

# Phân Tích Cảm Xúc Dựa Trên Khía Cạnh Của Phản Hồi Khách Hàng Về Khách Sạn

Trần Quang Duy<sup>1,2,3</sup>, Trần Thị Mỹ Duyên<sup>1,2,3</sup>, Trần Quốc Khánh<sup>1,2,4</sup>, Nguyễn Gia Tuấn Anh<sup>1,2,4</sup>

<sup>1</sup> University of Information Technology, Ho Chi Minh City, Vietnam

<sup>2</sup> Vietnam National University, Ho Chi Minh City, Vietnam

<sup>3</sup>{21522013, 21522017}@gm.uit.edu.vn

<sup>4</sup>{khanhtq, anhnghgt}@uit.edu.vn

## Abstract

Trong những năm gần đây phân tích cảm xúc dựa trên khía cạnh (Aspect-Based Sentiment Analysis - ABSA) đã thu hút được nhiều sự quan tâm nghiên cứu và ứng dụng. Tuy nhiên, đây là một chủ đề còn mới và được quan tâm gần đây tại Việt Nam. Trong nghiên cứu này chúng tôi xây dựng, phát triển và đánh giá bộ dữ liệu trên tiếng Việt với chất lượng cao. Bộ dữ liệu bao gồm hơn 8,000 bình luận và hơn 36,000 khía cạnh từ các phản hồi của khách hàng về khách sạn trên nền tảng đặt phòng trực tuyến Traveloka. Và thử nghiệm bộ dữ liệu trên các mô hình máy học như Logistic Regression, Multinomial Naive Bayes, SVM và các mô hình học sâu như LSTM và Bi-LSTM. Và thu được kết quả tốt nhất trên mô hình Bi-LSTM với  $F1_{score}$  71.21%, các mô hình học máy cũng cho kết quả khá tốt như Linear SVC với  $F1_{score}$  69.96%. Ngoài ra chúng tôi cũng đặt ra các thách thức cần phải giải quyết tiếp theo để cải thiện chất lượng bộ dữ liệu và nâng cao hiệu quả mô hình trong tương lai.

## 1 Giới Thiệu

Trong những năm gần đây, sự phát triển mạnh mẽ của các nền tảng đặt phòng trực tuyến đã làm chúng trở thành một phần không thể thiếu trong ngành du lịch và khách sạn. Những nền tảng này đã giúp cho những doanh nghiệp có lượng lớn dữ liệu thu từ các phản hồi của khách hàng. Theo số liệu từ nghiên cứu (Blomberg-Nygaard and Anderson, 2016), trong 80% khách du lịch tìm kiếm khách sạn trên trang web, sẽ có hơn 50% trong số họ đặt phòng qua trang web hoặc ứng dụng. Do đó nếu khách sạn đưa ra các đánh giá của các khách hàng trước đó theo khía cạnh mà khách hàng quan tâm sẽ giúp khách hàng tìm kiếm dịch vụ thuận lợi hơn, cung cấp cho khách hàng những thông tin cần thiết khách quan nhất có thể.

Tuy nhiên các đánh giá từ khách hàng thường chứa đựng nhiều khía cạnh khác nhau như chất lượng dịch vụ, phòng ngủ, các tiện nghi của khách

sạn, etc. Một khách hàng có thể thích chất lượng của phòng nhưng không thích thái độ phục vụ của khách sạn. Vì thế nên chúng tôi sử dụng phân tích cảm xúc dựa trên khía cạnh bằng cách đi phân tích cảm xúc tương ứng với từng khía cạnh chứa trong bình luận của khách hàng.

Với những lý do trên, chúng tôi quyết định đi xây dựng một bộ dữ liệu trên tiếng Việt phục vụ cho việc phân tích cảm xúc dựa trên khía cạnh bình luận khách hàng về khách sạn và áp dụng các mô hình máy học để phục vụ cho việc tự động xác định các khía cạnh bình luận của khách hàng. Mục tiêu là để cung cấp cho các nhà quản lý khách sạn thông tin chi tiết và chính xác về những khía cạnh mà khách hàng đánh giá cao hoặc không hài lòng. Đồng thời, bộ dữ liệu cũng giúp khách hàng có được những thông tin cần thiết về khía cạnh mà họ quan tâm, từ đó họ có thể lựa chọn khách sạn phù hợp nhu cầu và mong đợi của mình.

Trong bài báo cáo này, chúng tôi tập trung xây dựng một bộ dữ liệu chất lượng cao, có tính cập nhật để phục vụ cho bài toán phân tích cảm xúc dựa trên khía cạnh với lĩnh vực khách sạn, trình bày các phương pháp tiếp cận của bài toán và phân tích lỗi để ra hướng phát triển cho bộ dữ liệu.

