

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN  
KHOA KHOA HỌC MÁY TÍNH**



**BÁO CÁO ĐỒ ÁN**  
**ĐỀ TÀI: ASPECT-BASED SENTIMENT ANALYSIS**  
**CHO BÌNH LUẬN TIẾNG VIỆT VỀ CHỦ ĐỀ**  
**SMARTPHONE**  
**TRÊN CÁC TRANG THƯƠNG MẠI ĐIỆN TỬ**

LỚP: CS114.N21

**GV HƯỚNG DẪN**  
PGS.TS. LÊ ĐÌNH DUY  
THS. PHẠM NGUYỄN TRƯỜNG AN

**NHÓM SINH VIÊN THỰC HIỆN: KHNT**

STT	HỌ VÀ TÊN	MSSV
1	ĐOÀN NHẬT SANG (TRƯỞNG NHÓM)	21522542
2	TRƯỜNG VĂN KHẢI	21520274
3	ĐẶNG PHƯỚC SANG	21521377

**TP.HCM, NGÀY 9 THÁNG 7 NĂM 2023**

## **CẬP NHẬT SAU KHI VẤN ĐÁP**

- Chỉnh lại format theo khoá luận tốt nghiệp
  - Font chữ Unicode: Times New Roman, kích thước(size) 13pt.
  - Dẫn dòng (line spacing): 1.5 lines.
  - Lề trên 3 cm, lề dưới 3.5 cm, lề trái 3.5 cm, lề phải 2 cm.
  - Bắt đầu đánh số trang từ phần Tóm tắt nội dung đã làm: đánh số trang ở giữa bên dưới, theo hệ thống số Ả-rập
- Thêm danh mục ảnh và danh mục bảng ở trang thứ 2 kể từ trang này.
- [Trích dẫn theo kiểu IEEE](#) – tài liệu tham khảo trang 26

## MỤC LỤC

<b>1.</b>	<b>TÓM TẮT NỘI DUNG ĐÃ LÀM .....</b>	<b>1</b>
<b>2.</b>	<b>MỞ ĐẦU.....</b>	<b>3</b>
2.1.	Lý do chọn đề tài .....	3
2.2.	Ngữ cảnh ứng dụng .....	3
2.3.	Input và output.....	4
<b>3.</b>	<b>PHÂN CÔNG CÔNG VIỆC .....</b>	<b>5</b>
<b>4.</b>	<b>THU THẬP DỮ LIỆU .....</b>	<b>6</b>
4.1.	Cách xây dựng bộ dữ liệu UIT - ViSFD.....	6
4.2.	Khó khăn của việc thu thập dữ liệu: .....	8
4.3.	Lý do chọn bộ dataset UIT-ViSFD:.....	9
4.4.	Thống kê và tiền xử lý dữ liệu trên UIT-ViSFD .....	9
<b>5.</b>	<b>TRAIN VÀ ĐÁNH GIÁ MODEL .....</b>	<b>13</b>
5.1.	Hướng tiếp cận và chọn model để huấn luyện.....	13
5.2.	Tổng quan về Bi-LSTM, Phobert-based method.....	14
5.2.1.	Phobert-based method .....	14
5.2.2.	Bi-LSTM.....	17
5.3.	Đánh giá mô hình .....	19
<b>6.</b>	<b>THÁCH THỨC.....</b>	<b>22</b>
<b>7.</b>	<b>ỨNG DỤNG VÀ HƯỚNG PHÁT TRIỂN .....</b>	<b>24</b>
7.1.	Ứng dụng .....	24
7.2.	Hướng phát triển trong tương lai .....	24
<b>8.</b>	<b>TÀI LIỆU THAM KHẢO .....</b>	<b>26</b>
8.1.	Tài liệu tiếng việt.....	26
8.2.	Tài liệu tiếng anh .....	26

## DANH MỤC HÌNH

Figure 4.1: IAA của 6 giai đoạn con.....	8
Figure 4.2: Preprocess pipeline.....	12
Figure 5.1: PhoBERT Multi-task Approach Architecture .....	15
Figure 5.2: PhoBERT Multi-task with Multi-branch Approach Architecture.....	17
Figure 5.3: Bi-LSTM ArchitectureArchitecture.....	17
Figure 7.1: Local Web demo sử dụng Sreamlit.....	24

## DANH MỤC BẢNG

Table 3.1: Danh sách thành viên.....	5
Table 3.2: Bảng phân công công việc.....	5
Table 4.1: Các loại Aspect và ý nghĩa của từng Aspect.....	7
Table 4.2: Bảng thống kê tổng quan trên train/val/test set.....	10
Table 4.3: Bảng phân phối của aspect và polarity trên train/test/val set .....	10
Table 5.1: Kết quả so sánh 2 phương pháp dựa trên PhoBERT trên 2 dataset trong hội nghị VLSP 2018.....	14
Table 5.2: Kết quả chạy thực nghiệm trên tất cả 3 phương pháp .....	20

## 1. TÓM TẮT NỘI DUNG ĐÃ LÀM

Tìm hiểu cách xây dựng của một số bộ dataset ở các hội nghị lớn như SemEval, VLSP 2018 và một số bài báo khoa học như UIT-ViSFD. Lựa chọn bộ dữ liệu UIT-ViSFD để thực hiện đề tài.

Quan sát bộ dữ liệu UIT-ViSFD, để đưa ra một số thao tác tiền xử lý dữ liệu:

- Xóa bỏ các tag html
- Chuyển đổi các ký tự unicode 1252 sang utf-8
- Chuẩn hoá các từ tiếng Việt theo quy tắc đặt dấu thanh của chữ Quốc ngữ kiểu cũ [1]
- Chuyển đổi các từ viết tắt phổ biến thành từ tiếng việt có đủ nghĩa hơn.  
VD: tks, thks, thanks, thank  $\Rightarrow$  cảm ơn
- Chuyển đổi emoji thành các từ tiếng việt có nghĩa tương tự.
- Sử dụng word segmentation trên câu để gom nhóm được các từ tiếng Việt một cách đủ nghĩa.
- Xóa các ký tự không cần thiết, tránh gây nhiễu đến ý nghĩa của câu.

$\Rightarrow$  Đây của bộ dữ liệu sau khi loại bỏ một vài thông tin không cần thiết và một số thao tác tiền xử lý trên label

(<https://github.com/dnsang1611/CS114.N21-KHNT/tree/master/final-project/dataset/preprocessed-data>)

Sử dụng 3 mô hình Deep learning để huấn luyện (nhấp vào tên mô hình để đến file colab tương ứng):

- Multi-task method with PhoBERT (Visualize data và các thao tác tiền xử lý sẽ được chú thích kỹ ở file này)

(<https://colab.research.google.com/drive/1ENWKZkOcVvIaPyTX1fuWieTkqNZLVzGN#scrollTo=lDED4mxv521k>)

- Multi-task multi-branch with PhoBERT (Các thao tác tiền xử lý sẽ được chú thích kỹ ở file này)

([https://colab.research.google.com/drive/199Y\\_ezuwgHqSxh5Wd2ZMUvfIU3MFfTy4#scrollTo=xNSQwkZBvo1x](https://colab.research.google.com/drive/199Y_ezuwgHqSxh5Wd2ZMUvfIU3MFfTy4#scrollTo=xNSQwkZBvo1x))

- Bi-LSTM với Elmo embedding model pretrained on Vietnamese (Các thao tác tiền xử lý sẽ được chú thích kỹ ở file này)

(<https://colab.research.google.com/drive/1Xvc57wfwMDTcvX3Cv0sn502SK6qSZv1G#scrollTo=XUbPef5-Xi36>)

Tiến hành phân tích, đánh giá, nhận xét performance của 3 model trên tập test, từ đó so sánh chọn ra model tốt nhất.

