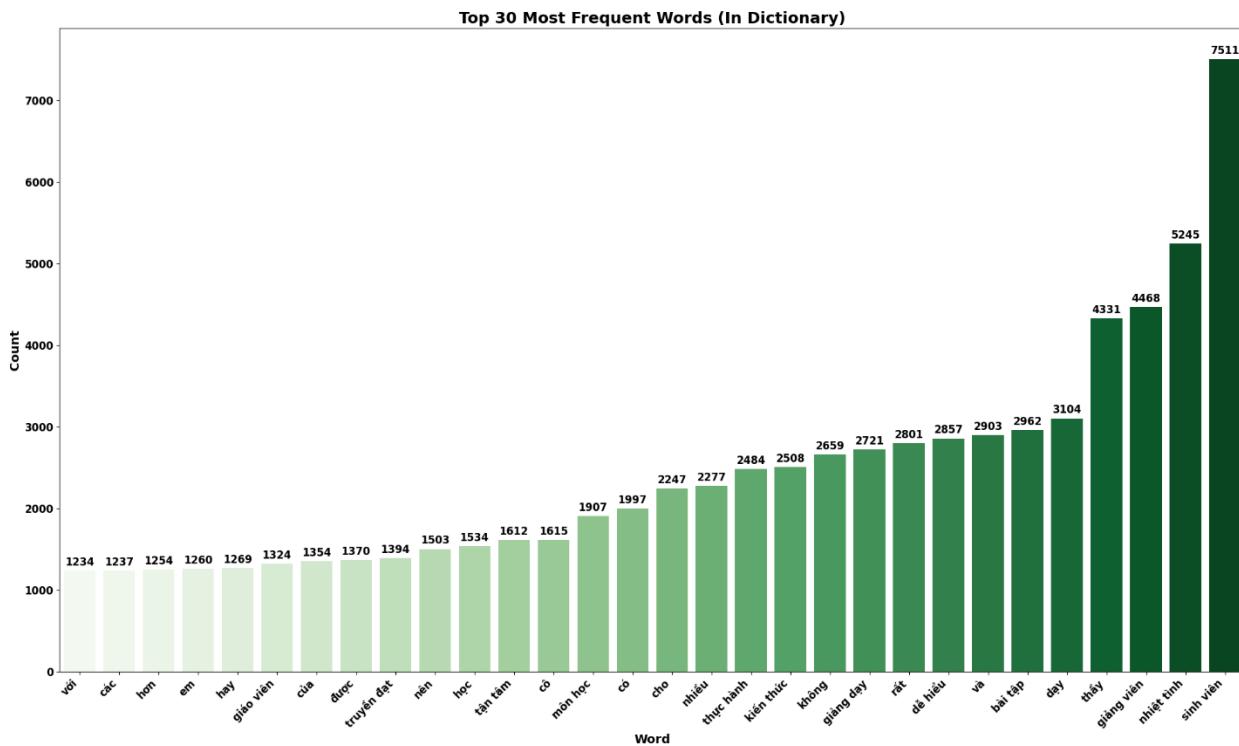
Text length by Sentiment (>50)



- Phân bố dữ liệu cho thấy rõ về quan hệ giữa số lượng chữ trong câu và cảm xúc của câu.
- Số lượng chữ của câu càng dài, khả năng mang cảm xúc tích cực ít lại và hầu như các câu chứa hơn 40 chữ thuộc cảm xúc tiêu cực.

2.3.3. Thống kê tần suất xuất hiện của các từ

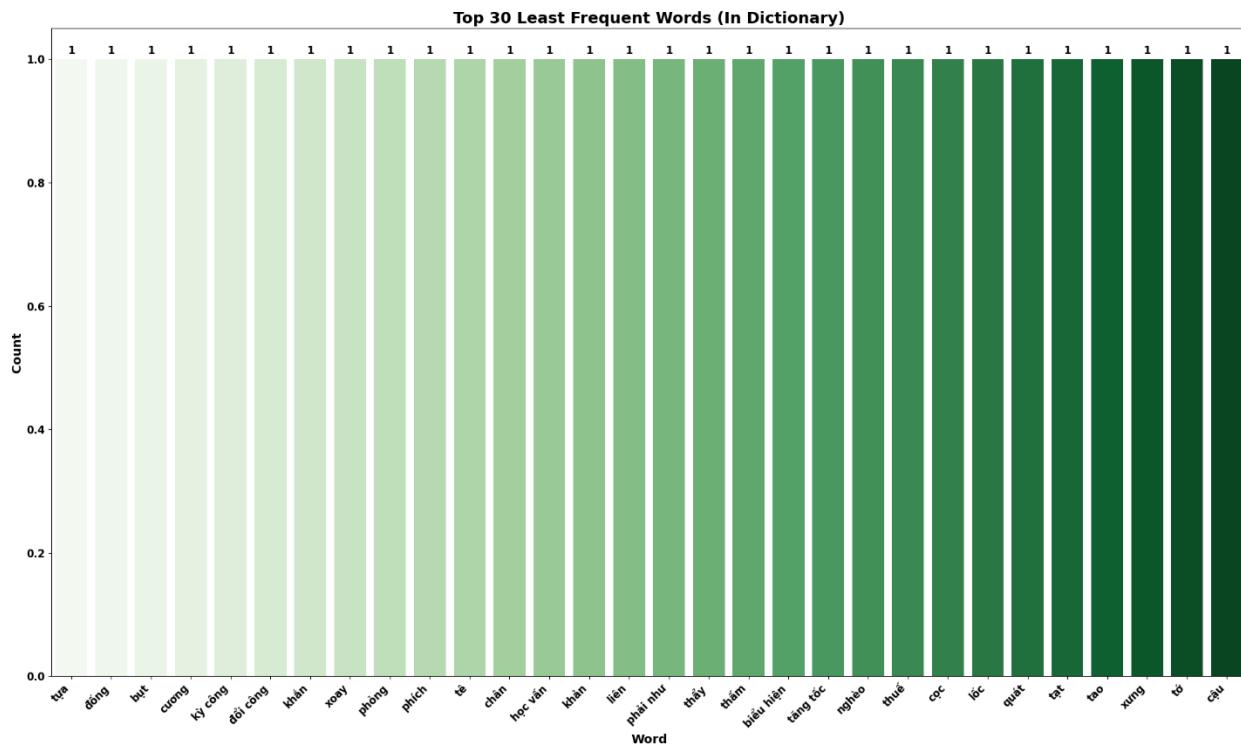
2.3.3.1. 30 từ có trong từ điển tiếng Việt và xuất hiện nhiều nhất



- Từ mang tính chức năng/ ngữ pháp chiếm đa số: “có”, “và”, “rất”. Điều này là đặc trưng của dữ liệu phản hồi tự nhiên, cho thấy sinh viên diễn đạt dưới dạng câu đầy đủ, không phải từ khóa rời rạc.
- Từ liên quan trực tiếp đến môi trường giáo dục xuất hiện với tần suất cao. Hầu hết phản hồi đều chứa chữ “giảng viên”, “thầy”, “cô” để nói người giảng dạy lớp thuộc môn học đó. Phản ánh trọng tâm feedback tập trung vào giảng viên và hoạt động giảng dạy – học tập, đúng với ngữ cảnh dữ liệu.
- Sự xuất hiện nhiều của “rất”, “nhiều”, “để hiểu”, “nhiệt tình” cho thấy:
 - Sinh viên thường đánh giá mức độ (rất, nhiều)

- Có xu hướng nhận xét về sự dễ hiểu, sự nhiệt tình của giảng viên
 - ⇒ Đây là các tín hiệu quan trọng cho phân tích cảm xúc (sentiment analysis).

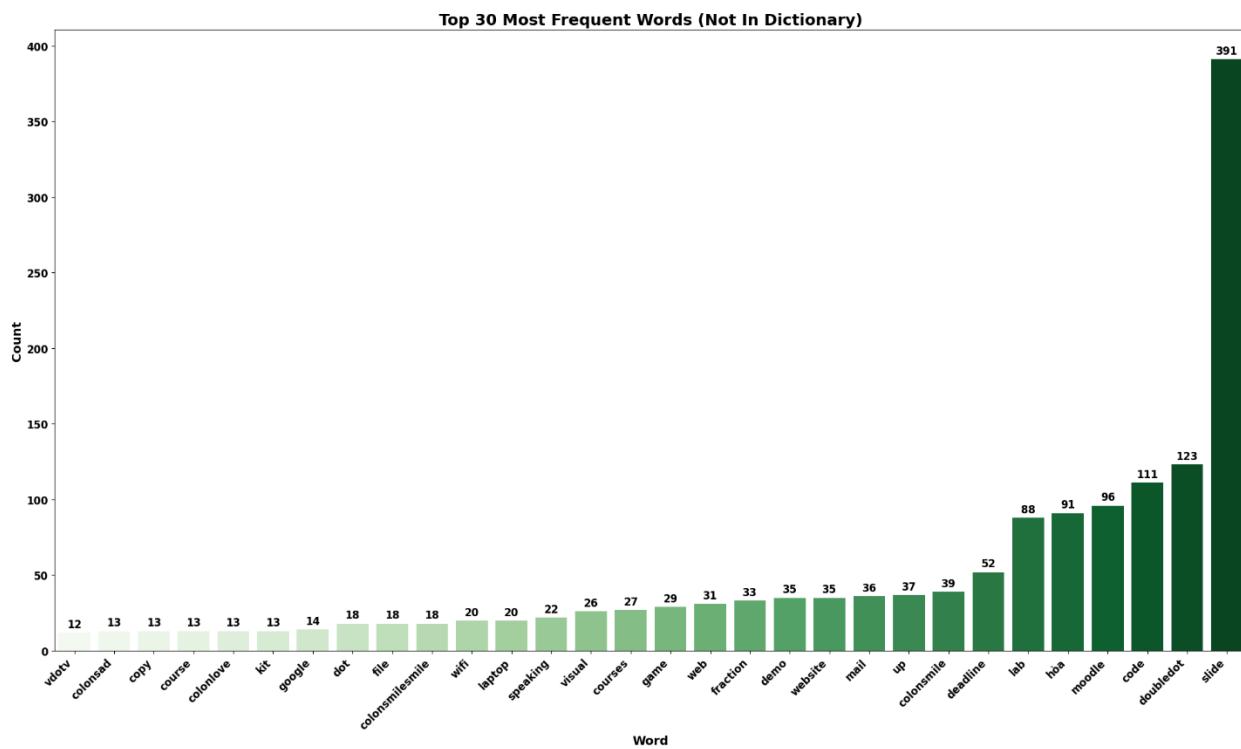
2.3.3.2. 30 từ có trong từ điển tiếng Việt và xuất hiện ít nhất



- Tất cả các từ đại diện xuất hiện trong biểu đồ đều có tần suất = 1, đây là các từ rất hiếm, chỉ xuất hiện đúng một lần trong toàn bộ tập dữ liệu. Điều này cho thấy vốn từ vựng đa dạng, nhưng phân bố tần suất rất lệch (long-tail distribution), đặc trưng của dữ liệu văn bản tự nhiên.
- Từ có thể mang tính cảm xúc hoặc đánh giá cá nhân “té”, “nghèo”, “cộc”, “lóc”, “quát” có thể phản ánh trải nghiệm tiêu cực cá biệt, không đại diện cho xu hướng chung
- Từ mang tính hành động / trạng thái cụ thể “xoay”, “phích”, “biểu hiện”, “thăm” thường gắn với ngữ cảnh rất riêng, không phổ quát.
- Các từ này:
 - Đóng góp rất ít về mặt thống kê.

- Vẫn có khả năng đóng góp thông tin nếu đặt trong ngữ cảnh đầy đủ.
- Không đủ tần suất để học quy luật tổng quát và dễ gây nhiễu trong các mô hình dựa trên Bag-of-Words hoặc TF-IDF
- Tuy nhiên các mô hình học sâu có thể giải quyết các từ này nếu kèm theo ngữ cảnh, đúng với thách thức của ngôn ngữ tự nhiên.

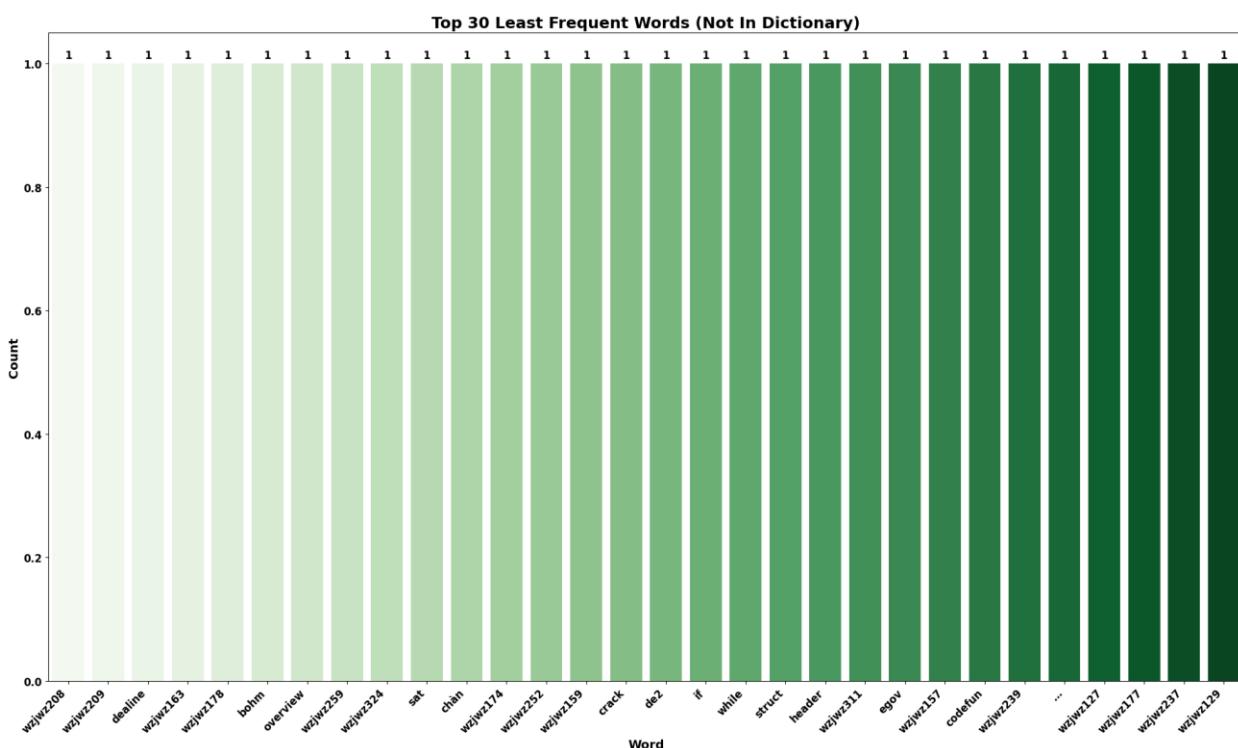
2.3.3.3. 30 từ không có trong từ điển tiếng Việt và xuất hiện nhiều nhất



- Phân bố rất lệch, trong đó:
 - Từ “slide” (391) vượt trội hoàn toàn
 - Nhóm tiếp theo: “doubleot” (~123), “code” (~111), “moodle” (~96), “hoa” (~91), “lab” (~88)
 - ⇒ Điều này ^đnh^đnh định nh^đnh từ ngoài từ điển kh^đng ph^đai nhiễu ng^đu nh^đnh, m^đ là các kh^đnh ni^đem trung t^đm trong feedback.
 - Các thuật ngữ liên quan đến học tập như slide, code, lab, moodle, web, website, laptop, ... có thể phản ánh:

- Tài liệu giảng dạy
 - Hệ thống giảng dạy và trao đổi tài liệu học tập
 - Nội dung thực hành, công nghệ sử dụng trong môn học
 - ...
- ⇒ Đây là từ khóa chủ đề (topic keywords) quan trọng.
- Từ mượn tiếng Anh demo, mail, up, copy, file, dot, kit, ... rất phổ biến trong ngôn ngữ sinh viên, nhưng không nằm trong từ điển tiếng Việt chuẩn.
 - Xuất hiện các từ colonsmile, colonsmilesmile, doubleot có thể là do lỗi kỹ thuật.