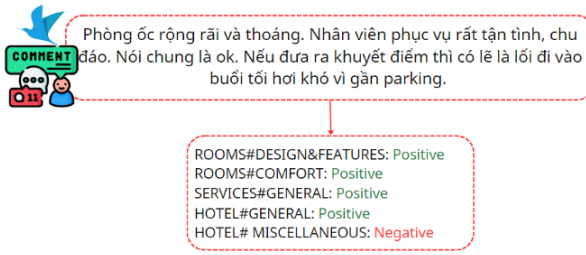
Phần còn lại của báo cáo được tổ chức như sau. Trong phần 2 chúng tôi trình bày về định nghĩa bài toán. Tiếp theo là phần 3 trình bày về các công trình liên quan đến bài toán, phần 4 trình bày chi tiết về quy trình xây dựng bộ dữ liệu. Ở phần 5 và phần 6 chúng tôi trình bày về hướng tiếp cận bài toán và kết quả các thử nghiệm. Cuối cùng ở phần 7 đưa ra kết luận và hướng phát triển cho bài toán.

## 2 Định Nghĩa Bài Toán

Mục tiêu bài toán của chúng tôi được mô tả ở hình 1 với đầu vào  $X$  là phản hồi của khách hàng về khách sạn, đầu ra là  $\{y_1, y_2, \dots, y_n\}$  với  $y$  là cảm xúc của từng khía cạnh trong phản hồi khách hàng.

**Input:** Một phản hồi của khách hàng  $A$  về khách sạn sau khi sử dụng dịch vụ của khách sạn  $B$ .

**Output:** Là một hoặc nhiều khía cạnh của khách sạn *B* đi kèm với cảm xúc của khách hàng *A* về khía cạnh đó.



Hình 1: Mô Tả Bài Toán

Bảng 1: Các thực thể và khía cạnh được xác định

	General	Prices	Design&Features	Cleanliness	Comfort	Quality	Style&Options	Miscellaneous
Hotel	X	X	X	X	X	X		X
Rooms	X	X	X	X	X	X		X
Room&Amenities	X	X	X	X	X	X		X
Facilities	X	X	X	X	X	X		X
Service	X	X	X	X	X	X		X
Location	X							
Food&Drinks		X				X	X	X

Các khía cạnh được tạo nên từ 07 thực thể mô tả về khách sạn như *Hotel*, *Rooms*, *Rooms&Amenities*, *Facilities*, *Service*, *Location*, *Food&Drinks* kết hợp cùng với 08 thuộc tính của các thực thể trên như *General*, *Prices*, *Design&Features*, *Cleanliness*, *Comfort*, *Quality*, *Style&Options*, *Miscellaneous* và tạo thành 34 khía cạnh tương ứng như trong bảng 1. Tương ứng với mỗi khía cạnh chúng tôi đánh giá trên 03 cảm xúc *Positive*, *Negative* và *Netrual*.

### 3 Các công trình nghiên cứu liên quan

Chủ đề phân tích cảm xúc dựa trên khía cạnh những năm gần đây đã thu hút nhiều sự chú ý. Để tìm ra giải pháp tốt nhất cho vấn đề này, một số hội thảo trong và ngoài nước như SemEval 2014 (Pontiki et al., 2014), SemEval 2015 (Pontiki et al., 2015), SemEval 2016 (Pontiki et al., 2016) và VLSP 2018 (Nguyen et al., 2019) đã tổ chức các nhiệm vụ cho bài toán ABSA.

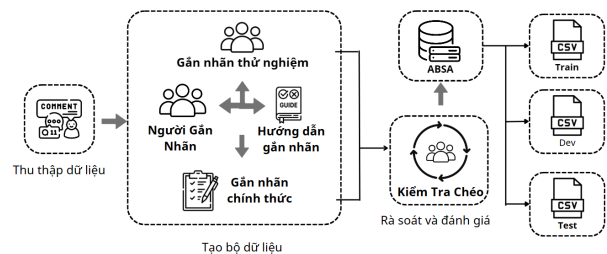
Trên tiếng Việt cũng có nhiều nghiên cứu cho ra các bộ dữ liệu về phân tích cảm xúc dựa trên khía cạnh trong nhiều lĩnh vực khác nhau như nhà hàng,

khách sạn, e-commerce. Trong hội thảo VLSP-2018 (Nguyen et al., 2019) đã công bố bộ dữ liệu SA-VLSP 2018 để phân tích cảm xúc dựa trên khía cạnh về lĩnh vực khách sạn và nhà hàng. Ngoài ra còn có (Thanh et al., 2021) đã sử dụng phát hiện đoạn văn cho phân tích cảm xúc dựa trên khía cạnh trên bộ dữ liệu UIT-ViSD4SA trên lĩnh vực điện thoại thông minh, (Tran et al., 2022) tiến hành đánh giá của người Việt về các sản phẩm làm đẹp (Son) trên nền tảng Shopee, (Nguyen et al., 2018) đã công bố bộ dữ liệu UIT-VSFC gồm hơn 16,000 câu phản hồi từ sinh viên, được sử dụng để phân loại cảm xúc và phân loại chủ đề.