Xây dựng demo bằng streamlit dựa trên model tốt nhất.

## **2. MỞ ĐẦU**

### **2.1. Lý do chọn đề tài**

Theo thống kê, trong năm 2022, lượng người dùng smartphone tại Việt Nam đạt khoảng 62,8 triệu người và tính đến năm 2026, ước đạt 67,3 triệu người (Theo vietnambiz.vn 09/05/2023) [2] cho thấy nhu cầu sở hữu smartphone ngày càng tăng. Với sự phát triển mạnh mẽ của thương mại điện tử, việc mua smartphone đã trở nên dễ dàng hơn và người dùng thường có xu hướng để lại những bình luận của mình trên các trang thương mại điện tử này. Hiện nay, các trang thương mại điện tử bán smartphone phổ biến ở Việt Nam như Shopee, Thegioidiong, ... vẫn đang dùng hệ thống sao để cung cấp cái nhìn tổng quan về sản phẩm cho người dùng và doanh nghiệp. Tuy nhiên, để biết được những bình luận đó đang thể hiện độ như thế nào và mặt nào của sản phẩm thì họ vẫn phải đọc từng bình luận, điều này gây mất nhiều thời gian. Từ đó, nhóm quyết định chọn đề tài aspect-based sentiment analysis cho bình luận tiếng Việt trên chủ đề smartphone để có thể giúp tự động đánh giá cảm xúc của khách hàng về một số khía cạnh cụ thể của smartphone và dịch vụ bán hàng. Điều này tạo cơ sở cho việc vẽ các biểu đồ thống kê, chức năng lọc các bình luận về một khía cạnh nào đó hay cụ thể hơn là khía cạnh gì với cảm xúc như thế nào để người dùng có cái nhìn tổng quát một cách nhanh chóng, từ đó đưa ra quyết định mua hàng hoặc nâng cao chất lượng dịch vụ, sản phẩm.

Trong bài báo cáo này, chúng tôi sẽ đề xuất 3 phương pháp là sử dụng Bi-LSTM và 2 phương pháp còn lại dựa trên PhoBERT cho ABSA tiếng Việt.

### **2.2. Ngữ cảnh ứng dụng**

Thông thường, trước khi mua một món hàng hay quyết định sử dụng dịch vụ online nào đó, mọi người thường có xu hướng tìm kiếm các lời khuyên, review từ những người đã mua món hàng hoặc sử dụng dịch vụ đó. Kèm theo sự phát triển vượt trội của các trang thương mại điện tử, các diễn đàn đánh giá sản phẩm cùng ngày

càng hoạt động sôi nổi. Do đó, số lượng đánh giá ngày càng tăng và trở thành nguồn tài nguyên quý giá cho khách hàng và cho doanh nghiệp. Đối với khách hàng, nguồn dữ liệu này cung cấp thông tin về sản phẩm và những lời khuyên hữu ích giúp họ tránh mua các sản phẩm hoặc đăng ký dịch vụ không phù hợp. Mặt khác, đánh giá của người dùng cũng là thông tin quý giá cho doanh nghiệp, nếu sử dụng hiệu quả nguồn dữ liệu này có thể giúp cho doanh nghiệp nâng cao chất lượng sản phẩm, xác định chính xác nhu cầu khách hàng.

### **2.3. Input và output**

Input: Một hay nhiều câu bình luận (kiểu chuỗi).

Output: Danh sách các khía cạnh liên quan đến điện thoại hoặc dịch vụ (SCREEN, CAMERA, FEATURES, BATTERY, PERFORMANCE, STORAGE, DESIGN, PRICE, GENERAL, SER&ACC) có trong câu cùng với thái độ (positive, negative, neutral) ứng với từng khía cạnh đó.

Những khía cạnh trên được tham khảo từ [3]



### 3. PHÂN CÔNG CÔNG VIỆC

Họ và tên	MSSV	Email
Đoàn Nhật Sang	21522542	21522542@gm.uit.edu.vn
Trương Văn Khải	21520274	21520274@gm.uit.edu.vn
Đặng Phước Sang	21521377	21521377@gm.uit.edu.vn

*Table 3.1: Danh sách thành viên*

	N.Sang	Khải	P. Sang
Tìm hiểu cách xây dựng bộ dữ liệu	+	+	
Các thao tác tiền xử lý	+	+	+
Phân tích bộ dữ liệu ViSFD	+		
Bi-LSTM	+	+	
Multi-task multi-branch approach PhoBERT	+		+
Multi-task approach với PhoBERT	+		+
Đánh giá kết quả	+	+	+
Hoàn thành nội dung báo cáo	+	+	+
Demo	+	+	

*Table 3.2: Bảng phân công công việc*

## 4. THU THẬP DỮ LIỆU

### 4.1. Cách xây dựng bộ dữ liệu UIT - ViSFD

UIT-ViSFD (UIT - Vietnamese smartphone feedback dataset) là một bộ dữ liệu về aspect-based sentiment analysis được công bố trong bài báo [3] được chấp nhận tại hội nghị KSEM 2021.

Nhóm tác giả thu thập thông tin phản hồi bằng văn bản từ khách hàng trên một trang web thương mại điện tử lớn ở Việt Nam. Để đảm bảo dữ liệu đa dạng và có giá trị, họ thu thập phản hồi từ top-10 thương hiệu điện thoại thông minh phổ biến được sử dụng tại Việt Nam. Có nhiều bình luận dài, đánh giá lan man và đánh giá mâu thuẫn, không rõ ràng đối với việc hiểu để xác định đúng nhãn của chúng. Vì vậy, các ý kiến dài hơn 250 tokens (chiếm một tỷ lệ rất nhỏ) sẽ bị xóa. Đồng thời, tác giả cũng xóa những bình luận có quá nhiều lỗi chính tả trong đó, khiến người đọc gặp khó khăn trong hiểu và chú giải chính xác.

Dataset gồm 10 aspect cụ thể, các giá trị có thể của polarity ứng với các aspect này là {negative, positive, neutral}. Ngoài ra, có 1 aspect đặc biệt là OTHERS dùng để chỉ bình luận chứa những thông tin không liên quan đến 10 aspects trên hoặc không đánh giá sản phẩm, và aspect này sẽ không được đánh nhãn polarity.