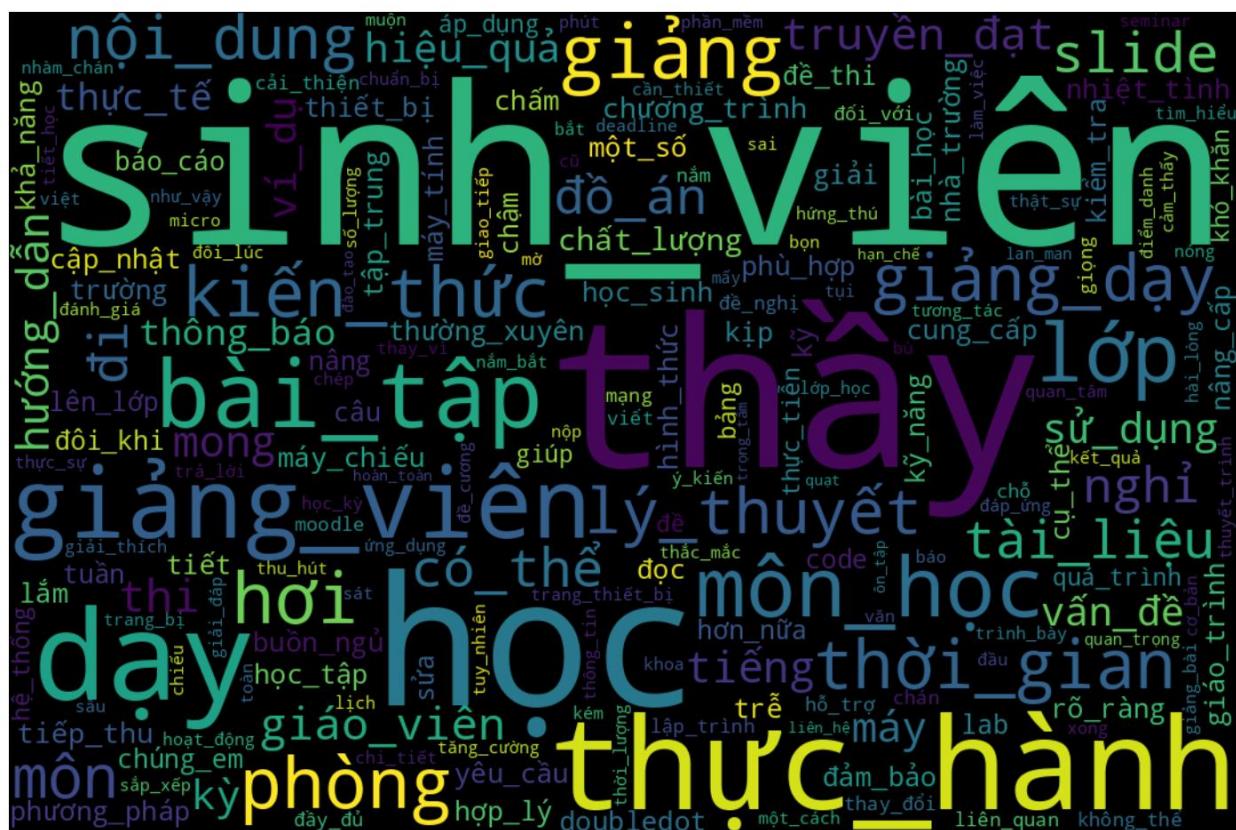
2.3.3.4. 30 từ không có trong từ điển tiếng Việt và xuất hiện ít nhất



- Có những từ ngữ liên quan đến học tập như “if”, “while”, ... cho thấy có những câu ngữ cảnh nói đến nội dung môn học.
- Hầu hết là các từ ẩn danh để ẩn đi tên tiếng cụ thể.
 - Có giá trị để mô hình học sâu hiểu rằng đây là các danh từ riêng của một người khi đặt trong đúng ngữ cảnh.

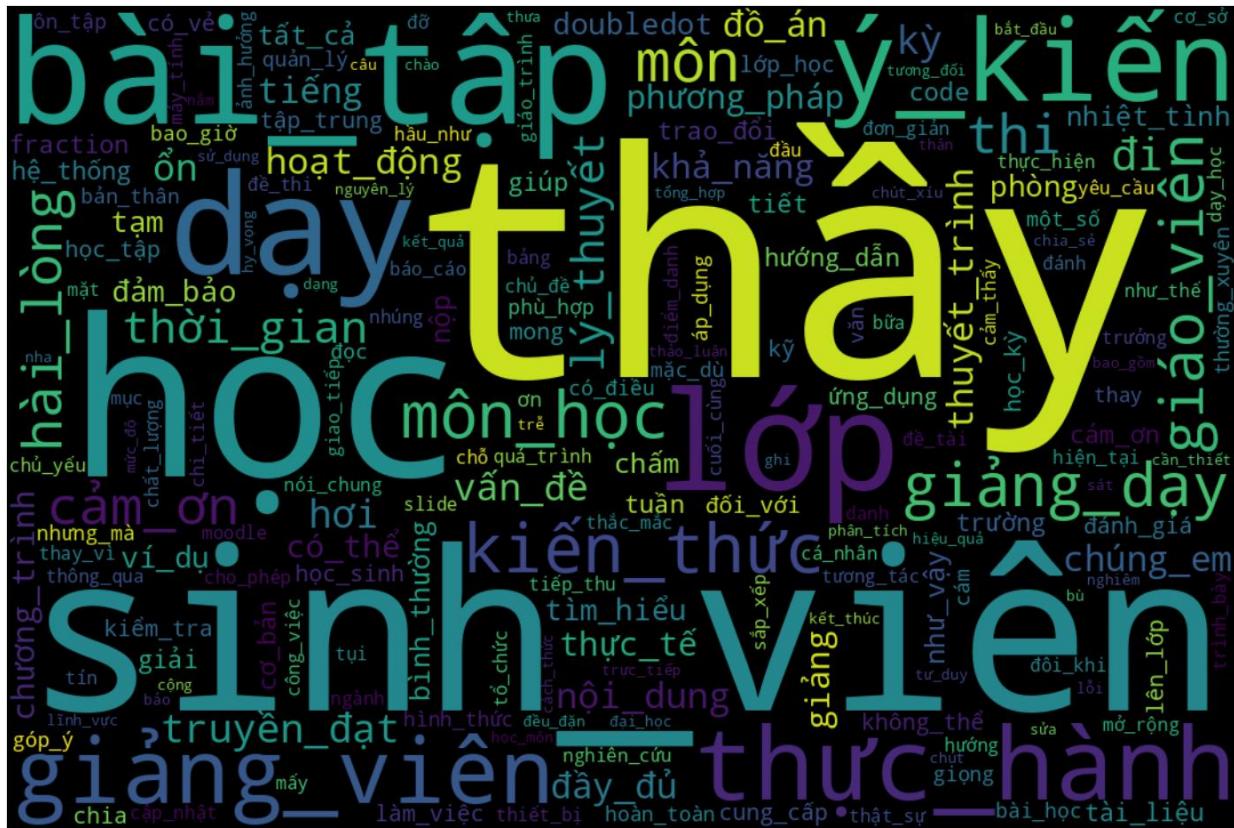
2.3.4. Trực quan số lượng từ trong từng phân loại bằng WordCloud

2.3.4.1. Các từ trong phân loại tiêu cực



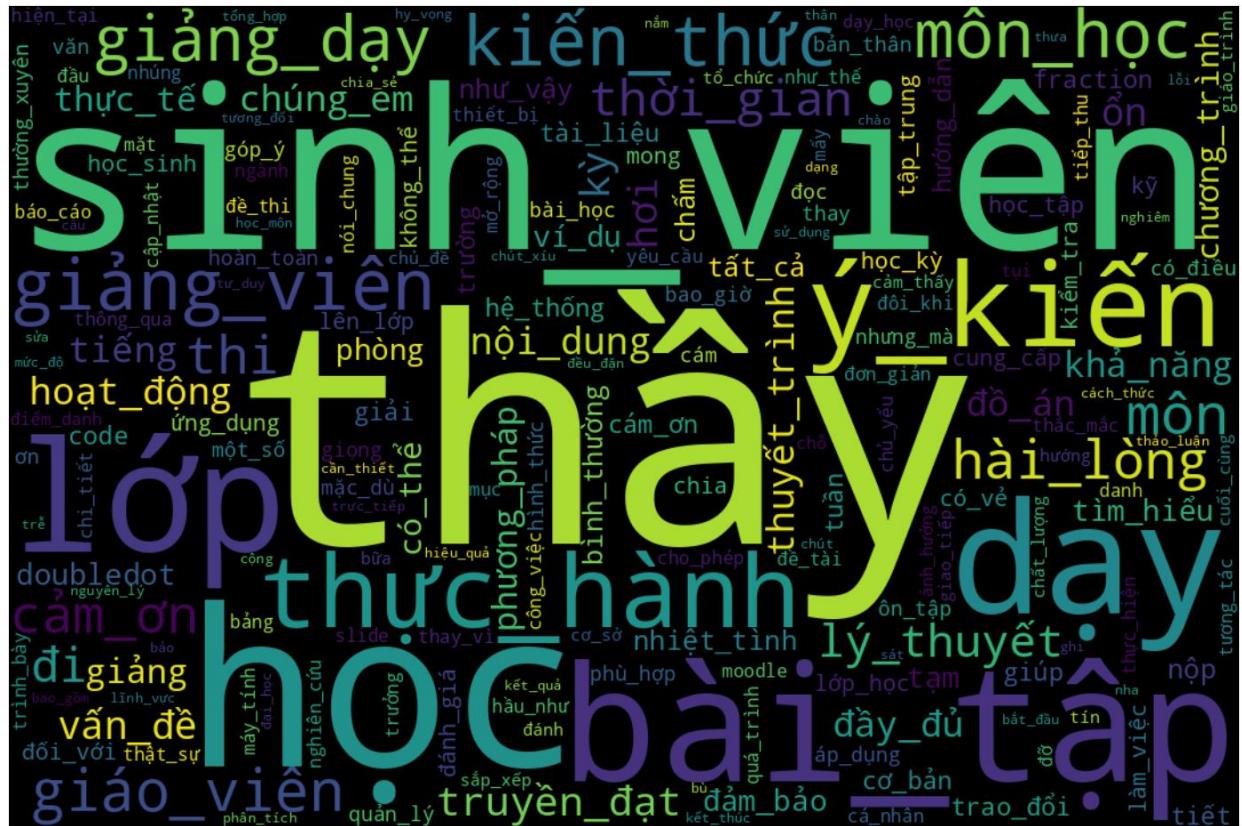
- Ngoài các từ phổ biến như “sinh viên”, “thầy”, ... có xuất hiện các từ mang ý nghĩa thể hiện cảm xúc tiêu cực rõ ràng.
 - Các từ như “*buồn ngủ*”, “*trễ*”, “*lan man*”, ... phản ánh tiêu cực cách truyền đạt hoặc cách làm việc của người phụ trách giảng dạy.
 - “đè nghị”, “mong”, “cải thiện”, ... phản ánh mong muốn của người học đối với môn học hoặc người giảng viên trong ngữ cảnh cụ thể.

2.3.4.2. Các từ trong phân loại trung tính



- Ngoài các từ phổ biến hầu hết các từ đều mang tính mong muốn bổ sung hoặc đánh giá một cách trung tính, ít xuất hiện các từ mang ý nghĩa thật sự đủ mạnh để phân thành tích cực hoặc tiêu cực.

2.3.4.3. Các từ trong phân loại tích cực



- Mang các từ chưa hàm ý tích cực như “*cần thiết*”, “*chi tiết*”, “*cảm ơn*” ... phản ánh tốt về thái độ và cảm nghĩ của sinh viên sau môn học.
 - Các từ “*kỹ*”, “*nghiêm túc*”, “*hiệu quả*” thể hiện cách nhìn nhận của sinh viên đối với cách làm việc của giảng viên trong quá trình giảng dạy.

2.4. Phân tích một số mẫu trong bộ ngũ liệu

Để đánh giá sâu hơn về đặc điểm ngôn ngữ và những thách thức thực tế đối với mô hình phân loại, nhóm đã thực hiện phân tích chi tiết trên **70 mẫu phản hồi** điển hình được trích xuất từ bộ ngữ liệu. Các mẫu này được phân loại dựa trên độ khó trong việc nhận diện cảm xúc đối với các thuật toán học máy.

2.4.1. Nhãn Tiêu cực (Negative – Nhãn 0)

Mức độ	Câu ví dụ	Cơ sở phân loại	Giải thích chi tiết
Dễ	"giảng viên truyền đạt kiến thức chưa tốt ."	"chưa tốt"	Câu ngắn, từ phủ định nhẹ nhàng, rõ ràng.
Dễ	"giảng viên dạy quá nhanh ."	"quá nhanh"	Mô tả một khiếm khuyết cụ thể, mang tính phàn nàn trực tiếp.
Dễ	" không nhiệt tình , không cập nhật slide ."	"không nhiệt tình", "không cập nhật"	Hai tín hiệu phủ định mạnh xuất hiện song song.
Dễ	" không nắm được thực lực , bài tập quá sức với sinh viên ."	"không nắm được", "quá sức"	Ý phàn nàn cụ thể, cấu trúc câu rõ ràng.
Dễ	"lên lớp không đầy đủ ."	"không đầy đủ"	Phủ định một hành động mong đợi của giảng viên.
Dễ	"phản thực hành , chúng em không được hướng dẫn đầy đủ ."	"không... đủ"	Thể hiện sự thiếu sót về mặt đào tạo một cách trực diện.
Dễ	"những vấn đề cô nói thường nhanh , dài và khó hiểu ."	"khó hiểu", "nhanh", "dài"	Sử dụng các tính từ tiêu cực diễn hình trong sự phạm.
Dễ	"phương pháp dạy của thầy chưa phù hợp , gây khó hiểu cho sinh viên ."	"chưa phù hợp", "khó hiểu"	Tín hiệu tiêu cực xuất hiện ở cả hai vế của câu.
Dễ	"phòng học không đáp ứng đủ nhu cầu học tập của sinh viên , máy chiếu khá mờ ."	"không", "khá mờ".	Các từ thể hiện sự tiêu cực phàn nàn về cơ sở vật chất.
Dễ	" không hiểu bài cho lắm."	"không hiểu bài".	Từ mang ý nghĩa tiêu cực rõ ràng.
Trung bình	" cần thêm vài ví dụ cụ thể hơn và có bài tập để rèn luyện"	Mong muốn cải thiện.	Không có từ tiêu cực; mô hình phải hiểu "cần thêm" nghĩa là hiện tại đang "thiếu".
Trung bình	"khi giảng thầy chưa chỉ rõ trọng tâm của bài để sinh viên nắm bắt và dễ tiếp thu hơn ,	Phàn nàn về kỹ năng truyền đạt.	Câu dài, nhiều nội dung chê nhẹ; yêu cầu hiểu

	giọng thầy còn nhỏ , đều đẽo .”		tổng thể thay vì từ khóa đơn lẻ.
Trung bình	"phòng thực hành cần được nâng cấp . "	Gián tiếp nói cơ sở vật chất chưa tốt.	Không có từ cảm xúc; đòi hỏi sự suy luận về trạng thái của đối tượng được nhắc đến.
Trung bình	“nội dung môn học cần rõ ràng hơn nữa để biết môn học học cái gì và khả năng tiếp nhận kiến thức của sinh viên .”	Hàm ý nội dung hiện tại chưa rõ.	Cách diễn đạt gián tiếp, mang tính chất đóng góp xây dựng nên dễ bị nhầm lẫn.
Trung bình	" cần quan tâm tới học sinh của mình hơn . "	Hàm ý giảng viên thiếu sự quan tâm.	Ngôn từ nhẹ nhàng, cảm xúc ẩn; mô hình phải hiểu được sự so sánh ngầm ("hơn").
Trung bình	“giảng viên nên hòa nhã , thoái mái để sinh viên có thể tương tác tốt với giảng viên , giảng viên nên diễn đạt câu từ một cách ngắn gọn dễ hiểu , nên dạy kỹ những phần kiến thức quan trọng .”	Liên tục dùng từ "nên" để chỉ ra thiếu sót.	Cấu trúc câu cầu khiến; mô hình dễ nhầm với trung tính nếu không hiểu mục đích gộp ý.
Trung bình	“ban đầu thì em có vẻ không hài lòng lắm về giáo viên khi học môn này , nhưng sau một khoảng thời gian thì thấy cũng ổn .”	Cấu trúc đối lập "nhưng".	Chứa cả hai luồng cảm xúc; mô hình cần nhận diện được sự chuyển đổi ý nghĩa ở về sau.
Khó	“hy vọng lần sau thầy dạy sẽ mang đến thêm một số bài tập kèm hướng dẫn để bọn em dễ dàng vận dụng kiến thức mình học hơn , khi thầy giảng xong , có thể bọn em nói bọn em hiểu nhưng bọn em không nhớ lâu được do không có chỗ để bọn em vận dụng , nên cuối cùng đến khi thi bọn em vẫn trả lời rằng bọn em không hiểu .”	Phàn nàn về hệ quả: "vẫn không hiểu".	Câu rất dài, kết hợp giữa lời góp ý mang tính xây dựng và sự phàn nàn về kết quả học tập.