Aspect	Mean	User comments	Sentiment
SCREEN	User comments express screen quality, size, colors, and display technology.	màn hình đẹp (a nice screen)	Pos
CAMERA	The comments mention the quality of a camera, vibration, delay, focus, and image colors.	điện thoại chụp hình mờ (the phone took blur picture)	Neg
FEATURES	The users refer to features, fingerprint sensor, wifi connection, touch and face detection of the phone.	nhận diện khuôn mặt chậm (the face detection is slow)	Neg
BATTERY	The comment describes battery capacity and battery quality.	pin trâu (long battery life)	Pos
PERFORMANCE	The reviews describe ramming capacity, processor chip, performance using, and smoothness of the phone.	cấu hình có thể chấp nhận được (acceptable configuration )	Neu
STORAGE	The comment mention storage capacity, the ability to expand capacity through memory cards.	bộ nhớ lớn (large storage)	Pos
DESIGN	The reviews refer to the style, design, and shell.	điện thoại thiết kế thô (rough design phone)	Neg
PRICE	The comments present the specific price of the phone.	giá cả ở mức trung bình (the price is at average)	Neu
GENERAL	The reviews of customers generally comment about the phone.	mọi thứ đều ok (everything is ok)	Pos
SER&ACC <sup>7</sup>	The comments mention sales service, warranty, and review of accessories of the phone.	nhân viên tư vấn nhiệt tình (shop assistants advice enthusiastic)	Pos

*Table 4.1: Các loại Aspect và ý nghĩa của từng Aspect*

*trong UIT-ViSFD (Nguồn: [3])*

Quá trình annotation: gồm 2 giai đoạn, 5 annotators

- Training phase: gồm có 6 giai đoạn con, dùng để đạt được inter-annotator agreement (IAA) cao và guideline nghiêm ngặt. Trong giai đoạn con đầu tiên, sau khi xây dựng guideline đầu tiên, các annotators sẽ đánh nhãn 200 comment cùng với nhau để hiểu được nguyên tắc đánh nhãn. Trong 5 giai đoạn con còn lại, nhóm tác giả sẽ lấy ra 200 bình luận ngẫu nhiên và mỗi người sẽ phải đánh nhãn cho 200 bình luận này. Nếu có sự bất đồng ý kiến, các annotators sẽ thảo luận và đưa ra quyết định thông qua việc bỏ phiếu, và cập nhật guideline. Ở mỗi giai đoạn con, để đảm bảo chất lượng của dataset nhóm tác giả sẽ tính IAA bằng hệ số Cohen's Kappa giữa các thành viên trên việc đánh nhãn aspect và polarity. Công thức của nó như sau:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Trong đó  $k$  là IAA,  $Pr(a)$  là sự đồng thuận quan sát được giữa các annotators,  $Pr(e)$  là xác suất giả định của các khả năng đồng thuận. Đến khi IAA đạt trên 80% và hoàn thiện guideline đánh nhãn, các annotators sẽ chuyển sang Labeling phase.

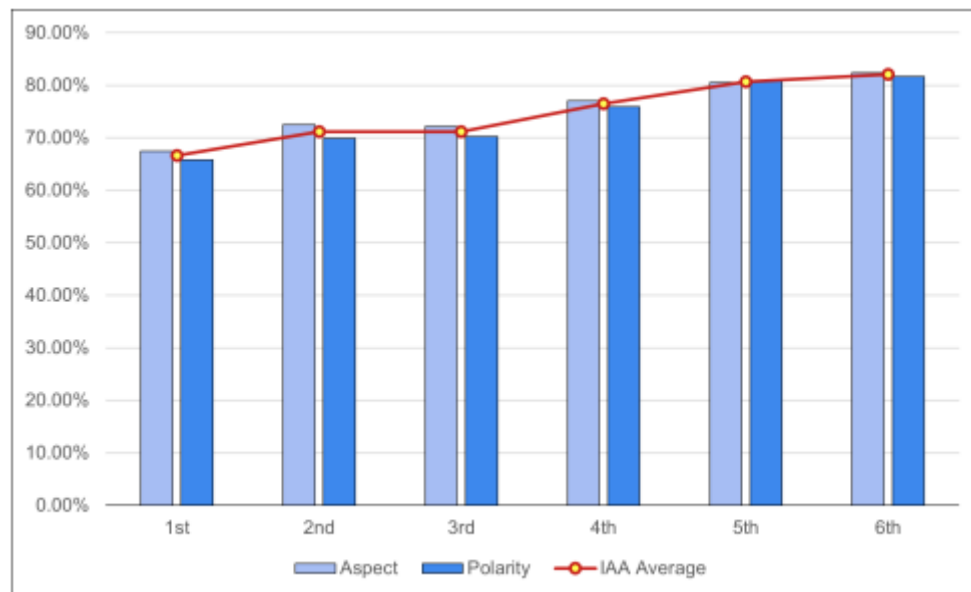


Figure 4.1: IAA của 6 giai đoạn con (Nguồn: [3])

- Labeling phase: Các annotators sẽ tiến hành đánh nhãn một cách độc lập. Nếu gặp những trường hợp khó, bất đồng ý kiến thì sẽ nhóm sẽ thảo luận để đưa ra nhãn cuối cùng và hoàn thiện guideline đánh nhãn.

#### 4.2. Khó khăn của việc thu thập dữ liệu:

Việc thu thập dữ liệu sẽ gặp khó khăn trong giai đoạn đánh nhãn dữ liệu. Thực tế, nó phụ thuộc nhiều vào phán đoán chủ quan của người chú thích. Hơn nữa, khó để đạt được sự đồng thuận chung trong cách chú thích của một nhóm người cùng thực hiện. Thậm chí một cá nhân không phải lúc nào cũng nhất quán trong cách họ gán nhãn dữ liệu. Có một số lý do cho điều này, chẳng hạn như độ khó vốn có của nội dung, quan điểm chủ quan của cá nhân hoặc đơn giản là chất lượng chú thích kém.

Để giảm thiểu tối đa khả năng mâu thuẫn, điều quan trọng là phải có một nhóm gồm nhiều người chú thích để kiểm tra chéo giữa các lượt gán nhãn và thực hiện nhiều lượt kiểm tra, đối chiếu giữa các thành viên.

#### **4.3. Lý do chọn bộ dataset UIT-ViSFD:**

UIT-ViSFD là một bộ dữ liệu có số lượng lớn mẫu dữ liệu, được tiến hành đánh nhãn theo một quá trình nghiêm ngặt nên đảm bảo được chất lượng tốt. Bên cạnh đó, nó cũng phù hợp với đề tài mà nhóm đã chọn: bình luận về smartphone trên trang thương mại điện tử.

Mục đích của nhóm cũng là thử nhiều phương pháp để có thể đánh giá và từ đó chọn ra phương án tốt nhất để demo sản phẩm.

#### **4.4. Thống kê và tiền xử lý dữ liệu trên UIT-ViSFD**

Dataset gồm 11,122 bình luận, bao gồm 5 cột index (số thứ tự của hàng), comment (bình luận), n\_star (Số sao đánh giá), date\_time (ngày lập bình luận), và label (gồm tập các cặp aspect#polarity).