Khó	“sử dụng từ ngữ không phù hợp , có thái độ phân biệt với một vài sinh viên , thường nói chuyện với các sinh viên nữ .”	"không phù hợp", "phân biệt".	Chứa các vấn đề đạo đức/xã hội; mô hình cần hiểu hàm nghĩa nằm ngoài phạm vi học thuật thông thường.
Khó	“cả 3 , 4 buổi thực hành theo hình thức 2 môn này có thầy cô này tham gia hay hướng dẫn buổi nào đâu mà đánh giá , chỉ có mỗi thầy wzjwz43 thôi .”	Chỉ trích sự vắng mặt của giảng viên.	Sử dụng cấu trúc hội thoại tự nhiên, khẩu ngữ và thực thể riêng; đòi hỏi quy trình tiền xử lý cực tốt.
Khó	"chương trình học khá nặng , quá nhiều kiến thức."	Phản ánh áp lực tiêu cực: "quá nhiều".	Ranh giới mong manh giữa "mô tả thực trạng" và "phàn nàn"; mô hình dễ nhầm với trung tính.
Khó	“dạy không đủ bài giảng có thi , không giúp sinh viên mở rộng các bài tập liên quan đến bài giảng , chỉ dạy y chang theo slide mà không rõ sinh viên đã hiểu chưa.”	Một loạt phủ định: "không đủ", "không giúp", "không rõ".	Câu dài với nhiều mệnh đề phức tạp; yêu cầu mô hình hiểu cấu trúc toàn câu để xác định cảm xúc tổng thể.
Khó	“ngược lại để thi nên cho những câu hỏi khó , nhưng thời gian thi rộng để sinh viên có thời gian động não suy nghĩ các giải tăng được tự duy suy luận và sáng tạo hơn là khả năng học thuộc bài như kiểu trên thích hợp cho ngành khác .”	"ngược lại" + "nên".	Sử dụng nhiều từ ngữ mang sắc thái tích cực nhưng hàm ý phàn nàn về cách ra đề thi hiện tại.
Khó	“đầu thì thông báo là sẽ lập nhóm khoảng 5 người để seminar lấy điểm giữa kỳ , sau đó không cho seminar nữa mà thông báo lập lò là " có thể " sẽ cho nhóm 2 người để là website cộng điểm , mới đây thì lại thông báo là mỗi người làm 1 website để cộng điểm .”	"không" + "lại".	Cấu trúc kể chuyện, mô hình phải nhận diện được sự thiếu nhất quán qua các mốc thời gian.

2.4.2. Nhãm Trung tính (Neutral – Nhãm 1)

Mức độ	Câu ví dụ	Cơ sở phân loại	Giải thích chi tiết
Dẽ	"thầy dạy đủ buổi ."	Mô tả sự việc khách quan.	Không chứa tính từ cảm xúc; thông tin thuần túy về số lượng buổi học.
Dẽ	"thầy dạy đủ để thi ."	Khẳng định mức độ hoàn thành.	Mô tả thực tế đáp ứng yêu cầu tối thiểu, không thể hiện sự hài lòng hay phàn nàn.
Dẽ	"giảng viên dạy được ."	Đánh giá ở mức chấp nhận.	Cụm "dạy được" mang tính nước đôi, không nghiêng về tích cực hay tiêu cực rõ rệt.
Dẽ	"dạy học ."	Cụm từ nêu chủ đề.	Chỉ là danh từ/động từ thuần túy, hoàn toàn không có ngữ cảnh cảm xúc đi kèm.
Dẽ	"dùng hệ thống we code"	Mô tả công cụ sử dụng.	Hoàn toàn không chứa từ chỉ cảm xúc hay đánh giá.
Dẽ	“bài tập trên lớp .”	Cụm danh từ	Chỉ nêu tên hoạt động, không có đánh giá tích cực hay tiêu cực.
Dẽ	“nghiên cứu khoa học .”	Cụm danh từ	Nêu lĩnh vực thực hiện, mang tính chất liệt kê chủ đề.
Dẽ	“làm bài tập .”	Cụm động từ	Mô tả hành động thực tế diễn ra trong tiết học, hoàn toàn khách quan.
Dẽ	“điểm .”	Danh từ đơn	Nêu đối tượng quan tâm nhưng không kèm theo thái độ hài lòng hay thất vọng.
Dẽ	“thời gian và cách thức thầy dạy .”	Cụm danh từ	Liệt kê các khía cạnh quan sát được, không có tính từ bổ nghĩa cảm xúc.

Dễ	“dạy với slide tiếng anh .”	Mô tả công cụ	Cung cấp thông tin thực tế về giáo trình, không mang ý khen hay chê.
Dễ	“thuyết trình trên lớp .”	Mô tả hình thức	Xác nhận phương pháp học tập thực tế đang diễn ra tại lớp.
Dễ	“không có .”	Trạng thái trống	Câu trả lời xác nhận không có ý kiến gì thêm, mang tính thông báo sự việc.
Trung bình	"tạm ổn ."	Đánh giá ở mức vừa phải.	Cảm xúc mờ nhạt; mô hình dễ nhầm với Tích cực nhẹ.
Trung bình	"ngoài ra còn nhiều kiến thức ngoài ."	Mô tả thêm thông tin nội dung.	Không khen/chê rõ ràng; cần hiểu ngữ cảnh để không nhầm là Tích cực.
Trung bình	"không ý kiến ."	Trạng thái trung lập tuyệt đối.	Từ phủ định "không" để khiếu nại mô hình nhận diện sai thành Tiêu cực.
Trung bình	“dưới đây là hình thầy cung cấp toàn bộ tài liệu trong quá trình học .”	Chỉ dẫn thông tin	Một câu giới thiệu bằng chứng khách quan, không mang sắc thái biểu cảm rõ rệt.
Trung bình	“không có hoạt động giảng dạy nào không hài lòng .”	Phủ định của phủ định	Cấu trúc khẳng định sự ổn định: không có điểm xấu nhưng cũng không nêu điểm tốt cụ thể.
Khó	"thầy có vẻ hơi bận rộn , gấp thầy có vẻ khó ."	Mô tả tình trạng khách quan.	Dễ bị hiểu là làm lùi phản nàn (Tiêu cực); mô hình cần phân biệt giữa mô tả thực trạng và chê trách.
Khó	"có thể nhiều bạn thích nhưng bản thân mình thấy thầy nói một vấn đề hơi dài dòng ."	Kết hợp Khen ("nhiều bạn thích") và Chê ("dài dòng").	Sentiment pha trộn; yêu cầu mô hình hiểu sự cân bằng giữa hai chiều cảm xúc đối lập.

Khó	"thầy dạy có tâm , nhiệt tình nhưng mà nói hơi nhanh chút xíu ."	Chứa cả Khen ("nhiệt tình") và Góp ý ("hơi nhanh").	Hai hướng cảm xúc đối lập; đòi hỏi mô hình nắm bắt được ngữ nghĩa cân bằng để đưa về nhãn Trung lập.
Khó	"em sẽ nợ môn này , nhưng em sẽ học lại ở các học kỳ kế tiếp ."	Đối lập sự việc	"Nợ môn" nghe có vẻ tiêu cực nhưng "học lại" thể hiện kế hoạch cá nhân; tổng thể là trung tính.
Khó	"ví dụ đê tài của em là doubledot các nguyên lý cơ bản , bao gồm 7 nguyên tắc của bohm (1983) ."	Nội dung chuyên môn	Câu chứa nhiều thuật ngữ, tên riêng, năm xuất bản; mô hình dễ bị nhiễu do cấu trúc quá dài.

2.4.3. Nhãn Tích cực (Positive – Nhãn 2)

Mức độ	Câu ví dụ	Cơ sở phân loại	Giải thích chi tiết
Dễ	"thầy dạy tốt và nhiệt tình."	Từ khóa tích cực rõ ràng: "tốt", "nhiệt tình".	Câu ngắn, khen trực tiếp, không có mâu thuẫn hay ngữ cảnh phức tạp.
Dễ	"Tiết học thú vị, giảng viên dạy hay, rất dễ hiểu."	Ba tín hiệu tích cực trực tiếp: "thú vị", "hay", "dễ hiểu".	Từ khóa cảm xúc rõ, không yêu cầu suy luận ngữ cảnh hay ngữ nghĩa sâu.
Dễ	"nhiệt tình với sinh viên , bài giảng hay ."	Sử dụng tổ hợp từ khóa: "Nhiệt tình" và "hay".	Các từ thể hiện rõ sự tích cực trực tiếp.
Dễ	"thầy đảm bảo phần kiến thức truyền đạt ."	Từ khóa: "đảm bảo".	Thể hiện rõ sự tích cực và tin cậy.
Dễ	"giảng viên giảng bài kỹ ."	Từ khóa: "kỹ".	Từ mang ý nghĩa tích cực rõ ràng về phương pháp.

Dễ	"thân thiện , dễ thương ."	Từ khóa: "thân thiện" và "dễ thương".	Hai từ mang ý nghĩa khen ngợi tính cách.
Dễ	"giảng viên dạy có tâm huyết ."	Từ khóa: "tâm huyết".	Từ mang ý nghĩa tích cực trực tiếp về tinh thần trách nhiệm.
Dễ	"sự tận tình của giảng viên ."	Từ khóa: "sự tận tình".	Từ ngữ mang sắc thái tích cực rõ ràng.
Dễ	"thân thiện , hòa đồng , giảng dạy có kỹ lưỡng ."	Từ khóa: "thân thiện", "hòa đồng", "kỹ lưỡng".	Các từ mang ý nghĩa khen ngợi phong cách và phương pháp.
Dễ	"giảng dạy kỹ ."	Từ khóa: "kỹ".	Thể hiện sự tích cực rõ ràng trong cách truyền đạt.
Dễ	"thầy khá là nhiệt tình trong khi giảng bài , khi sinh viên hỏi bài , giáo viên rất nhiệt tình ."	Từ khóa: "rất nhiệt tình".	Các từ mang ý nghĩa tích cực rõ rệt và lặp lại.
Dễ	"giảng dạy nhiệt tình tâm huyết ."	Từ khóa: "nhiệt tình" và "tâm huyết".	Các từ mang ý nghĩa tích cực trực diện về thái độ.
Trung bình	"Giảng viên nhiệt tình, thân thiện, giảng dạy hiệu quả, dễ tiếp thu, liên tục update bài tập lên Moodle để sinh viên dễ dàng làm bài."	Nhiều yếu tố tích cực kết hợp: "nhiệt tình", "thân thiện", "hiệu quả", "dễ tiếp thu".	Câu dài, chứa nhiều yếu tố khen thuộc các khía cạnh khác nhau; cần mô hình tổng hợp nội dung.
Trung bình	"Thầy dạy rất hay, truyền cảm tốt, phong cách giải toán có một không hai, thiếu buổi nào là bù buổi đó."	Tín hiệu tích cực đa chiều: "rất hay", "truyền cảm", "có một không hai", "bù buổi".	Tích cực đa chiều về cả phong cách và thái độ; mô hình cần hiểu ngữ cảnh hành vi (bù buổi = tận tâm).

Trung bình	"Giảng viên nhiệt tình trong giảng dạy, giảng dạy dễ hiểu, nhiều ví dụ, bài tập để sinh viên thực hành."	Các cụm từ tích cực: "nhiệt tình", "dễ hiểu", "nhiều ví dụ", "thực hành".	Yêu cầu nắm được cấu trúc và quan hệ bổ nghĩa để tổng hợp tín hiệu từ nhiều phần của câu.
Trung bình	"Giảng viên rất nhiệt tình trong việc giảng dạy, luôn dành thời gian trả lời những câu hỏi của sinh viên."	Thái độ tốt: "Rất nhiệt tình" và "luôn dành thời gian trả lời".	Không có từ khóa cảm xúc mạnh ở về sau; mô hình phải suy ra hành vi hỗ trợ sinh viên là tích cực.
Trung bình	"thầy giải nhiều bài tập cho sinh viên giúp sinh viên củng cố kiến thức sau khi học lý thuyết."	Logic về câu: "giúp... củng cố"	Mô tả một hành động có lợi; mô hình phải hiểu hệ quả "củng cố kiến thức" là tích cực.
Trung bình	"thầy có hướng dẫn đồ án rất tận tình, luôn giải đáp các câu hỏi và thắc mắc của sinh viên về đồ án rất nhanh."	Phản hồi nhanh	Sự kết hợp giữa thái độ ("tận tình") và hiệu suất ("nhanh") trong công việc.
Khó	"Phong cách dạy rất hay, rất dễ hiểu, tận tâm, nhiệt tình."	Một loạt các từ tích cực mạnh: "hay", "dễ hiểu", "tận tâm", "nhiệt tình".	Chứa cường độ cảm xúc cao ("rất"); mô hình cần phân biệt mức độ nhấn mạnh.
Khó	"Thầy có cách tiếp cận với sinh viên, thầy dạy và làm rõ những vấn đề sinh viên đang thắc mắc, làm sinh viên có cảm giác thích học thầy."	Mô tả ảnh hưởng tích cực: "làm rõ vấn đề", "thích học thầy".	Câu dài, sentiment ẩn trong ngữ cảnh hành vi; yêu cầu suy luận cảm xúc gián tiếp từ kết quả học tập.
Khó	"Boss cuối, thầy dạy quá hay, cả kiến thức môn học lẫn kiến thức về cuộc sống."	Sắc thái mạnh ("quá hay") và ẩn dụ ("boss cuối").	Chứa ẩn dụ và ngôn ngữ phi chuẩn ("boss cuối"); mô hình phải hiểu văn phong sinh viên để suy ra sự ngưỡng mộ.

Khó	"Giảng viên nhiệt tình, không lạm dụng slide, thầy cho sinh viên tự thảo luận, làm bài rồi mới sửa."	Mô tả phong cách hiện đại: "không lạm dụng slide", "cho thảo luận", "sửa bài".	Gồm nhiều mệnh đề mang cảm xúc gián tiếp, không dùng từ "hay", "tốt"; mô hình phải suy luận ý khen qua hành vi giảng dạy.
Khó	"thầy dạy hay , nhiệt tình , không bao chê bài sinh viên dù mặc dù không tốt mà hướng sinh viên đến suy nghĩ tích cực hơn , có gắng hơn ."	Phủ định của tiêu cực	Sử dụng cụm "không bao chê bài" và "dù mặc dù không tốt" để khen; mô hình dễ nhầm sang tiêu cực vì nhiều từ "không".
Khó	"những điều này nghe tưởng chừng không liên quan đến môn học , nhưng lại giúp cho sinh viên phần nào có định hướng được công việc tương lai của mình ."	Cấu trúc đối lập	Về đầu mang sắc thái hoài nghi ("không liên quan"), nhưng về sau khẳng định giá trị ("giúp định hướng").