Dataset được chia làm train, val, test theo tỷ lệ 7:1:2. Trung bình, mỗi bình luận sẽ có 3 aspect và có chiều dài khoảng 36 token (chưa dùng word segmentation). Có sự mất cân bằng khá lớn giữa các aspects, nhiều nhất là GENERAL (6936) và ít nhất là STORAGE (132). Và polarity cũng có sự mất cân bằng, nhiều nhất là Positive chiếm 56,13% tiếp đến là negative chiếm 31,7% và thấp nhất là neutral chiếm 12,17%. Việc mất cân bằng dữ liệu này sẽ gây ra nhiều thách thức cho các mô hình machine learning.

	Train	Dev	Test
Number of Comments	7,786	1,112	2,224
Number of Tokens	283,460	39,023	80,787
Number of Aspects	23,597	3,371	6,742
Average number of aspects per sentence	3.3	3.2	3.3
Average length per sentence	36.4	35.1	36.3

Table 4.2: Bảng thống kê tổng quan trên train/val/test set

Aspect	Train			Dev			Test			Total
	Pos	Neu	Neg	Pos	Neu	Neg	Pos	Neu	Neg	
BATTERY	2,027	349	1,228	303	51	150	554	92	368	5,122
CAMERA	1,231	288	627	172	36	88	346	71	171	3,030
DESIGN	999	77	302	135	12	40	274	28	96	1,963
FEATURES	785	198	1,659	115	33	233	200	52	459	3,734
GENERAL	3,627	290	949	528	34	127	1,004	83	294	6,936
PERFORMANCE	2,253	391	1,496	327	45	210	602	116	454	5,894
PRICE	609	391	316	72	144	36	162	328	79	2,882
SCREEN	514	56	379	62	12	47	136	17	116	1,339
SER&ACC	1,401	107	487	199	13	78	199	27	167	2,678
STORAGE	59	107	21	11	1	2	18	3	6	132
Total	13,505	2,903	7,464	1,924	381	1,011	3,495	817	2,210	

Table 4.3: Bảng phân phối của aspect và polarity trên train/test/val set

Do aspect OTHERS, index, date\_time, n\_star không có nhiều giá trị đối với mục tiêu của nhóm, nên chúng tôi xin bỏ các thành phần này ra khỏi dataset.

Một vài đặc điểm của dữ liệu:

- Có khá nhiều từ viết tắt, sai chính tả  $\Rightarrow$  Sử dụng từ điển teencode chứa các từ viết tắt thông dụng đồng thời bổ sung các từ viết tắt nhóm quan sát được trên tập train để ánh xạ về từ đúng chính tả.

VD1: "Mới mua hồi sáng về dùng đc một lúc màn hình bị đơ , lag, pin tụt nhanh khoảng **5p** cho 1% pin khi **k** sử dụng bất kì ứng dụng nào trong khi đó **mk** để ở chế độ tiết kiệm pin , được trang bị sạc nhanh nhưng sạc vẫn chậm

hơn 2 tiếng cho một lần sạc đầy từ 20% - 100%, mic lúc thì nói **ngk** nghe dc lúc thì **k** 😞"

VD2: “Thậm tệ hơn là **vao wep** tải dữ liệu quá kém.**man hình** trắng bộp.ko có gì. Phải kéo xuống 1 cái **mí** tải **dc wep**”

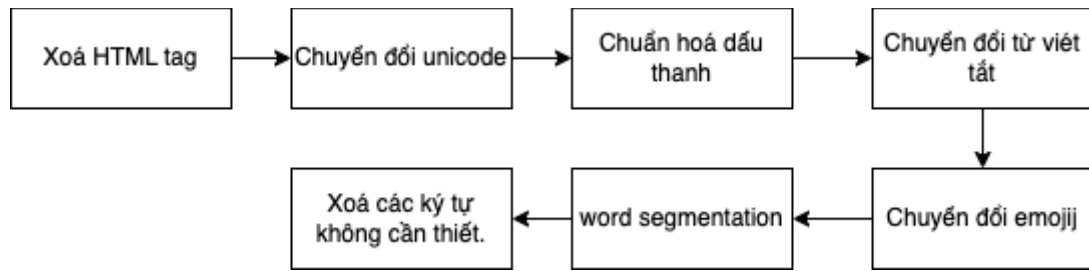
- Có sử dụng icon  $\Rightarrow$  Sử dụng từ điển để ánh xạ icon theo thành từ có nghĩa tương ứng, sau đó xoá những icon không có trong từ điển này.
- Các từ gồm nhiều tiếng chưa được liên kết với nhau  $\Rightarrow$  Sử dụng word segmentation của VnCoreNLP.
- Các ký tự đặc biệt không có ý nghĩa  $\Rightarrow$  Xoá các ký tự không cần thiết

VD1: “Loa trong bị lỗi không nghe dc chỉ nghe được loa ngoài thôi~~~~~”

VD2: Cảm ứng sản phẩm kém. Không kiểm tra thiết bị \*#0\*# được. Như các dòng cảm ứng đời đầu và hay bị treo máy. Tôi mới mua 2 tháng. Với mức 3 triệu thì không xứng với tiền bỏ ra

- Dữ liệu được xử lý sạch các nhãn html, nhưng cũng có thể còn sót lại vài chỗ, và khi demo sản phẩm, người dùng có thể dùng những data crawl từ những trang web nhưng chưa loại bỏ hoàn toàn các tag html, nên nhóm cũng sẽ dùng thêm thao tác để xoá đi những tag html này trước khi đưa vào model.
- Chuyển đổi các ký tự unicode 1252 sang unicode utf-8, chuẩn hóa cách từ theo quy tắc đặt dấu thanh của chữ Quốc ngữ kiểu cũ để tránh những sai khác không cần thiết giữa các từ trừ ý nghĩa của chúng.

Từ các đặc điểm trên, nhóm sẽ đưa ra preprocess pipeline như sau:



*Figure 4.2: Preprocess pipeline*



## 5. TRAIN VÀ ĐÁNH GIÁ MODEL

### 5.1. Hướng tiếp cận và chọn model để huấn luyện

Trong bài báo cáo đồ án này, chúng tôi sẽ giải quyết vấn đề trên hai nhiệm vụ chính là: Aspect Category Detection (ACD) and Sentiment Polarity Classification (SPC). Chúng tôi quyết tiếp cận bài toán bằng việc sử dụng phương pháp Deep Learning với việc chạy thực nghiệm dựa trên các kỹ thuật như: Bi-LSTM (Bidirectional Long Short-Term Memory) và PhoBERT làm mô hình ngôn ngữ được tiền huấn luyện cho tiếng Việt theo hai cách: Multi-task and Multi-task with Multi-branch approach.

Đối với kỹ thuật Bi-LSTM: Dựa trên việc tìm tòi nghiên cứu bài toán ABSA trên các lĩnh vực khác, chúng tôi khá ấn tượng đối việc sử dụng kỹ thuật Bi-LSTM để phân loại ngữ nghĩa và khía cạnh trên sản phẩm son môi của công trình nghiên cứu [4]. Trong bài báo liên quan được công bố, model Bi-LSTM cho ra kết quả khá cao: 94.12% F1-score trên aspect detection và 81.44% F1-score trên sentiment classification. Đồng thời, việc cài đặt model Bi-LSTM cũng không phải là không thể đối với chúng tôi nên chúng tôi quyết định chọn Bi-LSTM là một trong 3 phương pháp để giải quyết và đánh giá trên bài toán này. Nhóm đã tham khảo sourcecode [5] của nhóm tác giả.

Đối với kỹ thuật PhoBERT: Cũng dựa trên việc tìm tòi nghiên cứu bài toán ABSA trên các lĩnh vực khác, chúng tôi phát hiện bài nghiên cứu "Multi-task Solution for Aspect Category Sentiment Analysis on Vietnamese Datasets" [6]. Trong bài nghiên cứu này, tác giả đã đề xuất phương pháp sử dụng PhoBERT làm mô hình ngôn ngữ được tiền huấn luyện cho tiếng Việt theo hai cách: Multi-task and Multi-task with Multi-branch approach để đồng thời xử lý trên cả hai nhiệm vụ ACD và SPC trên lĩnh vực Nhà Hàng và Khách Sạn. Mô hình với multi-task approach đạt được kết quả rất tốt khi được áp dụng tiền xử lý và được tác giả khẳng định là mô hình đạt được kết quả hàng đầu (SOTA) trong lĩnh vực Khách Sạn và Nhà Hàng của tập dữ liệu ABSA

VLSP 2018, ngoài ra mô hình còn lại cũng đạt được kết quả khả quan khi xấp xỉ với mô hình multi-task approach và cao hơn best submit trên bộ dataset Restaurant. Kết quả so sánh ở trong bảng 3. Vì vậy, chúng tôi quyết định áp dụng cả hai cách sử dụng PhoBERT này để áp dụng trên bộ dataset UIT-ViSFD trong đề án này. Nhóm đã tham khảo sourcecode [7] của nhóm tác giả

Task	Method	Hotel			Restaurant		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Aspect Detection	VLSP best submission	76.00	66.00	70.00	79.00	76.00	77.00
	Bi-LSTM+CNN	84.03	72.52	77.85	82.02	77.51	79.70
	BERT-based Hierarchical	-	-	82.06	-	-	<b>84.23</b>
	Multi-task	<b>87.45</b>	<b>78.17</b>	<b>82.55</b>	81.09	85.61	83.29
	Multi-task Multi-branch	63.21	57.86	60.42	80.81	87.39	83.97
Aspect + Polarity	VLSP best submission	66.00	57.00	61.00	62.00	60.00	61.00
	Bi-LSTM+CNN	76.53	66.04	70.90	66.66	63.00	64.78
	BERT-based Hierarchical	-	-	74.69	-	-	71.30
	Multi-task	<b>81.90</b>	<b>73.22</b>	<b>77.32</b>	<b>69.66</b>	<b>73.54</b>	<b>71.55</b>
	Multi-task Multi-branch	57.55	52.67	55.00	68.69	74.29	71.38

Table 5.1: Kết quả so sánh 2 phương pháp dựa trên PhoBERT trên 2 dataset trong hội nghị VLSP 2018 (Nguồn: [7])

## 5.2. Tổng quan về Bi-LSTM, Phobert-based method

### 5.2.1. Phobert-based method

#### 5.2.1.1. PhoBERT [8]

PhoBERT là một mô hình ngôn ngữ đơn ngôn ngữ quy mô lớn được tiền huấn luyện cho tiếng Việt. PhoBERT vượt trội hơn so với các phương pháp đơn ngôn ngữ và đa ngôn ngữ trước đó, đạt được hiệu suất mới nhất trên bốn tác vụ xử lý ngôn ngữ

tự nhiên tiếng Việt: gán nhãn từ loại, phân tích phụ thuộc, nhận dạng thực thể và suy luận ngôn ngữ tự nhiên.

PhoBERT được phân thành hai phiên bản: PhoBERTbase và PhoBERTlarge. Phương pháp tiền huấn luyện của PhoBERT dựa trên RoBERTa, tối ưu quy trình tiền huấn luyện BERT để đạt được hiệu suất ổn định hơn.

### 5.2.1.2. Kiến trúc Model

### 5.2.1.3. Multi-task Approach

Mô hình được xây dựng với đầu ra là một danh sách gồm C vector one-hot, trong đó C là số khía cạnh hiện có trong tập dữ liệu được sử dụng. Đối với tập dữ liệu VLSP, số khía cạnh C sẽ là 12 cho tập dữ liệu Điện thoại thông minh. Mỗi vector có bốn thành phần tương ứng với ba nhãn của các chiều cực, bao gồm Positive (Tích cực), Negative (Tiêu cực), Neutral (Trung lập), và một nhãn None để chỉ ra liệu đầu vào có chứa khía cạnh đó hay không để có thể có một chiều cực. Vì đây là một vector one-hot, bất kỳ nhãn nào tồn tại sẽ được gán giá trị 1, và các nhãn còn lại sẽ có giá trị 0.

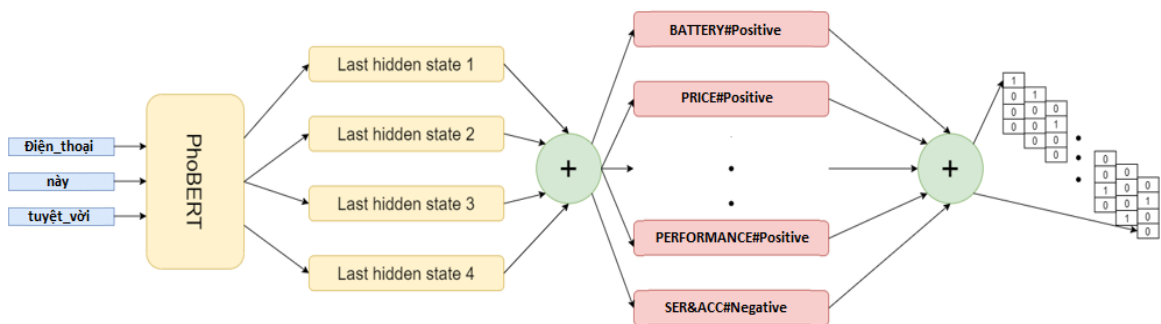


Figure 5.1: PhoBERT Multi-task Approach Architecture

Đặc trưng đã học (g) được truyền qua một tầng kết nối đầy đủ được tạo bởi việc ghép C tầng mật độ tương ứng với C vector one-hot. Do đó, chúng ta sẽ có một tầng mật độ gồm 30 neuron cho lĩnh vực Điện thoại di động (10 khía cạnh  $\times$  3 cực

tính), và sử dụng hàm softmax để tính điểm cho giá trị dự đoán  $y_{\text{pred}}$  cho mỗi khía cạnh A:

$$y_{\text{pred}} = \text{softmax}(W^{(a)} \cdot g + b^{(a)}).$$

Vì vậy, chúng ta có thể dự đoán một khía cạnh  $a$  và tính chất của nó trong một bước bằng cách:

$$\text{output}^{(a)} = \arg \max_i \hat{y}_i^{(a)} \quad \text{where } i = 0, 1, 2, 3.$$

Đối với vấn đề phân loại nhiều lớp ( $C > 2$ ), hàm mất mát sẽ là hàm cross-entropy như sau:

$$L(y, \hat{y}) = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \cdot \log(\hat{y}_{ij})$$

#### 5.2.1.4. Multi-task with Multi-branch Approach

Sự khác biệt duy nhất của phương pháp này so với phương pháp trên là nó sẽ phân nhánh thành nhiều mô hình con bằng cách sử dụng  $C$  lớp kết nối đầy đủ riêng biệt mà không ghép chúng thành một lớp duy nhất. Mỗi mô hình dự đoán mỗi nhiệm vụ một cách độc lập.