CHƯƠNG III: PHƯƠNG PHÁP

3.1. Bối cảnh ra đời

Trong những năm gần đây, phương pháp học chuyển tiếp (Transfer Learning) đã trở thành xương sống của NLP hiện đại. Các mô hình như BERT (Bidirectional Encoder Representations from Transformers) của Google hay RoBERTa của Facebook AI đã đặt ra các tiêu chuẩn mới (SOTA - State of the Art).

Mặc dù các phiên bản đa ngôn ngữ như mBERT (hỗ trợ 104 ngôn ngữ) hay XLM-R có khả năng xử lý tiếng Việt, chúng thường gặp hạn chế do:

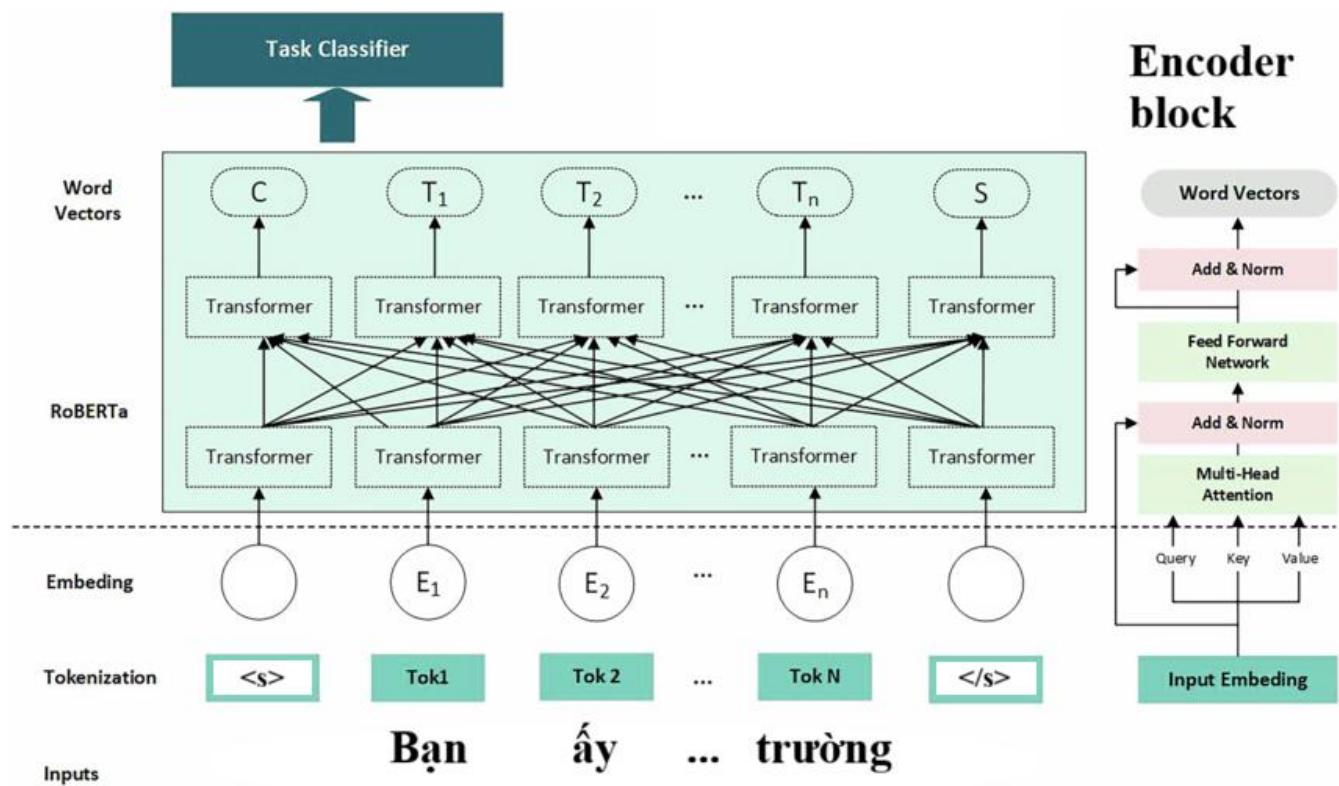
- Dữ liệu huấn luyện bị phân tán: Tiếng Việt chỉ chiếm một phần nhỏ trong kho dữ liệu khổng lồ của các mô hình đa ngôn ngữ.
- Sự khác biệt về ngôn ngữ học: Tiếng Việt là ngôn ngữ đơn lập, ranh giới từ không được xác định bằng khoảng trắng (như tiếng Anh) mà bằng ngữ nghĩa (từ ghép). Việc áp dụng trực tiếp tokenization của tiếng Anh vào tiếng Việt thường dẫn đến việc mô hình học các biểu diễn không tối ưu ở mức âm tiết (syllable) thay vì mức từ (word)

Từ đó PhoBERT ra đời với mục tiêu giải quyết các hạn chế trên bằng cách:

- Tập trung hoàn toàn vào dữ liệu tiếng Việt.
- Áp dụng kỹ thuật tiền xử lý đặc thù cho tiếng Việt (Word Segmentation).
- Cung cấp một mã nguồn mở (Open Source) mạnh mẽ cho cộng đồng nghiên cứu NLP tại Việt Nam.

3.2. Kiến trúc chung

Giới thiệu kiến trúc PhoBERT



PhoBERT là mô hình ngôn ngữ tiếng Việt tiên phong được xây dựng dựa trên kiến trúc tối ưu của RoBERTa (Robustly Optimized BERT Approach). Thay vì chỉ kế thừa cấu trúc, PhoBERT được huấn luyện chuyên sâu (pre-trained) trên một tập dữ liệu khổng lồ bao gồm 20GB văn bản thuần túy từ Vietnamese Wikipedia và Binhvq News Corpus. Nhờ sự kết hợp giữa thuật toán học sâu hiện đại và nguồn dữ liệu bản địa phong phú, PhoBERT kế thừa trọn vẹn sức mạnh xử lý ngữ cảnh của RoBERTa, chính vì vậy, PhoBERT không chỉ mang đầy đủ các đặc tính kiến trúc ưu việt của RoBERTa mà còn được tinh chỉnh để đạt được độ chính xác vượt trội trong các bài toán phân loại văn bản hay trích xuất thông tin.

PhoBERT được sử dụng trong đồ án này mang kiến trúc RobertaForSequenceClassification

*Kiến trúc tổng thể:

```
RobertaForSequenceClassification(  
    (roberta): RobertaModel(  
        (embeddings): RobertaEmbeddings(  
            (word_embeddings): Embedding(64001, 768, padding_idx=1)  
            (position_embeddings): Embedding(258, 768, padding_idx=1)  
            (token_type_embeddings): Embedding(1, 768)  
            (LayerNorm): LayerNorm((768,), eps=1e-05, elementwise_affine=True)  
            (dropout): Dropout(p=0.1, inplace=False)  
        )  
        (encoder): RobertaEncoder(  
            (layer): ModuleList(  
                (0-11): 12 x RobertaLayer(  
                    (attention): RobertaAttention(  
                        (self): RobertaSdpSelfAttention(  
                            (query): Linear(in_features=768, out_features=768, bias=True)  
                            (key): Linear(in_features=768, out_features=768, bias=True)  
                            (value): Linear(in_features=768, out_features=768, bias=True)  
                            (dropout): Dropout(p=0.1, inplace=False)  
                        )  
                        (output): RobertaSelfOutput(  
                            (dense): Linear(in_features=768, out_features=768, bias=True)  
                            (LayerNorm): LayerNorm((768,), eps=1e-05, elementwise_affine=True)  
                            (dropout): Dropout(p=0.1, inplace=False)  
                        )  
                    )  
                    (intermediate): RobertaIntermediate(  
                        (dense): Linear(in_features=768, out_features=3072, bias=True)  
                        (intermediate_act_fn): GELUActivation()  
                    )  
                    (output): RobertaOutput(  
                        (dense): Linear(in_features=3072, out_features=768, bias=True)  
                        (LayerNorm): LayerNorm((768,), eps=1e-05, elementwise_affine=True)  
                        (dropout): Dropout(p=0.1, inplace=False)  
                    )  
                )  
            )  
        )  
        (classifier): RobertaClassificationHead(  
            (dense): Linear(in_features=768, out_features=768, bias=True)  
            (dropout): Dropout(p=0.1, inplace=False)  
            (out_proj): Linear(in_features=768, out_features=3, bias=True)  
        )  
    )
```

RobertaModel là khối tổng hợp nhiều khối nhỏ, vai trò chính của khối là trích xuất đặc trưng ngôn ngữ của dữ liệu đầu vào, dựa trên quan hệ thứ tự từ, ngữ nghĩa, và ngữ nghĩa dài hạn trước khi cho đầu ra để làm đầu vào tiếp theo cho khối *RobertaClassificationHead*.

3.2.1. Lớp RobertaEmbeddings (embeddings)

Kiến trúc transformer không thể hiểu đầy đủ chỉ trên token số rời rạc. Vai trò nền tảng của khối để xử lý dữ liệu đầu vào dạng ma trận số, với mỗi hàng là một câu trong batch processing để ánh xạ câu đã được mã hóa vào không gian vector liên tục để cho mạng học sâu xử lý.

Khối chịu trách nhiệm tổng thể:

- Biến mảng token IDs mang các giá trị số rời rạc trở thành các vector số thực.
- Thể hiện ngữ nghĩa của từng từ, vị trí của từ trong câu, cấu trúc đầu vào tổng thể của câu.

3.2.1.1. Lớp Word Embeddings (word_embeddings)

Lớp Word Embeddings là thành phần đầu tiên và quan trọng nhất trong toàn bộ khối embedding của RoBERTa. Nhiệm vụ cốt lõi của lớp này là chuyển đổi các token rời rạc (token IDs) – vốn chỉ là các chỉ số nguyên không mang ý nghĩa toán học – thành các vector liên tục trong không gian embedding nhiều chiều, nơi mỗi chiều mã hóa một khía cạnh ngữ nghĩa hoặc cú pháp của từ.

Cụ thể, với mỗi token ID đầu vào, mô hình thực hiện một phép tra cứu bảng (embedding lookup) trong ma trận embedding kích thước:

$$V \times d$$

trong đó:

- V là kích thước từ điển (vocabulary size).
- d là chiều embedding của mô hình, tức là 768 đối với PhoBERT.

Kết quả là mỗi token được ánh xạ thành một vector d -chiều, đóng vai trò như biểu diễn ngữ nghĩa ban đầu của từ trước khi đi vào các tầng Encoder.

*Đặc điểm ngữ nghĩa và không gian vector

- Các token có ngữ nghĩa gần nhau (ví dụ: “giáo viên” – “giảng viên”) sẽ có vector gần nhau trong không gian embedding.
- Các token đa nghĩa (polysemy) không được phân biệt hoàn toàn ở bước này; embedding chỉ đóng vai trò là điểm khởi đầu ngữ nghĩa,

còn việc phân giải nghĩa cụ thể sẽ được đảm nhiệm ở các tầng Encoder thông qua cơ chế Self-Attention theo ngữ cảnh.

- Như vậy, Word Embedding trong RoBERTa không phải là biểu diễn “cuối cùng” của nghĩa từ, mà là biểu diễn ngữ nghĩa tĩnh ban đầu (context-independent embedding).

*Từ điển và tokenization trong PhoBERT

PhoBERT sử dụng từ điển có 64.001 token, được xây dựng dựa trên Byte-Pair Encoding (BPE). Cách tiếp cận này mang lại các lợi ích quan trọng:

- Hạn chế gần như hoàn toàn hiện tượng Out-Of-Vocabulary (OOV) trong tiếng Việt – một ngôn ngữ giàu biến thể hình thái và cách viết.
- Cho phép mô hình biểu diễn linh hoạt các từ mới, từ hiếm, từ mượn hoặc từ ghép dài bằng cách chia nhỏ thành các subword.
- Tăng khả năng tổng quát hóa, đặc biệt trong các tác vụ NLP thực tế như phân loại cảm xúc, trích xuất thực thể hoặc QA.

=> Tóm lại, Word Embeddings là cầu nối giữa ngôn ngữ tự nhiên và không gian toán học, đặt nền móng cho toàn bộ quá trình hiểu ngữ nghĩa sâu của RoBERTa.

3.2.1.2. Lớp Position Embeddings (position_embeddings)

Không giống như RNN hay CNN, kiến trúc Transformer không có khái niệm thứ tự tuần tự nội tại. Vì vậy, nếu chỉ sử dụng Word Embeddings, mô hình sẽ không thể phân biệt các câu có cùng tập từ nhưng khác trật tự. Đây chính là lý do lớp Position Embeddings được đưa vào.

* Vai trò cốt lõi

Lớp Position Embeddings chịu trách nhiệm mã hóa thông tin vị trí của từng token trong chuỗi, bằng cách ánh xạ mỗi vị trí (0, 1, 2, ..., max_length) thành một vector embedding cùng chiều với word embedding.

Embedding vị trí này:

- Được học trong quá trình huấn luyện (learned positional embeddings), không phải dạng hàm sin/cos cố định.
- Được cộng trực tiếp với Word Embeddings, tạo ra biểu diễn tổng hợp:

$$E_{final} = E_{word} + E_{position} + E_{token\ type}$$

* Ý nghĩa ngữ nghĩa và cú pháp

Nhờ Position Embeddings, mô hình có thể:

Phân biệt các cấu trúc cú pháp khác nhau dù cùng từ vựng.

Hiểu rõ mối quan hệ thứ tự, ví dụ:

- “tôi yêu bạn”
- “bạn yêu tôi”
- Dù dùng cùng các token, nhưng khác hoàn toàn về ý nghĩa.

Hỗ trợ Encoder học được các quan hệ xa-gần, phụ thuộc dài hạn (long-range dependencies) trong câu.

* **Tầm quan trọng trong tiếng Việt**

Trong tiếng Việt – nơi trật tự từ đóng vai trò lớn trong việc biểu đạt ý nghĩa – Position Embeddings càng trở nên quan trọng, vì chỉ cần thay đổi vị trí từ là ngữ nghĩa câu có thể đảo chiều hoàn toàn.

3.2.1.3. Lớp Token type (token_type_embeddings)

Token Type Embeddings (hay Segment Embeddings) ban đầu được thiết kế trong BERT nhằm phân biệt các đoạn văn khác nhau (ví dụ câu A và câu B trong bài toán Next Sentence Prediction).

* **Trong RoBERTa và PhoBERT:**

- RoBERTa loại bỏ hoàn toàn nhiệm vụ Next Sentence Prediction (NSP).
- Trong thực tế huấn luyện, tất cả token đều được gán token_type_id = 0.
- Lớp Token Type Embedding vẫn tồn tại trong kiến trúc để:
 - Giữ tính tương thích với BERT API.
 - Hỗ trợ tiềm năng mở rộng cho các bài toán đặc thù nếu cần.

* **Lý do thiết kế:**

Các nghiên cứu của RoBERTa cho thấy:

Việc phân biệt segment bằng Token Type Embedding không mang lại lợi ích rõ ràng cho chất lượng biểu diễn ngôn ngữ.

Loại bỏ NSP và đơn giản hóa segment giúp mô hình:

- Tập trung hoàn toàn vào Masked Language Modeling (MLM).
- Học ngữ nghĩa sâu hơn từ ngữ cảnh liên tục, thay vì các giả định nhân tạo về ranh giới câu.

Do đó, Token Type Embeddings trong RoBERTa mang tính hình thức hơn là chức năng thực tế, nhưng vẫn là một phần của tổng embedding.

3.2.1.4. Lớp LayerNorm (LayerNorm)

Sau khi cộng các embedding thành phần, vector embedding tổng hợp được đưa qua Layer Normalization.

* Vai trò kỹ thuật

LayerNorm thực hiện:

Chuẩn hóa từng vector embedding theo chiều đặc trưng:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

Giữ cho phân phối giá trị ổn định trong suốt quá trình lan truyền qua nhiều tầng Transformer.

* Lợi ích chính:

- Ổn định gradient, đặc biệt quan trọng với mô hình sâu như RoBERTa/PhoBERT (12 Encoder).
- Tăng tốc hội tụ khi huấn luyện.
- Giảm hiện tượng exploding/vanishing gradients.