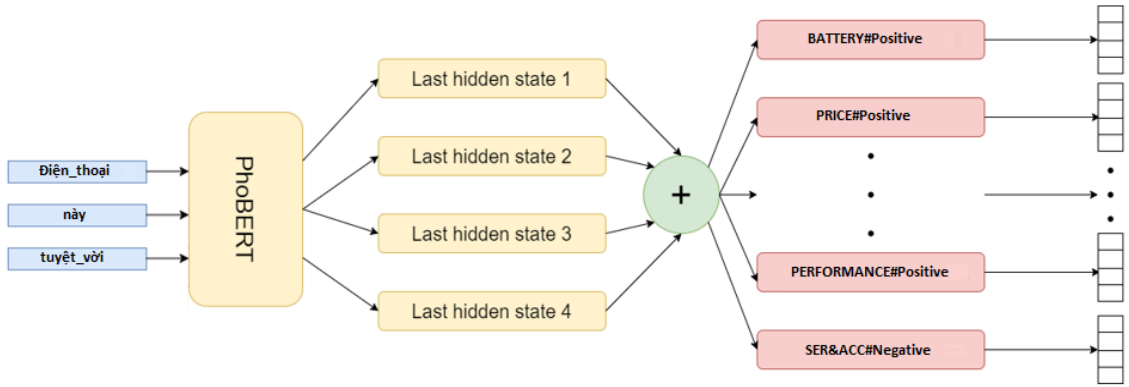


Figure 5.2: PhoBERT Multi-task with Multi-branch Approach Architecture

Mô hình sử dụng hàm softmax để tính toán các điểm số và có thể dự đoán đồng thời khía cạnh và tính chất của nó một cách tương tự như mô hình trên. Hàm mất mát cuối cùng sẽ bằng tổng các hàm mất mát trong tất cả các nhánh C:

$$L(W; X) = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \cdot \log(\hat{y}_{ij}).$$

### 5.2.2. Bi-LSTM

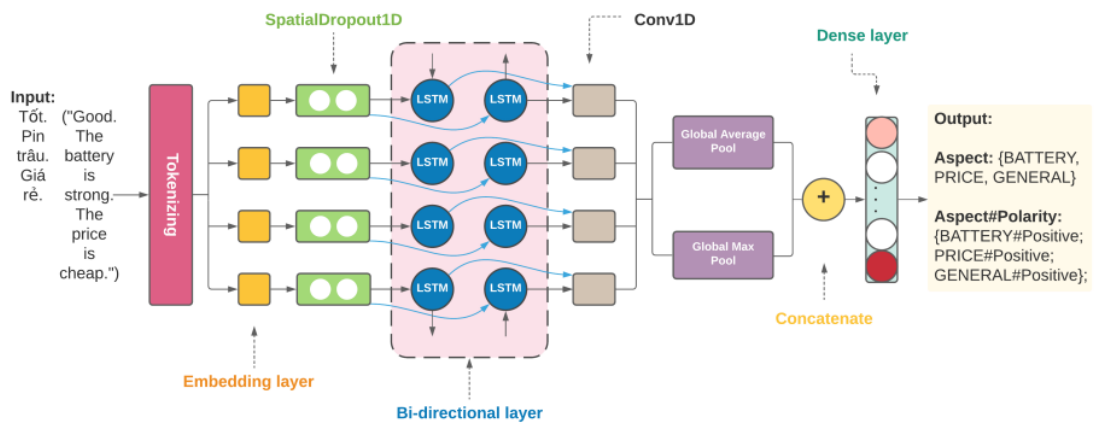


Figure 5.3: Bi-LSTM Architecture

Được truyền cảm hứng từ Bi-LSTM cho phân loại văn bản, chúng tôi đề xuất một phương pháp sử dụng mô hình Bi-LSTM cho ABSA tiếng Việt. Kiến trúc tổng quan được mô tả trong hình trên.

Kiến trúc này bao gồm một bộ phân tách từ (tokenizer), lớp nhúng (embedding layer), lớp SpatialDropout1D, lớp Bi-LSTM, lớp tích chập (convolutional layer), hai lớp pooling và một lớp dense. Đầu tiên, nhận xét sẽ trải qua bộ phân tách từ, chuyển đổi mỗi từ trong nhận xét thành một giá trị nguyên dựa trên chỉ mục từ vựng. Sau đó, chúng được xử lý thông qua lớp nhúng để chuyển đổi thành các vector đại diện. Kiến trúc sử dụng nhúng fastText làm đại diện đầu vào cho các từ. fastText có các đại diện từ tốt và mã hóa cho các từ hiếm không xuất hiện trong quá trình huấn luyện, đây là lựa chọn tốt cho các văn bản trên mạng xã hội tiếng Việt. Để giảm thiểu hiện tượng quá khớp, chúng tôi sử dụng SpatialDropout1D để giảm số lượng tham số sau mỗi lần huấn luyện. Chúng tôi sử dụng một lớp Bi-LSTM để trích xuất các đặc trưng trừu tượng, lớp này được tạo thành từ hai LSTM với kết quả đầu ra của chúng được xếp chồng lên nhau. Nhận xét được đọc tuần tự từ trái sang phải bởi một LSTM và từ phải sang trái bởi LSTM kia. Chúng tôi ghép các trạng thái ẩn của mỗi LSTM sau khi chúng đã xử lý từ cuối cùng của mình. Bi-LSTM sử dụng hai đơn vị LSTM riêng biệt, một cho hướng thuận và một cho hướng ngược lại. Lớp tích chập được sử dụng để chuyển đổi ma trận đa chiều từ Bi-LSTM thành ma trận một chiều. Lớp pooling bao gồm hai lớp song song: lớp tổng trung bình toàn cục (global average pool) và lớp tổng cực đại toàn cục (global max pool). Chức năng của lớp pooling được thực hiện để giảm kích thước không gian của biểu diễn. Ý tưởng là chọn phần tử lớn nhất và phần tử trung bình của bản đồ đặc trưng để trích xuất các đặc trưng quan trọng nhất từ lớp tích chập. Cuối cùng, nhãn của hai nhiệm vụ phụ cho nhiệm vụ ABSA được thu được sau khi chuẩn hóa trong lớp dense.