LayerNorm đảm bảo rằng embedding đầu vào của Encoder có thang đo phù hợp, giúp các cơ chế Self-Attention hoạt động hiệu quả và ổn định.

3.2.1.5. Lớp Dropout (Dropout)

Dropout là bước cuối cùng trong khối embedding trước khi dữ liệu đi vào Encoder.

* Cơ chế

- Ngẫu nhiên “tắt” một tỷ lệ nhỏ các chiều trong embedding vector trong quá trình huấn luyện.
- Không áp dụng trong pha inference.

* Vai trò

- Ngăn mô hình học quá phụ thuộc vào một số đặc trưng embedding cụ thể.
- Giảm hiện tượng overfitting, đặc biệt khi fine-tune trên tập dữ liệu nhỏ.

- Tăng khả năng tổng quát hóa của biểu diễn embedding.

Trong RoBERTa, Dropout được xem là một biện pháp regularization nhẹ nhưng hiệu quả, giúp embedding không trở nên quá khắt khe trước khi đi qua các tầng Attention phức tạp phía sau.

3.2.2. Khối RobertaEncoder (encoder)

Khối RoBERTaEncoder là trái tim của toàn bộ mô hình RoBERTa. Nếu khối Embedding chỉ mới tạo ra các vector mang nghĩa từ vựng ban đầu, thì Encoder chính là nơi ngữ nghĩa thực sự được hình thành thông qua ngữ cảnh.

Nói một cách khái quát, Encoder thực hiện quá trình:

- Biến “vector mang nghĩa của từng từ độc lập” thành “vector mang nghĩa của từ trong toàn bộ ngữ cảnh câu / đoạn văn”.
- Encoder gồm 12 layer chồng lên nhau.

Mỗi layer:

- Không làm thay đổi kích thước vector (ví dụ: luôn giữ 768 chiều).
- Chỉ làm thay đổi nội dung thông tin bên trong vector.

Mỗi layer Encoder giúp:

- Làm giàu ngữ nghĩa.
- Tăng mức độ trừu tượng.
- Mở rộng phạm vi ngữ cảnh mà mỗi token “nhìn thấy”.

3.2.2.1. Lớp RobertaAttention (attention)

Attention là cơ chế cốt lõi giúp Transformer – và do đó là RoBERTa – vượt trội so với các kiến trúc tuần tự truyền thống.

Attention cho phép mô hình:

- Xem xét toàn bộ các token khác trong chuỗi khi biểu diễn một token cụ thể.
- Học được:
 - Token nào quan trọng hơn.
 - Mức độ quan trọng là bao nhiêu.
 - Quan trọng trong ngữ cảnh hiện tại, không cố định.

Thay vì xử lý theo thứ tự như RNN, Attention xử lý song song, cho phép mô hình nắm bắt:

- Quan hệ xa (long-range dependencies).
- Quan hệ ngữ nghĩa phức tạp giữa các token không kèn nhau.
- Attention là cơ chế học quan hệ giữa token với token, dưới góc nhìn của từng token một, trong ngữ cảnh toàn cục.

3.2.2.1.1. Lớp RobertaSdpSelfAttention

Đây là khái niệm tính toán Self-Attention thực sự, nơi diễn ra phép toán Attention cốt lõi của mô hình.

Self-Attention nghĩa là:

- Query, Key, Value đều được sinh ra từ cùng một tập hidden states đầu vào.
- Mỗi token vừa là “nguồn truy vấn”, vừa là “nguồn để truy vấn thông tin” cho token khác.

RoBERTa sử dụng Scaled Dot-Product Attention, với các bước tuyến tính (Linear) để tạo ra Q, K, V.

3.2.2.1.1.1. Linear (query)

Lớp Linear (Query) có nhiệm vụ:

- Biến hidden state của mỗi token thành một Query vector.
- Query đại diện cho: “Token này đang tìm kiếm loại thông tin gì từ các token khác trong ngữ cảnh?”

Ví dụ:

- Với token “nó”, Query có thể tìm thông tin về chủ thể mà nó ám chỉ.
- Với token “rất”, Query có thể tìm thông tin về từ mà nó bô nghĩa.

* Đặc điểm kỹ thuật

Là một phép biến đổi tuyến tính:

$$Q = XW_Q$$

Không làm thay đổi chiều (vẫn là 768).

Chỉ thay đổi cách nhìn của token đối với ngữ cảnh.

3.2.2.1.1.2. Linear (key)

Lớp Linear (Key) biến hidden state thành Key vector.

* Vai trò trực giác

Key trả lời cho câu hỏi: “Token này có thông tin gì mà token khác có thể cần?”

Mỗi token “treo” thông tin của mình thông qua Key, để các Query từ token khác so khớp.

* Quan hệ với Query

Query và Key nằm trong cùng không gian vector (768 chiều).

Điều này cho phép thực hiện dot-product giữa Query của token i và Key của token j:

- Đo mức độ tương đồng.
- Mức độ “token i có quan tâm token j hay không”.

Công thức:

$$K = XW_K$$

3.2.2.1.1.3. Linear (value)

Lớp Linear (Value) tạo ra Value vector, đại diện cho nội dung thông tin thực sự sẽ được tổng hợp.

*Vai trò cốt lõi

- Key dùng để so khớp.
- Value dùng để truyền tải nội dung.

$$V = XW_V$$

*Công thức cốt lõi:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

Trong đó:

- QK^T : Đo độ tương đồng giữa mỗi Query với mọi Key.
- $\sqrt{d_k}$: Hệ số chuẩn hóa, tránh gradient quá lớn
- softmax : Để tạo phân phối chú ý, tổng trọng số chú ý trên toàn bộ token bằng 1.
- Nhân với V: Tổng hợp thông tin từ các token khác. Mỗi Value được lấy theo trọng số chú ý tương ứng.

3.2.2.1.4. Dropout (dropout)

Ngăn mô hình quá phụ thuộc vào một token cụ thể.

Dropout ở đây đóng vai trò regularization trực tiếp lên quan hệ giữa các token, chứ không chỉ trên vector đặc trưng.

3.2.2.1.2. Lớp RobertaSelfOutput (output)

Sau khi khôi Self-Attention hoàn tất việc tổng hợp thông tin ngữ cảnh, biểu diễn của mỗi token tuy đã “nhìn thấy” toàn bộ câu, nhưng chưa được sử dụng trực tiếp cho layer tiếp theo. Thay vào đó, kết quả Attention phải đi qua khôi Output, bao gồm Feed Forward Network (FFN) và các cơ chế ổn định huấn luyện.

Khôi này đóng vai trò: Chuyển hóa thông tin ngữ cảnh đã được trộn thành biểu diễn phi tuyến giàu ngữ nghĩa hơn, đồng thời giữ cho kiến trúc sâu hoạt động ổn định.

Trong mỗi Encoder layer của RoBERTa, FFN được áp dụng độc lập cho từng token, nhưng với cùng một tập trọng số, giúp mô hình học các phép biến đổi phức tạp trên không gian biểu diễn.

3.2.2.2. Intermediate (intermediate)

Khôi Intermediate là tầng mở rộng đầu tiên của Feed Forward Network, nơi mô hình:

- Tăng mạnh khả năng biểu diễn.
- Áp dụng biến đổi phi tuyến để học các tổ hợp đặc trưng phức tạp.

Nếu Attention chịu trách nhiệm trộn thông tin giữa các token, thì Intermediate chịu trách nhiệm:

- Xử lý sâu từng token sau khi đã được ngữ cảnh hóa.

Về mặt kỹ thuật, khôi Intermediate thực hiện:

$$H_{intermediate} = GELU(XW_{intermediate} + b_{intermediate})$$

Trong đó:

- X có kích thước 768.
- W mở rộng chiều từ 768 \rightarrow 3072 (gấp 4 lần).

Việc mở rộng chiều này mang ý nghĩa quan trọng:

- Tạo ra không gian biểu diễn lớn để mô hình học:
 - Các tương tác phi tuyến.
 - Các đặc trưng ngữ nghĩa tinh vi.

- Giảm hiện tượng “nút thắt cổ chai” (bottleneck) về biểu diễn.

3.2.2.3. RobertaOutput (output)

Sau khi qua Intermediate, biểu diễn token có kích thước 3072 sẽ được nén trở lại 768 chiều thông qua một lớp Linear thứ hai:

$$H_{output} = H_{intermediate} W_{output} + b_{output}$$

Việc nén này nhằm:

- Giữ kích thước thông nhất giữa các Encoder layer.
- Cho phép chồng nhiều layer mà không thay đổi cấu trúc tổng thể.

Sau đó áp dụng LayerNorm và Dropout giúp mô hình sâu vẫn học hiệu quả.

3.2.3. Khối RobertaClassificationHead (classifier)

Khối RoBERTaClassificationHead là phần head chuyên biệt cho bài toán phân loại trong kiến trúc RoBERTa. Khối này được gắn trực tiếp lên trên backbone RoBERTaEncoder, với nhiệm vụ chuyển đổi biểu diễn ngữ nghĩa tổng quát của câu thành logits cho các lớp đầu ra tương ứng với bài toán cụ thể (ví dụ: phân loại cảm xúc, phân loại chủ đề, đánh giá mức độ hài lòng,...).

Classification Head không học ngôn ngữ từ đầu, mà chỉ học cách đưa ra quyết định phân loại dựa trên biểu diễn ngữ nghĩa đã được học sẵn bởi RoBERTa trong giai đoạn pre-training.

Đầu vào của Classification Head thường là:

- Vector 768 chiều của token đặc biệt < s > ở layer Encoder cuối cùng.
- Vector này được xem như biểu diễn ngữ nghĩa toàn cục của câu hoặc đoạn văn, vì trong quá trình Self-Attention, token < s > đã tương tác với toàn bộ các token khác.

Do đó, vector < s > đóng vai trò như một bản tóm tắt ngữ nghĩa cô đọng của toàn bộ chuỗi đầu vào.

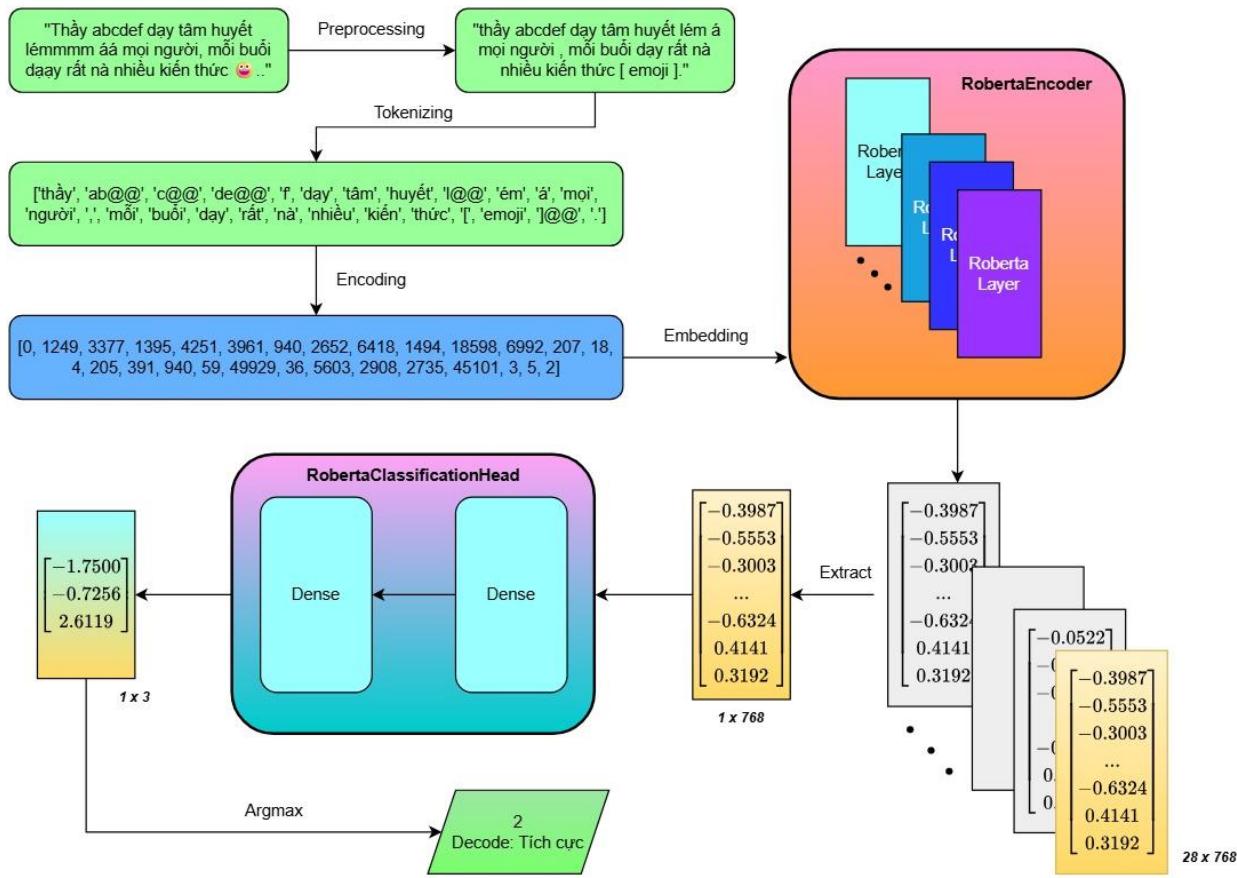
* Vai trò của lớp Dense trung gian

Điều chỉnh biểu diễn ngữ nghĩa tổng quát thành biểu diễn phù hợp với nhiệm vụ cụ thể.

Cho phép mô hình:

- Tái tổ chức thông tin ngữ nghĩa.
- Nhấn mạnh những đặc trưng quan trọng cho phân loại

3.3. Ví dụ cụ thể



Để có thể hiểu rõ hơn về cách hoạt động của từng khối trong kiến trúc PhoBERT-base, hình minh họa trên hình thể hiện một pipeline được sử dụng trong đồ án với một ví dụ cụ thể để giải bài toán phân loại cảm xúc văn bản tiếng Việt.

Toàn bộ quy trình có thể được mô tả như một chuỗi chuyển đổi từ dữ liệu thô sang kết quả dự đoán, gồm các khối chính: Preprocessing → Tokenizing → Encoding → Embedding → RoBERTaEncoder → Extract → RobertaClassificationHead → Logits → Argmax/Decode. Mỗi khối đảm nhiệm một vai trò riêng và đóng góp vào việc biến câu tiếng Việt thành vector số, học ngữ cảnh, rồi quy đổi thành nhãn cảm xúc.

Trong ví dụ minh họa, câu đầu vào mang nội dung khen ngợi giáo viên: "Thầy abcdef dạy tâm huyết lémmmm áá mọi người, mỗi buổi dạy rất nà nhiều kiến thức 😊 ."

Đây là một câu tiếng Việt không chuẩn hóa hoàn toàn, chứa hiện tượng kéo dài ký tự, emoji và dấu câu không nhất quán—những đặc điểm phổ biến trong dữ liệu văn bản do người dùng tạo (user-generated content).

3.3.1. Tiền xử lý văn bản

Bước đầu tiên trong pipeline là tiền xử lý nhằm giảm nhiễu và đưa văn bản về dạng chuẩn hóa. Các thao tác thường bao gồm:

- Chuẩn hóa các ký tự in hoa thành ký tự thường.
- Chuẩn hóa dấu câu.
- Thay thế emoji “” bằng flag “[emoji]”.
- Rút gọn các ký tự kéo dài (“lémmmm” → “lém”).
- Chuẩn hóa khoảng trắng.

Sau bước này, câu được chuyển thành:

“thầy abcdef dạy tâm huyết lém á mọi người , mỗi buổi dạy rất nà nhiều kiến thức [emoji].”

3.3.2. Tokenization với subword (BPE – Byte Pair Encoding)

Văn bản sau preprocessing được đưa qua bộ tokenizer của PhoBERT, vốn dựa trên cơ chế BPE tương tự RoBERTa. Khác với tokenization theo từ hoàn chỉnh, BPE phân tách câu thành các đơn vị con (subword), cho phép mô hình xử lý hiệu quả các từ hiếm hoặc chưa từng xuất hiện trong tập huấn luyện.