Nhóm sử dụng Elmo embedding model pretrained on Vietnamese của [9] khởi tạo embedding layer của mô hình.

### 5.3. Đánh giá mô hình

Sau khi tiến hành training, chúng tôi tiến hành đánh giá các mô hình. Do dữ liệu có sự mất cân bằng giữa các aspect cũng như polarity, nếu chỉ dựa vào Accuracy sẽ không đánh giá chính xác. Chúng tôi sử dụng ba thang đo (metrics) chính là Precision, Recall và F1-score.

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

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

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

Do Aspect Category Detection (ACD) và Sentiment Polarity Classification (SPC) đều là các bài toán multi-class. Vì thế chúng tôi sử dụng macro-average cho Precision, Recall và F1-score để tiến hành đánh giá các mô hình. Đây là bảng kết quả:

Task	Method	Precision	Recall	F1-score
Aspect Detection	Bi - LSTM	0.6747	0.4491	0.4885
	Multi-task	0.8754	0.8855	0.8765
	Multi-task with Multi-branch	0.8664	0.8723	0.8501
Aspect + Polarity	Bi - LSTM	0.6252	0.3724	0.3833
	Multi-task	0.7036	0.6452	0.6377
	Multi-task with Multi-branch	0.6789	0.5847	0.5786

Table 5.2: Kết quả chạy thực nghiệm trên tất cả 3 phương pháp

Mô hình Bi-LSTM có kết quả thấp hơn so với hai mô hình sử dụng PhoBERT với f1-score chỉ là 0.4885 ở Aspect Detection và 0.3833 ở Aspect + Polarity, qua đó chúng tôi không dùng mô hình này để demo sản phẩm. Với hai mô hình còn lại, mô hình Multi-task cho hiệu quả tốt hơn ở cả hai task trên cả 3 thang đo. Qua đó cho thấy PhoBert-based Multi-task Approach là hướng tiếp cận tốt nhất. Ngoài ra, phương pháp này tốt hơn so với phương pháp tốt nhất trong bài báo công bố UIT - ViSFD (0.8685 macro-f1 với aspect detection và 0.6306 macro-f1 với aspect#polarity khi xét trên 10 aspect chính).



Đối với task Aspect Detection, có sự mất cân bằng dữ liệu giữa các aspects. Các aspects thiểu số như DESIGN, SCREEN, STORAGE có số lượng dữ liệu rất thấp khiến cho kết quả dự đoán của các mô hình không tốt.

Đối với task Aspect + Polarity, cũng có sự mất cân bằng giữa các nhãn khi positive chiếm tỷ lệ cao nhất và neutral chiếm tỷ lệ thấp nhất. Qua đó, dự đoán của các mô hình đối với nhãn neutral ở tất cả Aspect đều thấp. Ngoài ra, kết quả của task Aspect Detection cũng góp phần ảnh hưởng, bởi Aspect Detection là một phần của task thứ hai. Kết quả không tốt của các aspects tối thiểu khiến cho kết quả Recall trên Aspect + Polarity của chúng rất thấp.

Vì vậy sự mất cân bằng dữ liệu đã ảnh hưởng đến kết quả của các mô hình trên cả hai task. Đồng thời, kết quả task Aspect Detection cũng góp phần ảnh hưởng lên kết quả task Aspect + Polarity, dễ dàng nhận thấy khi kết quả của task thứ hai thấp hơn so với task thứ nhất.

## 6. THÁCH THỨC

Từ góc độ ngôn ngữ học, nhiệm vụ tự động xác định các khía cạnh và tình cảm liên quan đặt ra nhiều thách thức, bởi vì chúng ta thường thấy các hiện tượng ngôn ngữ phức tạp không dễ diễn giải và hiểu. Đặc biệt, xác định đúng khía cạnh có thể được coi là nhiệm vụ khó nhất của phân tích này, bởi vì khách hàng có thể bày tỏ ý kiến của họ về rất nhiều khía cạnh (ví dụ: đối với một sản phẩm: giá cả, chất lượng, hiệu suất hoặc thiết kế), tùy thuộc vào tình huống, trong khi thực hiện phân tích cảm tính dựa trên khía cạnh, chúng tôi gặp phải những thách thức sau:

- Khi tình cảm phụ thuộc vào mục tiêu:

Trong khi mô tả sản phẩm hoặc dịch vụ, khách hàng cũng có thể tham chiếu đến các thực thể khác. Ví dụ: "Sản phẩm này có giá tốt; nhưng cái mà anh tôi mua có thiết kế đẹp hơn." Trong ví dụ, nó không đủ để mô hình có thể truy xuất tất cả các khía cạnh được đề cập trong văn bản và cảm xúc liên quan của chúng (PRICE#POSITIVE và DESIGN#POSITIVE). Điều quan trọng là mô hình chỉ có thể chọn cặp price#positive, vì đó là cặp liên quan đến sản phẩm đang được đề cập, Không nên chọn cặp design#positive vì nó đang đề cập đến một thực thể khác. Từ quan điểm tổng quát hơn, mô hình cần có khả năng phân biệt “mục tiêu” (khía cạnh chính của phân tích hiện tại) với các thực thể khác.

- Khi tình cảm không mang nghĩa đen:

Cho đến nay, chúng tôi đã thấy các ví dụ về văn bản phản hồi trong đó cảm xúc được truyền đạt ngầm (chỉ hiểu được nhờ vào hiểu biết về sản phẩm), được truyền đạt rõ ràng (nhưng đôi khi không đề cập đến mục tiêu) không nên hiểu theo nghĩa đen. "100% pin về 0 sau 3h sử dụng, thật sự tuyệt vời!" Mặc dù bản thân cụm từ “thực sự tuyệt vời!” rõ ràng là tích cực, nhưng ngữ cảnh cho thấy khách hàng muốn truyền đạt cảm xúc ngược lại, do đó đưa ra phản hồi tiêu

cực. Những trường hợp này phụ thuộc nhiều vào ngữ cảnh và sự giải thích của con người.

- Do bộ dữ liệu thu thập được có sự mất cân bằng khá lớn giữa các aspect cũng như giữa các aspect#polarity, nên sẽ gây khó khăn cho các mô hình machine learning để dự đoán tốt trên các nhóm ít mẫu dữ liệu.

## 7. ỨNG DỤNG VÀ HƯỚNG PHÁT TRIỂN

### 7.1. Ứng dụng

Trong bài báo cáo này, chúng tôi sử dụng thư viện “streamlit” của Python để tạo giao diện cho sự tương tác của người dùng.

Người dùng có thể chọn nhập 1 câu văn bản hoặc 1 bộ dataset dưới dạng .csv để dự đoán. Sau đó người dùng chỉ cần nhấn nút để hiển thị kết quả phân tích ngữ nghĩa và một số biểu đồ tròn thống kê số lượng.

Source code của demo ở [đây](#).