Kết quả tokenization có dạng mảng gồm 26 phần tử:

`['thầy', 'ab@@@', 'c@@@', 'de@@@', 'f', 'dạy', 'tâm', 'huyết', 'l@@@', 'ém', 'á', 'mọi', 'người', ',', 'mỗi', 'buổi', 'dạy', 'rất', 'nà', 'nhiều', 'kiến', 'thức', '[', 'emoji', ']@@@', '.']`

Chuỗi “abcdef” bị tách thành các mảnh nhỏ như “ab@@@”, “c@@@”, “de@@@”, “f”. Đây là hành vi chuẩn của BPE đối với các chuỗi hiếm hoặc không mang ý nghĩa từ vựng rõ ràng. Ưu điểm cốt lõi của subword tokenization là giúp mô hình tránh được vấn đề out-of-vocabulary (OOV), đồng thời vẫn xây dựng được biểu diễn cho từ mới dựa trên các thành phần cấu tạo.

Tokenizer đã giữ lại các từ tiếng Việt quan trọng như “thầy”, “dạy”, “tận”, “tâm”, “huyết”, “kiến”, “thức”, ... cho thấy PretrainedTokenizer đã được huấn luyện trên dữ liệu tiếng Việt chứa lượng từ vựng đủ để phân tách câu một cách hợp lý cho cho xử lý ngôn ngữ tiếng Việt.

3.3.3. Encoding sang token IDs

Sau quá trình tokenize, mỗi token được ánh xạ sang một số nguyên (token ID) dựa trên từ điển cố định của PhoBERT: Kết quả sau bước encoding thu được mảng 28 phần tử:

$[0, 1249, 3377, 1395, 4251, 3961, 940, 2652, 6418, 1494, 18598, 6992, 207, 18, 4, 205, 391, 940, 59, 49929, 36, 5603, 2908, 2735, 45101, 3, 5, 2]$

Bước encoding chuyển dữ liệu từ dạng ký hiệu rời rạc (symbolic representation) sang dạng số học, là điều kiện tiên quyết để mô hình học sâu có thể xử lý.

3.3.4. Embedding layer

Mảng token IDs sẽ đi qua Embeddings block, tại đây từ mảng số sẽ được mã hóa, biến các số rời rạc thành vector ngữ nghĩa liên tục, đồng thời cũng mã hóa thứ tự của token, tiếp đó sẽ đi qua lớp chuẩn hóa và thu được tensor 3 chiều, trong đó chiều đầu tiên là batch, tensor 2 chiều X chính là embeddings của câu đầu vào:

$$X \in \mathbb{R}^{28 \times 768}$$

Ở giai đoạn này, embedding chủ yếu mã hóa thông tin từ vựng và hình thái của token một cách độc lập. Ví dụ, embedding của “thày” khác với “cô”, “tâm” khác với “tệ”. Tuy nhiên, embedding đơn lẻ chưa đủ để suy luận cảm xúc ở cấp độ câu, vì ý nghĩa thực sự phụ thuộc mạnh vào ngữ cảnh.

3.3.5. RoBERTa Encoder và Self-Attention

Embedding đầu vào được đưa qua khối RoBERTa Encoder, bao gồm 12 layer Encoder xếp chồng nối tiếp. Mỗi layer sử dụng cơ chế self-attention đa đầu (multi-head self-attention) để cập nhật biểu diễn của từng token dựa trên toàn bộ câu.

Qua self-attention, khi mô hình xử lý token “huyết” trong cụm “tâm huyết”, nó có thể phân bổ trọng số chú ý không chỉ cho token “tâm” liền kề, mà còn cho các token liên quan về mặt ngữ nghĩa như “dạy”, “thày”, “kiến thức” hoặc thậm chí token biểu diễn emoji. Nhờ đó, vector của mỗi token trở thành contextual embedding, phản ánh vai trò và ý nghĩa của token trong ngữ cảnh tổng thể.

Các layer encoder ở độ sâu khác nhau học các mức trừu tượng khác nhau: từ quan hệ cục bộ và hình thái (layer nông), đến cấu trúc cụm từ và quan hệ cú pháp (layer trung gian), và cuối cùng là ý nghĩa tổng quát phục vụ cho nhiệm vụ phân loại (layer sâu).

Việc encoder có nhiều layer giúp mô hình học biểu diễn theo độ sâu. Các layer đầu thường học quan hệ cục bộ và hình thái từ; các layer giữa học cấu trúc cụm từ; layer sâu học nghĩa tổng quát và tín hiệu cho downstream task. Đầu ra cuối cùng của encoder, trừ chiều batch, là tensor 2 chiều:

$$H \in \mathbb{R}^{28 \times 768}$$

3.3.6. Biểu diễn cấp độ câu và Classification Head

Do bài toán sentiment là phân loại ở cấp độ câu, vậy nên mô hình rút gọn tensor kích thước 28×768 bằng cách trích xuất duy nhất vector đại diện cho toàn bộ câu, chính là vector của token đặc biệt ở đầu câu (tương đương token [CLS] trong kiến trúc BERT), vốn được thiết kế để tổng hợp thông tin ngữ cảnh toàn câu.

Vector này được đưa vào RobertaClassificationHead, bao gồm các tầng fully connected (Dense). Head này đóng vai trò ánh xạ biểu diễn ngữ nghĩa giàu thông tin từ encoder sang không gian nhãn của bài toán.

Tầng Dense cuối cùng sinh ra vector logits kích thước 1×3 :

$$[-1.7500, -0.7256, 2.6119]$$

3.3.7. Dự đoán và giải mã nhãn

Tương ứng với ba lớp cảm xúc. Logits là các giá trị chưa chuẩn hóa; lớp có giá trị lớn nhất được xem là dự đoán của mô hình.

Sau khi áp dụng phép argmax lên vector logits, mô hình dự đoán lớp có chỉ số 2, và bước decode ánh xạ chỉ số này thành nhãn cảm xúc “Tích cực”.

Kết quả này phù hợp với nội dung câu, vốn chứa nhiều tín hiệu khen ngợi như “tâm huyết”, “nhiều kiến thức” cùng với emoji mang sắc thái tích cực. Điều này cho thấy pipeline đã khai thác hiệu quả các đặc trưng ngữ nghĩa và cảm xúc, đồng thời khẳng định khả năng của PhoBERT/RoBERTa trong việc mô hình hóa ngôn ngữ tiếng Việt cho bài toán phân loại cảm xúc.

CHƯƠNG IV: CÀI ĐẶT VÀ THỬ NGHIỆM

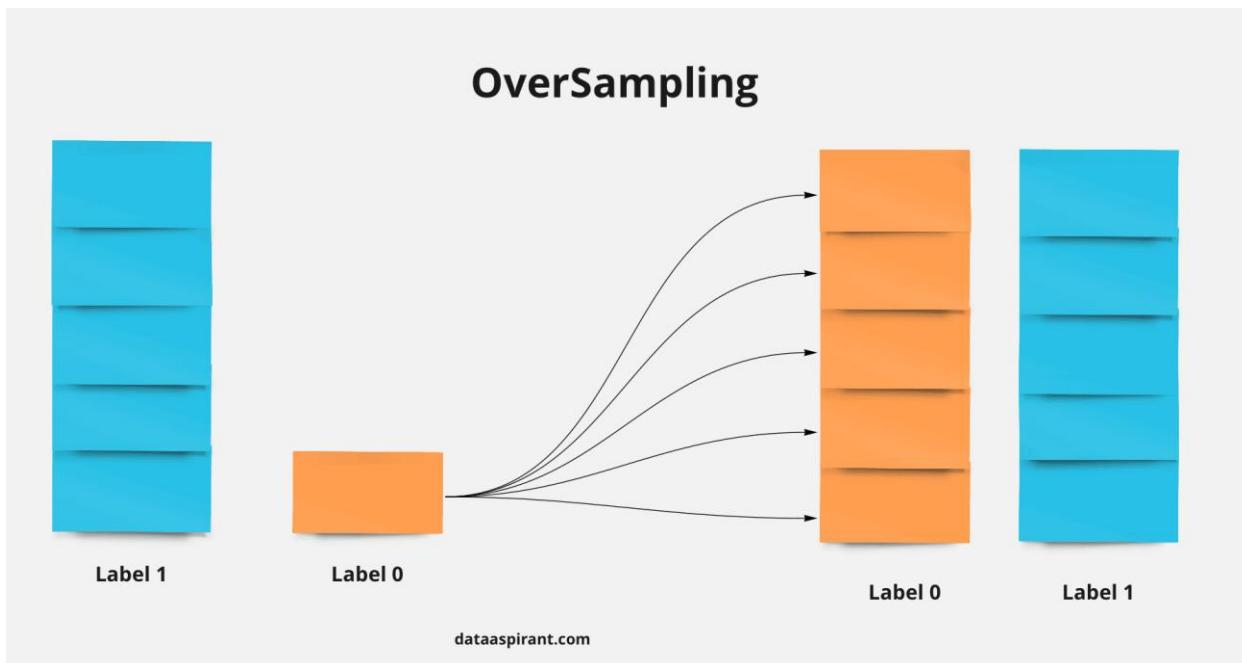
4.1. Chia tập train và xử lý mất cân bằng dữ liệu

Đầu tiên nhóm nghiên cứu sẽ chia bộ dữ liệu ban đầu thành 3 tập train, validation và tập test với tỉ lệ lần lượt là 70/10/20 với mục tiêu là tránh tình trạng rò rỉ dữ liệu và đánh giá mô hình một cách chính xác nhất.

Ngoài ra như đã đề cập ở phần đầu cuốn báo cáo, bộ dữ liệu nhóm sử dụng có sự mất cân bằng trầm trọng đối với nhãn Neutral so với 2 nhãn còn lại. Chính vì thế nhóm quyết định sử dụng kỹ thuật oversampling để cân bằng bộ dữ liệu

Oversampling là một kỹ thuật xử lý mất cân bằng dữ liệu (class imbalance) trong bài toán phân loại, bằng cách tăng số lượng mẫu của lớp thiểu số để phân bổ các nhãn trở nên cân bằng hơn trong quá trình huấn luyện. Khi tập dữ liệu bị lệch nhãn mô hình thường có xu hướng “thiên vị” dự đoán về lớp chiếm đa số, dẫn đến việc dự đoán kém cho các lớp ít mẫu. Oversampling giúp mô hình được “nhìn thấy” lớp thiểu số nhiều hơn, từ đó cải thiện khả năng học đặc trưng và tăng các chỉ số như recall/F1-score ở lớp thiểu số.

Và với nội dung của bài toán này thì nhóm đã sử dụng kỹ thuật Random Sampling, tức là sao chép mẫu Neutral và Negative cho đến khi số lượng bằng với mẫu Positive.



4.2. Tham số huấn luyện

Sau khi đã chuẩn bị bộ dữ liệu hoàn tất thì nhóm sẽ tiến hành thiết lập các tham số huấn luyện như sau:

Epochs	5
Learning_rate	1e-5
Scheduler_type	Cosine
Warm_up_ratio	0.03
Weight_decay	0
Train_batch_size	32
Grad_accum_steps	1
Test_batch_size	128
Dropout	0.1
Early_stopping_patience	3

Trong quá trình fine-tune mô hình vinai/phobert-base cho bài toán phân loại cảm xúc, nhóm sử dụng một bộ tham số huấn luyện nhằm cân bằng giữa khả năng hội tụ, độ ổn định và hạn chế overfitting. Do PhoBERT-base là mô hình Transformer đã được tiền huấn luyện sẵn, các hyperparameter khi fine-tune thường cần lựa chọn “nhẹ tay” hơn so với mô hình huấn luyện từ đầu. Vì vậy, việc thiết lập learning rate nhỏ và kết hợp scheduler là yếu tố quan trọng để mô hình tiếp thu tri thức mới từ tập dữ liệu mới nhưng vẫn giữ được biểu diễn ngôn ngữ đã học trong giai đoạn pretraining.

Trước hết, nhóm thiết lập epochs = 5, tức là toàn bộ tập huấn luyện được lặp lại 5 lần. Với các mô hình ngôn ngữ tiền huấn luyện như PhoBERT, số epoch quá lớn có thể khiến mô hình bị overfit, đặc biệt khi dữ liệu huấn luyện không quá lớn hoặc có nhiễu. Việc chọn 5 epoch là mức tương đối phổ biến trong fine-tuning, đủ để mô hình học được các đặc trưng đặc thù của bài toán sentiment, nhưng vẫn không làm mô hình “học thuộc” dữ liệu. Đồng thời, nhóm kết hợp thêm cơ chế early stopping patience = 3 để tăng tính an toàn trong trường hợp mô hình không còn cải thiện trên tập validation. Cụ thể, nếu mô hình không cải thiện sau 3 epoch liên tiếp, quá trình huấn luyện sẽ dừng sớm nhằm tránh overfitting và tiết kiệm thời gian tính toán.

Tiếp theo, learning rate = 1e-5 là một mức learning rate nhỏ và phù hợp cho fine-tune Transformer. Đây là một lựa chọn khá “an toàn” bởi nếu learning rate quá lớn (ví dụ 3e-5 hoặc 5e-5 trong một số tình huống) mô hình có thể bị mất ổn định, gradient cập nhật quá mạnh làm giảm chất lượng biểu diễn ngôn ngữ đã tiền huấn luyện. Với bài toán phân loại cảm xúc, learning rate 1e-5 thường giúp mô

hình điều chỉnh dần dần để phù hợp dữ liệu mà vẫn giữ được khả năng hiểu ngữ cảnh tiếng Việt.

Nhóm áp dụng scheduler type = cosine, tức là điều chỉnh learning rate theo dạng cosine decay. Cơ chế này giúp learning rate giảm dần theo tiến trình huấn luyện, trong đó giai đoạn đầu learning rate tương đối cao hơn để mô hình học nhanh, và giai đoạn cuối learning rate giảm để mô hình tinh chỉnh tham số ổn định hơn. Cosine scheduler thường được dùng phổ biến cho các mô hình Transformer vì giúp tránh dao động lớn ở cuối quá trình fine-tune, từ đó cải thiện khả năng hội tụ và tăng độ ổn định của loss/metric.

Bên cạnh đó, tham số warm up ratio = 0.03 nghĩa là trong khoảng 3% số bước đầu tiên của quá trình training, learning rate sẽ được tăng dần từ rất nhỏ lên mức learning rate tối đa. Warmup đóng vai trò quan trọng đối với Transformer vì ở giai đoạn đầu, nếu learning rate nhảy lên quá nhanh thì mô hình có thể cập nhật gradient mạnh dẫn đến mất ổn định, đặc biệt khi head phân loại còn chưa “thích nghi” với dữ liệu mới.

Đối với weight decay = 0, tức là nhóm không áp dụng regularization dạng L2 weight decay lên tham số mô hình. Trong một số bài toán, weight decay giúp giảm overfitting bằng cách hạn chế tham số trở nên quá lớn. Tuy nhiên, với fine-tuning Transformer, weight decay đôi khi cần được cân nhắc theo kích thước dữ liệu và mục tiêu tối ưu. Việc đặt weight decay = 0 có thể được hiểu là nhóm ưu tiên giữ nguyên độ linh hoạt của mô hình khi học dữ liệu downstream. Dù vậy, trong các nghiên cứu mở rộng, có thể thử thêm weight decay nhỏ (ví dụ 0.01) để đánh giá tác động lên khả năng tổng quát hóa.

Về batch size, nhóm sử dụng train batch size = 32 và test batch size = 128. Trong đó batch size huấn luyện 32 là một mức phù hợp để cân bằng giữa hiệu quả học và giới hạn bộ nhớ GPU. Batch size càng lớn có thể giúp gradient ổn định hơn, tuy nhiên lại tiêu tốn nhiều tài nguyên hơn. Batch size cho tập test lớn hơn (128) là hợp lý vì inference không cần lưu gradient, do đó có thể tăng batch size để đánh giá nhanh hơn. Ngoài ra, grad accum steps = 1 cho thấy nhóm không sử dụng gradient accumulation; tức là mỗi batch huấn luyện sẽ cập nhật tham số một lần.