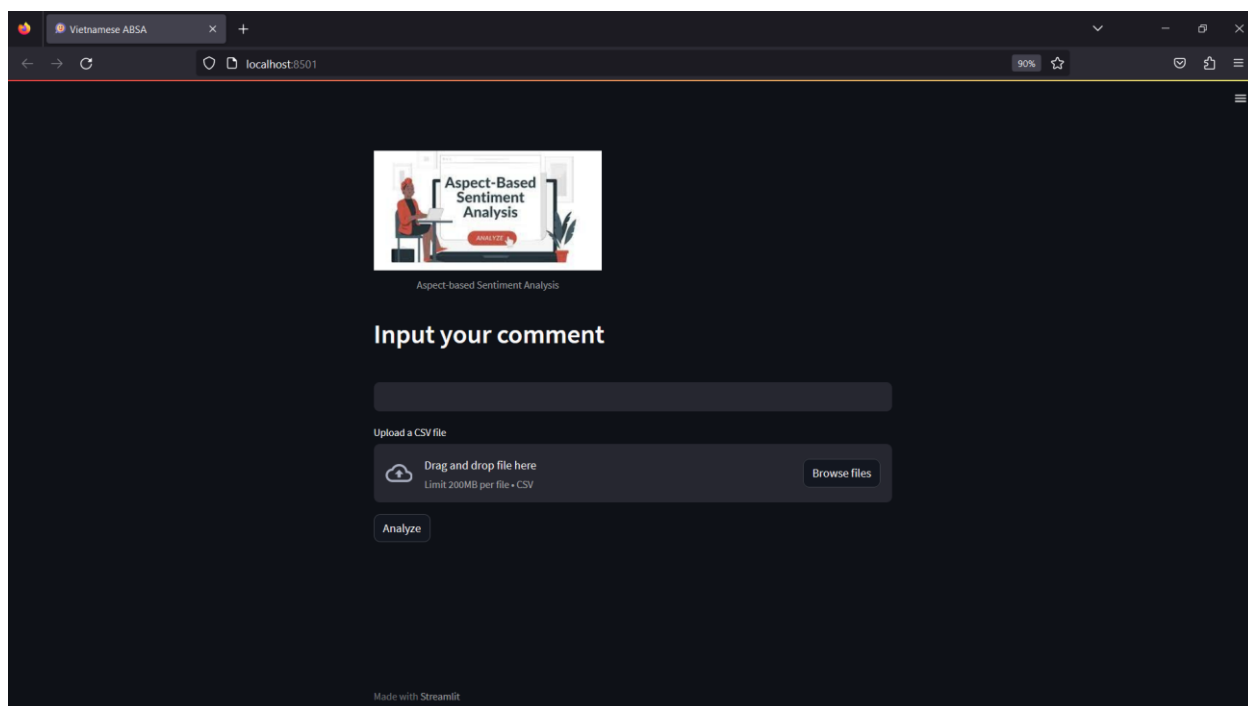


Figure 7.1: Local Web demo sử dụng Sreamlit

### 7.2. Hướng phát triển trong tương lai

Để mô hình được ứng dụng rộng rãi trong thực tế, ta cần cải tiến một số mặt sau:

- Về thuật toán: Ứng dụng các phương pháp khác cho vấn đề này như các mô hình: Graph models, BERT variants. Thay đổi Embedding classes for multi-task model: PhoW2V, fastText, multi Embedding, vELECTRA.
- Về data: Sử dụng các phương pháp tăng cường dữ liệu cho tập huấn luyện như thêm các data mới chất lượng, dịch ngôn ngữ, viết lại câu, tương đồng từ vựng, v.v... để tăng hiệu quả của mô hình. Đặc biệt cần chú ý tăng cường các aspect và polarity có số lượng mẫu dữ liệu thấp để mô hình học được đầy đủ hơn về các mặt này.

Ngoài ra, chúng tôi sẽ cố gắng xây dựng 1 mô hình End-to-End ABSA platform. Mục đích của việc xây dựng platform này để tạo ra một Dashboard phục vụ cho việc phân tích ngữ nghĩa dựa trên cảm xúc và tiếp cận gần hơn đối với việc ứng dụng thực tiễn của bài phân tích này vào lĩnh vực kinh doanh.

## 8. TÀI LIỆU THAM KHẢO

### 8.1. Tài liệu tiếng Việt

- [1] V.T.T.Lan, “Quy tắc đặt dấu thanh trong tiếng Việt”, 22/08/2017. [Trực tuyến]. Địa chỉ: <https://vndoc.com/quy-tac-dat-dau-thanh-trong-tieng-viet-129192>
- [2] vietnambiz.vn, “Lượng người dùng smartphone tại Việt Nam ước đứng thứ hai Đông Nam Á vào năm 2026, thuộc top nhiều nhất thế giới”, 09/05/2023, [Trực tuyến]. Địa chỉ: <https://vietnambiz.vn/luong-nguoi-dung-smartphone-tai-viet-nam-uoc-dung-thu-hai-dong-nam-a-vao-nam-2026-thuoc-top-nhieu-nhat-the-gioi-202359104330221.htm>

### 8.2. Tài liệu tiếng Anh

- [3] P.L.Luong, P.H.Phuc, N.T.T.Kim ... N.V.Kiet, “SA2SL: From Aspect-Based Sentiment Analysis to Social Listening System for Business Intelligence”, *The 14th International Conference on Knowledge Science, Engineering and Management (KSEM 2021), Tokyo, Japan, August 14-16, 2021*, Place of Publication: Springer, 2021, 647–658.
- [4] T.Q.Linh, L.D.T.Phan and D.T.Hop, “Aspect-based Sentiment Analysis for Vietnamese Reviews about Beauty Product on E-commerce Websites”, *the 36th Pacific Asia Conference on Language, Information and Computation (PACLIC 2022), Manila, Philippines, October 20-22, 2022*, Place of Publication: De La Salle University, 2022, 767–776.
- [5] T.Q.Linh, “Aspect-based Sentiment Analysis for Vietnamese Reviews about Beauty Product on E-commerce Websites”, May 16, 2023. [Online]. Available: <https://github.com/linh222/Aspect-based-Sentiment-Analysis-for->

- [6] D.H.Quan, N.D.D.Anh and D.T.Hop, “Multi-task Solution for Aspect Category Sentiment Analysis on Vietnamese Datasets”, *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia, 16-18 June 2022*, Place of Publication: IEEE, 2022.
- [7] D.H.Quan, “Aspect-based Sentiment Analysis for Vietnamese”, August 28, 2022. [Online]. Availabe: <https://github.com/ds4v/absa-vlsp-2018> [Accessed July 2, 2023].
- [8] N.Q.Dat, “PhoBERT: Pre-trained language models for Vietnamese”, April 24, 2023. [Online]. Availabe: <https://github.com/VinAIRsearch/PhoBERT> [Accessed July 2, 2023].
- [9] V.X.Son, Thanh Vu, Son N. Tran and Lili Jiang, “ETNLP: A Toolkit for Extraction, Evaluation and Visualization of Pre-trained Word Embeddings”, October 17, 2022. [Online]. Availabe: <https://github.com/vietnlp/etnlp> [Accessed July 2, 2023].