Cuối cùng, nhóm đặt dropout = 0.1, đây là giá trị dropout phổ biến trong kiến trúc BERT/RoBERTa. Dropout giúp giảm overfitting bằng cách “tắt ngẫu nhiên” một phần neuron trong quá trình huấn luyện, buộc mô hình không phụ thuộc quá nhiều vào một vài đặc trưng cụ thể. Với bài toán sentiment và dataset

không quá lớn, dropout 0.1 thường tạo sự cân bằng tốt giữa học hiệu quả và tổng quát hóa.

Tổng hợp lại, bộ hyperparameter trên thể hiện hướng thiết lập tập trung vào việc fine-tune ổn định cho PhoBERT-base với learning rate nhỏ, scheduler cosine, warmup nhẹ và cơ chế early stopping để tránh overfitting. Đây là cấu hình phù hợp cho bài toán phân loại cảm xúc tiếng Việt, đồng thời có thể làm nền tảng cho các thử nghiệm mở rộng như điều chỉnh weight decay, tăng epochs hoặc áp dụng gradient accumulation khi mở rộng dữ liệu trong các hướng phát triển tiếp theo.

CHƯƠNG V: KẾT QUẢ ĐẠT ĐƯỢC

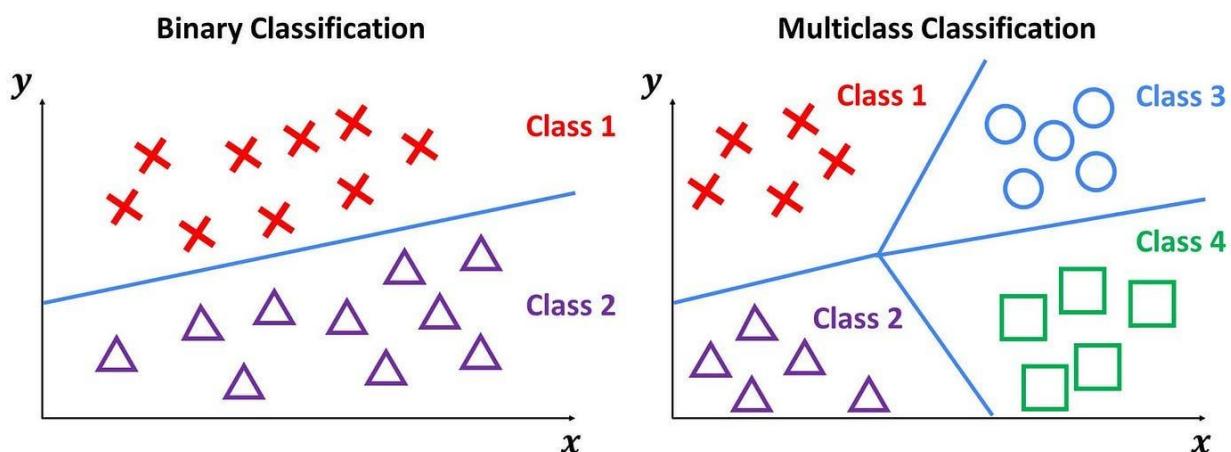
5.1. Kết quả

Ngoài việc huấn luyện mô hình PhoBERT-base, nhóm cũng đã tiến hành sử dụng các mô hình máy học truyền thống (Logistic Regression, SVM, XGBoost) kết hợp với kỹ thuật embedding TF-IDF để có thể đánh giá một cách chính xác hơn về mức độ hiệu quả của PhoBERT-base. Việc lựa chọn các mô hình truyền thống này nhằm tạo ra một hệ quy chiếu (baseline) phổ biến trong bài toán phân loại văn bản, từ đó giúp kiểm tra xem mô hình ngôn ngữ tiền huấn luyện như PhoBERT-base có thực sự mang lại cải thiện rõ rệt hay không. Đồng thời, các mô hình Machine Learning cổ điển vẫn có giá trị ứng dụng thực tiễn trong nhiều hệ thống vì ưu điểm dễ triển khai, tốc độ huấn luyện nhanh và hoạt động ổn định khi dữ liệu không quá phức tạp.

Trong các bài toán xử lý ngôn ngữ tự nhiên, văn bản thô ban đầu tồn tại ở dạng chuỗi ký tự, vì vậy không thể đưa trực tiếp vào các mô hình ML truyền thống vốn chỉ xử lý dữ liệu số. Do đó, nhóm sử dụng TF-IDF (Term Frequency – Inverse Document Frequency) như một phương pháp biểu diễn văn bản cơ bản nhưng hiệu quả. TF-IDF biến mỗi câu thành một vector số, trong đó mỗi chiều tương ứng với một từ (hoặc n-gram) trong tập từ vựng. Giá trị của mỗi chiều phản ánh mức độ quan trọng của từ đó trong câu hiện tại so với toàn bộ tập dữ liệu. Nói cách khác, TF-IDF vừa xem xét số lần xuất hiện của từ trong câu (Term Frequency), vừa giảm trọng số của những từ xuất hiện quá phổ biến ở nhiều câu (Inverse Document Frequency). Đây là một cơ chế hợp lý vì trong sentiment analysis, các từ khóa như “tệ”, “không hiểu”, “chán”, “hài lòng”, “nhiệt tình”, “tâm huyết” thường mang tính phân biệt cao giữa các nhãn, trong khi các từ rất phổ biến như “là”, “và”, “của”, “ở”, “trong” gần như không giúp ích nhiều cho việc phân loại cảm xúc.

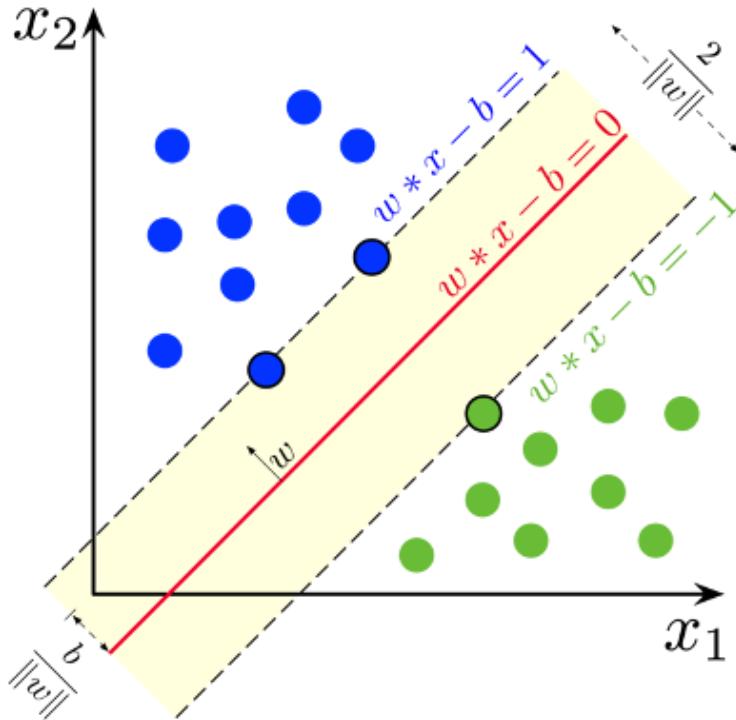
Quy trình áp dụng TF-IDF trong đồ án thường được thực hiện theo các bước: đầu tiên là tiền xử lý cơ bản (chuẩn hóa chữ thường, loại ký tự thừa, có thể xử lý stopwords tùy vào thiết kế), sau đó dùng TF-IDF Vectorizer để học từ vựng trên tập huấn luyện và chuyển từng câu thành vector đặc trưng dạng “sparse” (thưa). Việc biểu diễn sparse xuất hiện do mỗi câu chỉ chứa một số ít từ so với toàn bộ từ vựng, nên hầu hết các chiều trong vector là 0. Điều này phù hợp với Logistic Regression và SVM vì đây là các mô hình có khả năng xử lý tốt dữ liệu nhiều chiều và sparse. Với tiếng Việt, TF-IDF có thể áp dụng ở mức word-level hoặc kết hợp thêm n-gram (ví dụ unigram + bigram) để mô hình học được các cụm từ quan trọng như “không tốt”, “không hài lòng”, “rất hay”, “hoi nhanh”, “quá tệ”. Đây là điểm quan trọng vì sentiment trong tiếng Việt đôi khi không nằm ở một từ đơn lẻ mà nằm ở cả cụm từ hoặc câu trúc phủ định.

Sau khi văn bản được chuyển thành vector TF-IDF, nhóm tiến hành huấn luyện các mô hình phân loại truyền thống. Logistic Regression là một mô hình tuyến tính phổ biến và được xem là baseline mạnh trong phân loại văn bản. Mô hình học một bộ trọng số cho từng đặc trưng TF-IDF, sau đó tính tổng trọng số để quyết định nhãn đầu ra. Với bài toán nhiều lớp (Negative/Neutral/Positive), Logistic Regression thường sử dụng Softmax (multinomial) để ước lượng xác suất thuộc từng lớp. Ưu điểm lớn của Logistic Regression là khả năng huấn luyện nhanh, dễ kiểm soát overfitting bằng regularization (L1 hoặc L2), đồng thời khá dễ giải thích vì có thể xem các từ nào đang có trọng số cao cho từng nhãn. Trong thực tế, Logistic Regression thường hoạt động tốt khi dữ liệu sentiment có nhiều từ khóa trực tiếp, ví dụ như các câu khen/chê rõ ràng.

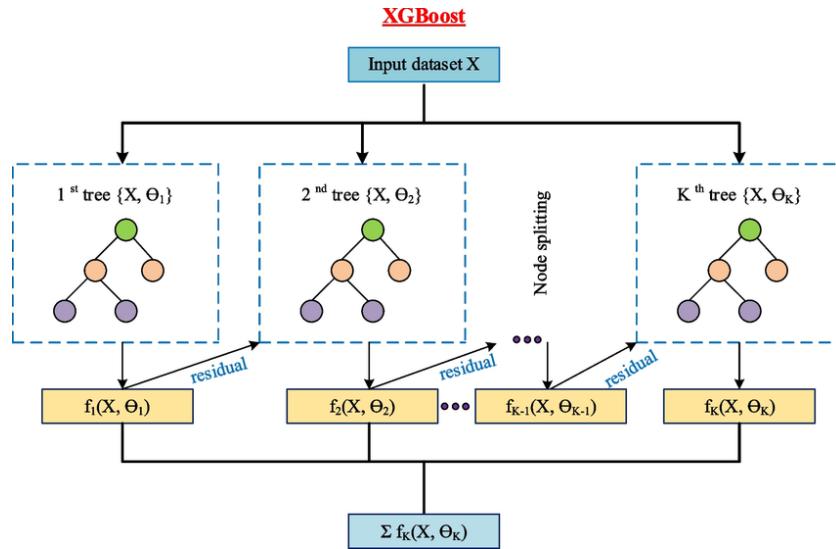


SVM (Support Vector Machine) là một mô hình rất mạnh cho dữ liệu TF-IDF, đặc biệt khi số chiều lớn. Thay vì tối ưu xác suất như Logistic Regression, SVM tối ưu một siêu phẳng phân tách các lớp sao cho khoảng cách biên (margin) là lớn nhất. Trực giác ở đây là mô hình không chỉ cố gắng phân loại đúng mà còn cố gắng phân tách “an toàn” nhất giữa các lớp, giúp tăng khả năng tổng quát hóa. Với văn bản tiếng Việt, SVM thường đạt kết quả tốt vì nó tận dụng tốt các đặc trưng sparse và không đòi hỏi quá nhiều dữ liệu để học. Tuy nhiên, hạn chế của SVM cũng tương tự Logistic Regression: mô hình chỉ thấy sự xuất hiện của từ/cụm từ trong văn bản, chứ không thực sự “hiểu” ngữ cảnh. Điều này khiến SVM gặp khó khăn ở những câu có cấu trúc phức tạp như phủ định kép (“không có gì không

hài lòng”), câu chuyển hướng (“hay nhưng hơi nhanh”), hoặc câu mang tính tu từ và ẩn dụ.



Ngoài các mô hình tuyến tính, nhóm cũng thử nghiệm XGBoost, một phương pháp ensemble dựa trên boosting và thường đạt hiệu năng cao trong nhiều bài toán dữ liệu dạng bảng. XGBoost hoạt động bằng cách kết hợp nhiều cây quyết định nhỏ, trong đó mỗi cây mới sẽ học để sửa lỗi của các cây trước. Khi áp dụng với TF-IDF, XGBoost có thể học được các tương tác phi tuyến giữa các đặc trưng (ví dụ một số cụm từ khi đi kèm nhau sẽ tạo cảm xúc mạnh hơn so với đứng riêng lẻ). Điều này giúp XGBoost đôi khi cải thiện Macro F1 so với các mô hình tuyến tính, đặc biệt trong những trường hợp dữ liệu có pattern không tuyến tính. Tuy nhiên, do TF-IDF có số chiều rất lớn và dữ liệu thưa, XGBoost có thể tốn thời gian huấn luyện hơn và cần tinh chỉnh hyperparameter cẩn thận để tránh overfitting. Đồng thời, dù mạnh hơn ở mặt phi tuyến, XGBoost vẫn không giải quyết triệt để vấn đề thiểu ngữ cảnh, vì bản chất TF-IDF không mang thông tin thứ tự hay quan hệ phụ thuộc giữa các token.



Một ưu điểm đáng chú ý khi sử dụng TF-IDF + ML truyền thống là hệ thống thường có tốc độ huấn luyện và suy luận nhanh hơn, dễ triển khai trên các thiết bị có tài nguyên hạn chế. Với các ứng dụng cần phản hồi nhanh hoặc triển khai trên CPU thông thường, Logistic Regression và SVM có thể là lựa chọn phù hợp. Ngoài ra, các mô hình này cũng phù hợp để làm baseline so sánh vì pipeline đơn giản và dễ tái lập. Trong đồ án, việc so sánh với các mô hình truyền thống giúp chứng minh rằng PhoBERT-base không chỉ “phức tạp hơn” mà còn “mang lại hiệu quả tốt hơn” nhờ khả năng học biểu diễn ngữ cảnh (contextual embeddings) thay vì chỉ dựa vào tần suất xuất hiện từ.

Tuy vậy, các mô hình dựa trên TF-IDF cũng bộc lộ hạn chế trong các trường hợp câu không chứa từ khóa sentiment trực tiếp hoặc chứa cấu trúc ngữ pháp phức tạp. Ví dụ, câu “bắt đầu buổi học đúng giờ” hoặc “cung cấp bài tập đa dạng” là những nhận xét tích cực nhưng lại không chứa nhiều từ cảm xúc mạnh như “rất hay”, “tuyệt vời”, nên mô hình TF-IDF có thể dễ nhầm sang Neutral. Ngược lại, các câu có phủ định hoặc từ khóa tiêu cực xuất hiện trong bối cảnh tích cực cũng có thể gây nhầm lẫn. Đây là điểm mà PhoBERT-base thường có lợi thế vì mô hình học được ngữ cảnh và quan hệ giữa các token nhờ self-attention, từ đó phân biệt tốt hơn các trường hợp “không” mang vai trò phủ định hay mang tính tu từ.

Tổng kết lại, nhóm lựa chọn TF-IDF kết hợp Logistic Regression, SVM và XGBoost nhằm xây dựng một bộ baseline đa dạng cả về tuyến tính lẫn phi tuyến, đảm bảo việc đánh giá PhoBERT-base khách quan hơn. TF-IDF đóng vai trò là phương pháp biểu diễn văn bản cổ điển, giúp chuyển đổi dữ liệu từ dạng

câu chữ sang dạng vector đặc trưng để các mô hình ML truyền thống có thể học và phân loại. Trong khi đó, Logistic Regression và SVM thể hiện thế mạnh về tốc độ và sự ổn định với dữ liệu sparse, còn XGBoost bổ sung khả năng học các quan hệ phi tuyến giữa đặc trưng. Tuy nhiên, các mô hình này vẫn bị giới hạn bởi việc TF-IDF không mang tính ngữ cảnh, do đó khó xử lý tốt các câu có sắc thái phức tạp. Đây cũng chính là cơ sở để nhóm lựa chọn PhoBERT-base làm mô hình trọng tâm, nhằm tận dụng sức mạnh của biểu diễn ngữ cảnh và cải thiện hiệu năng tổng thể trên bài toán phân loại cảm xúc tiếng Việt.

Bảng kết quả đạt được:

Model	Macro F1	Accuracy
TF-IDF + Logistic Regression	0.73	0.87
TF-IDF + SVM	0.71	0.89
TF-IDF + XGBOOST	0.74	0.86
vinai/phobert-base	0.7737	0.8920

NHẬN XÉT:

Dựa trên bảng kết quả, có thể thấy các mô hình Machine Learning truyền thống sử dụng đặc trưng TF-IDF như Logistic Regression, SVM và XGBoost đều cho hiệu năng khá tốt, với Accuracy dao động từ 0.86–0.89 và Macro F1 khoảng 0.71–0.74. Trong đó, TF-IDF + XGBoost đạt Macro F1 cao nhất trong nhóm ML (0.74), cho thấy mô hình cây tăng cường có khả năng khai thác đặc trưng tốt hơn trong một số trường hợp. Tuy nhiên, mô hình vinai/phobert-base vẫn cho kết quả nổi bật hơn khi đạt Macro F1 = 0.7737 và Accuracy = 0.8920, vượt trội so với các baseline TF-IDF. Điều này cho thấy PhoBERT có lợi thế rõ rệt nhờ khả năng học biểu diễn ngữ cảnh (contextual embedding), giúp xử lý tốt hơn các câu mang sắc thái phức tạp, phụ thuộc vào ngữ nghĩa và cấu trúc câu, thay vì chỉ dựa vào tần suất từ như TF-IDF.

5.2. Phân tích các trường hợp đúng/sai

Để có thể hiểu rõ hơn những ưu điểm cũng như hạn chế của mô hình sau khi train, nhóm cũng đã tiến hành đi phân tích một số trường hợp mô hình đoán đúng và đoán sai.

Đầu tiên là các trường hợp mô hình dự đoán đúng nhãn:

Câu ví dụ	Nhân đúng	Nhân dự đoán
“sinh viên không tiếp thu kịp cung như không hiểu gì.”	Negative	Negative
“chưa giỏi chuyên môn cho lắm.”	Negative	Negative
“giảng viên đảm bảo thời gian trên lớp , tạo điều kiện trong quá trình thực hành và thi thực hành.”	Positive	Positive
“giáo viên rất vui tính”	Positive	Positive
“cô max có tâm”	Positive	Positive
“giáo viên không giảng dạy kiến thức , hướng dẫn thực hành trong quá trình học”	Negative	Negative
“thầy dạy nhiệt tình và tâm huyết”	Positive	Positive
“thầy nhiệt tình giảng lại cho học sinh”	Positive	Positive
“có đôi lúc nói hơi nhanh làm sinh viên không theo kịp”	Negative	Negative
“giảng dạy nhiệt tình , liên hệ thực tế khá nhiều , tương tác với sinh viên tương đối tốt”	Positive	Positive

Nhìn vào 10 câu ví dụ trong bảng, có thể thấy mô hình dự đoán đúng chủ yếu nhờ các câu đều chứa những từ khóa cảm xúc rất rõ ràng, và đa số câu có xu hướng cảm xúc nhất quán, ít xuất hiện các câu mang hàm ý hoặc chứa các từ khóa của các nhân khác. Hãy cùng phân tích từng câu để thấy rõ luận điểm này:

Nhóm câu mang nhân Positive

Câu “giáo viên rất vui tính” có cụm “rất vui tính” là một biểu thức khen ngợi rõ ràng. Đây là kiểu câu mà mô hình dễ dự đoán vì cảm xúc nằm ngay ở một cụm tính từ tích cực, không cần suy luận nhiều từ ngữ cảnh.

Câu “cô max có tâm” tuy có yếu tố ngôn ngữ mang (“max”), nhưng từ khóa quan trọng là “có tâm” thường mang ý nghĩa đánh giá cao thái độ, trách nhiệm và sự tận tình. Vì vậy mô hình có thể nhận diện đây là câu tích cực.

Câu “thầy dạy nhiệt tình và tâm huyết” chứa hai cụm cảm xúc tích cực mạnh là “nhiệt tình” và “tâm huyết”. Hai từ này thường xuất hiện trong các bình luận khen giáo viên/giảng viên, nên chúng đóng vai trò như tín hiệu phân loại rất chắc. Đặc biệt, đây là những từ mang tính “đánh giá thái độ giảng dạy”, nên liên hệ với nhãn Positive gần như trực tiếp.

Câu “thầy nhiệt tình giảng lại cho học sinh” có từ “nhiệt tình” và mô tả hành động hỗ trợ “giảng lại” tạo cảm giác được giúp đỡ. Do đó, mô hình không chỉ dựa vào từ khóa mà còn dựa vào ngữ cảnh hành động để cung cấp dự đoán Positive.

Câu “giảng dạy nhiệt tình, liên hệ thực tế khá nhiều, tương tác với sinh viên tương đối tốt” là câu dài hơn, nhưng cảm xúc lại càng rõ ràng hơn nhờ nhiều cụm tích cực xuất hiện liên tiếp. “nhiệt tình” biểu hiện thái độ tốt, “liên hệ thực tế khá nhiều” thể hiện chất lượng bài giảng, và “tương tác … tương đối tốt” là đánh giá cao sự giao tiếp với sinh viên. Vì trong câu có nhiều tín hiệu tích cực cùng hướng, mô hình dễ “đi đến kết luận” nhãn Positive.

Câu “giảng viên đảm bảo thời gian trên lớp, tạo điều kiện trong quá trình thực hành và thi thực hành” có các cụm “đảm bảo thời gian” và “tạo điều kiện” đều mang nghĩa tích cực, thể hiện việc giảng viên thực hiện tốt trách nhiệm và hỗ trợ người học. Đây là dạng câu mà mô hình có thể dự đoán đúng nhờ học được rằng các động từ/cụm từ mang nghĩa hỗ trợ và đảm bảo quyền lợi thường nghiêng về Positive.

Nhóm câu mang nhãn Negative

Câu “giáo viên không giảng dạy kiến thức, hướng dẫn thực hành trong quá trình học” có từ mang tính phủ định “không” đặt trước các nội dung cốt lõi như “giảng dạy kiến thức” và “hướng dẫn thực hành”. Khi phủ định rơi vào những phần “đáng lẽ phải có” trong một khóa học, câu này trở thành phần nàn nghiêm trọng, nên mô hình dễ phân loại Negative.

Câu “có đôi lúc nói hơi nhanh làm sinh viên không theo kịp” có cấu trúc gộp ý nhưng vẫn rõ tiêu cực vì có hậu quả “không theo kịp”. Cụm này mang nghĩa “không hiểu bài”, thường là trải nghiệm khó chịu của người học, nên rất hay được gán Negative. Vì vậy mô hình dự đoán đúng là hợp lý.

Câu “sinh viên không tiếp thu kịp cũng như không hiểu gì” có mức độ tiêu cực cao hơn do dùng liên tiếp hai cụm phủ định “không tiếp thu kịp” và “không hiểu gì”. Đây là dạng câu tiêu cực rõ ràng nhất vì thể hiện cảm giác thất bại trong

việc học. Việc nhắc đến “không hiểu gì” thường là tín hiệu mạnh, nên mô hình có xu hướng gán Negative.

Câu “chưa giỏi chuyên môn cho lắm” là câu ngắn, nhưng từ khóa “chưa giỏi” thể hiện chê trực diện về năng lực. Trong phân loại cảm xúc, các cấu trúc đánh giá năng lực kèm phủ định/giảm mức độ như “chưa giỏi”, “không ổn”, “không tốt” thường bị gán nhãn Negative vì mang ý bất mãn hoặc không hài lòng.

Tiếp theo hãy xét các trường hợp mà mô hình dự đoán sai nhãn:

Câu ví dụ	Nhãn đúng	Nhãn dự đoán
“chỗ tiếng việt giống như là copy nguyên văn từ google dịch vậy.”	Neutral	Negative
“ không có điều gì không hài lòng.”	Positive	Neutral
“cô cho tài liệu học tập là một trang web , lên đó tự học và làm đồ án.”	Negative	Neutral
“cung cấp bài tập đa dạng”	Positive	Neutral
“bắt đầu buổi học đúng giờ”	Positive	Negative
“trong trường macbok thầy số hai thì không có máy nào số một”	Positive	Negative
“tính điểm thi đua các nhóm.”	Positive	Neutral
“giảng bài thu hút , dí dỏm”	Positive	Negative
“nói tiếng anh lưu loát.”	Positive	Neutral
“ấn tượng nhất doubledot dạy không cần máy chiếu hay laptop , nhưng lượng bài giảng có thể nói là “khủng ” , nhìn thầy ghi bài giảng lên bảng mà bất ngờ.”	Positive	Negative

Nhìn vào các ví dụ dự đoán sai, có thể thấy phần lớn lỗi của mô hình không đến từ việc mô hình “không hiểu tiếng Việt”, mà chủ yếu do các câu thuộc nhóm khó phân loại cảm xúc vì mang tính mơ hồ, thiếu từ khóa cảm xúc rõ ràng, có yếu tố so sánh/châm biếm hoặc chứa cấu trúc phủ định dễ gây nhầm lẫn. Ngoài ra, một số câu có ý nghĩa thực tế là tích cực hoặc trung lập nhưng lại xuất hiện các từ có sắc thái tiêu cực (hoặc ngược lại), khiến mô hình bị “đánh lừa” khi dự đoán. Hãy cùng phân tích từng câu sai để hiểu hơn về mặt hạn chế này của mô hình:

Câu “chỗ tiếng việt giống như là copy nguyên văn từ google dịch vậy.”, mô hình có xu hướng gán nhãn tiêu cực vì câu chứa các từ/cụm từ mang sắc thái chê như “giống copy”, “google dịch”, thường đi kèm ý phê bình chất lượng. Vì vậy mặc dù câu không có ý nghĩa công kích mạnh nhưng mô hình đã được huấn luyện thiên về những từ mang sắc thái nêu trên.

Ở câu “không có điều gì không hài lòng.”, đây là câu khen mang ý nghĩa tích cực (tức là hoàn toàn hài long), nhưng có cấu trúc phủ định kép “không … không ..” khiến mô hình hiểu nhầm thành câu trung tính.

Ở câu “cô cho tài liệu học tập là một trang web , lên đó tự học và làm đồ án.” và câu “tính điểm thi đua các nhóm”, các câu này mang tính “kể lại sự việc” nên mô hình dự đoán trung tính vì không có bất kì từ nào thể hiện cảm xúc của người viết. Tuy nhiên nhãn đúng lại là Negative vì có thể ngầm phản ánh việc thiếu hướng dẫn trực tiếp, “tự học” bị hiểu như phàn nàn. Còn câu còn lại mang hàm ý tạo động lực học tập hoặc tổ chức hoạt động nhóm tốt nên được đánh nhãn Positive.

Câu “trong trường macbok thầy số hai thì không có máy nào số một” là một câu mang phong cách khen “thầy là số 1” nhưng được diễn đạt theo kiểu so sánh/ẩn dụ và có chứa từ phủ định “không có”. Cấu trúc “thầy số hai thì không có ai số một” thực chất nghĩa là thầy giỏi nhất, nhưng mô hình có thể bị nhiễu vì gặp từ phủ định “không có” và cấu trúc câu không phổ biến. Trường hợp này tương tự lỗi phủ định kép: câu mang ý khen mạnh nhưng lại chứa các dạng phủ định khiến mô hình hiểu nhầm là Negative.

Câu dài nhất cuối cùng là một ví dụ điển hình của câu khen mạnh nhưng có từ dễ gây hiểu nhầm. Câu chứa cụm “không cần máy chiếu hay laptop” và nhiều phủ định/so sánh, nhưng thực tế ý nghĩa lại là khen năng lực giảng dạy (không cần thiết bị vẫn dạy tốt, bài giảng “khủng”). Ngoài ra câu còn có từ “bất ngờ”, “khủng” – ngôn ngữ mạng – nếu mô hình không học tốt các từ slang này thì khả năng nhận diện Positive giảm. Khi câu dài và chứa nhiều thành phần, mô hình đôi khi bị “hút” bởi những token phủ định hoặc những đoạn mô tả gây nhiễu thay vì trọng tâm khen ngợi.

Ngoài ra, những câu “cung cấp bài tập đa dạng”, “bắt đầu buổi học đúng giờ”, “giảng bài thu hút, dí dỏm” hoặc “nói tiếng anh lưu loát.” thì từ ngữ cảm xúc thuộc nhãn đúng khá rõ ràng và nhất quán từ đầu đến cuối nhưng có thể mô hình vẫn được fine-tune đủ sâu nên vẫn còn sai sót trong những câu khá dễ dự đoán như vậy

NHẬN XÉT:

Tóm lại, qua các ví dụ đúng và các lỗi phân loại cho thấy PhoBERT-base hoạt động tốt với cảm xúc tường minh, nhưng giảm hiệu quả ở các câu có cảm xúc ẩn, phụ thuộc ngữ cảnh, hoặc dễ gây nhầm giữa Neutral và Negative/Positive. Đây là dạng lỗi thường gặp trong bài toán sentiment tiếng Việt, đặc biệt với dữ liệu phản hồi về giáo viên/môn học do văn phong thường mang tính “góp ý nhẹ” và ít dùng từ biểu cảm mạnh.

CHƯƠNG VI: CÁC HƯỚNG PHÁT TRIỂN TIẾP THEO

Trong tương lai, hệ thống có thể được cải thiện trước hết ở khâu tiền xử lý văn bản do dữ liệu đầu vào là câu do người dùng nhập vào nên thường chứa nhiều biến thể khó lường như viết tắt, sai chính tả, thiếu dấu, kéo dài ký tự quá mức hoặc xen kẽ tiếng Anh. Hiện tại pipeline đã xử lý tốt các trường hợp cơ bản như chuẩn hoá chữ thường, dấu câu, emoji và rút gọn ký tự lặp, tuy nhiên vẫn có thể mở rộng thêm bằng cách xây dựng tập quy tắc (lexicon) cho các dạng phủ định phổ biến (“ko/k/k0”, “hok”, “chả”,...), các từ nhấn mạnh mức độ (“cực kỳ”, “siêu”, “quá trời”), cũng như chuẩn hoá các biến thể tiếng lóng thường thấy ở lứa tuổi học sinh/sinh viên và từ lóng trên mạng Internet.

Bên cạnh đó, có thể phát triển thêm các cách xử lý đối với những token mang tính nhiễu hoặc không mang ý nghĩa ngữ nghĩa rõ ràng như chuỗi ký tự ngẫu nhiên, mã lớp học, tên viết tắt, hoặc các đoạn văn bản không liên quan. Trong pipeline hiện tại thì những từ đó đang bị tách thành nhiều subword do cơ chế BPE nên nếu những từ này có thể được thay thế bằng các token chuẩn hóa như [id] cho MSSV, mã số giáo viên hay [name] cho tên người sẽ làm tang tính tổng quát của mô hình hơn.

Ngoài fine-tuning theo cách thông thường, mô hình cũng có thể được cải thiện bằng cách áp dụng các kỹ thuật huấn luyện nâng cao như điều chỉnh learning rate theo từng tầng (layer-wise learning rate decay), early stopping, hoặc tăng cường regularization nhằm hạn chế overfitting khi dữ liệu gán nhãn không quá lớn.

REFERENCE

<https://aclanthology.org/2020.findings-emnlp.92/>

<https://doi.org/10.48550/arXiv.1907.11692>

<https://huggingface.co/blog/sentiment-analysis-python>

<https://ieeexplore.ieee.org/document/7860361>

<https://www.researchgate.net/publication/329645066 UIT-VSFC Vietnamese Students' Feedback Corpus for Sentiment Analysis>