Các hướng tiếp cận của bài toán cũng đa dạng từ các phương pháp máy học truyền thống, phương pháp học sâu cho đến phương pháp học chuyển tiếp. Trên các bộ dữ liệu tiếng Việt như SA-VLSP 2018 các đội đã sử dụng các phương pháp Linear SVM, Naive Bayes, Logistic Regression, Xgboost, (Dang et al., 2022) đã sử dụng mô hình Bert, PhoBert để giải quyết bài toán. Ngoài ra còn có các mô hình học sâu để giải quyết bài toán như BiLSTM, Bi-GRU, LSTM được (Tran et al., 2022) và (Luc Phan et al., 2021). Vì thế trong bài toán này chúng tôi với mục tiêu xây dựng bộ dữ liệu về phân tích cảm xúc dựa trên khía cạnh và thử nghiệm bài toán trên các mô hình học máy truyền thống và các mô hình học sâu.

## 4 Xây Dựng Bộ Dữ Liệu

Hình 2 mô tả quy trình xây dựng bộ dữ liệu của chúng tôi. Quy trình trải qua 03 bước chính gồm: Thu thập dữ liệu; Xây dựng, phát triển guideline và gắn nhãn; Rà soát đánh giá lại bộ dữ liệu.



Hình 2: Quy trình xây dựng bộ dữ liệu

### 4.1 Thu Thập Dữ Liệu

Có nhiều nền tảng đặt phòng trực tuyến tại Việt Nam hiện nay như: Agoda, Booking.com, Traveloka. Chúng tôi đã quyết định thu thập dữ liệu

trên nền tảng đặt phòng trực tuyến Traveloka<sup>1</sup>, một trong những nền tảng hỗ trợ đặt phòng phổ biến tại Việt Nam và là nơi chứa dữ liệu của 16,697 khách sạn tại Việt Nam. Và hỗ trợ cho việc crawl dữ liệu từ các trang web của chúng tôi là thư viện Selenium - một trong những thư viện hỗ trợ mạnh mẽ của Python<sup>2</sup>.

Thư viện này thích hợp cho những trang web sử dụng “Lazy loading”, đây là kỹ thuật tối ưu hóa cho các trang web khi người dùng cần tối đối tượng thì đối tượng mới được khởi tạo. Còn cung cấp nhiều các phương thức, thuộc tính để tăng tính linh và tùy chỉnh trong crawl dữ liệu. Chúng tôi tiến hành thu thập các thuộc tính review, hotel, hotel url của khách hàng từ trang web này.

## 4.2 Quy trình gắn nhãn

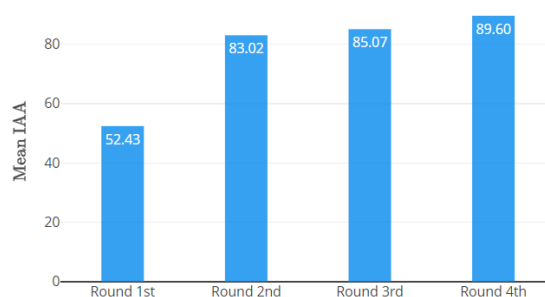
Trong quá trình xây dựng bộ dữ liệu cần một thước đo để đánh giá chất lượng của người gắn nhãn (Annotators). Độ đồng thuận gắn nhãn (Inter annotator agreement – IAA) là một độ đo nhằm đánh giá khả năng Annotator gắn cùng một giá trị (hay còn gọi là nhãn - label) cho một dữ liệu nhất định. Có rất nhiều độ đo để đánh giá IAA như  $F1_{score}$ , Cohen’s Kappa, Fleiss Kappa, Scotts. Với bộ dữ liệu này chúng tôi áp dụng độ đo  $F1_{score}$  cho việc đánh giá bộ dữ liệu.

**Giai đoạn 1:** Chúng tôi tiến hành xây dựng Guideline gắn mẫu để điều chỉnh Guideline cho phù hợp với bài toán và dữ liệu thu thập. Để làm quen với nhiệm vụ gắn nhãn, chúng tôi đã thực hiện thử nghiệm đánh nhãn trên một mẫu ngẫu nhiên gồm 50 điểm dữ liệu, với sự tham gia của cả hai người. Khi đạt được độ đồng thuận lớn hơn 85% với độ đo  $F1_{score}$  chúng tôi mới tiến hành gắn nhãn chính thức. Qua 04 lần thử nghiệm gắn nhãn, chúng tôi đã thu được kết quả như hình 6, cho thấy sự cải thiện rõ rệt qua từng lần thử nghiệm. Điều này đã tạo điều kiện cho chúng tôi tiếp tục thực hiện quá trình gắn nhãn chính thức một cách chính xác và hiệu quả.

**Giai đoạn 2:** Chúng tôi đã tiến hành quá trình gắn nhãn chính thức bằng cách phân chia công việc giữa hai người, mỗi người tiến hành gắn nhãn trên một tập dữ liệu riêng biệt và không trùng nhau.

**Giai đoạn 3:** Sau khi chúng tôi hoàn thành việc gắn nhãn chính thức sẽ tiến hành kiểm chéo những dữ liệu đã gắn nhãn để đảm bảo chất lượng của bộ dữ liệu, đối với các điểm dữ liệu không đảm bảo sẽ

tiến hành gắn nhãn lại.



Hình 3: Độ đồng thuận qua 04 lần thử nghiệm

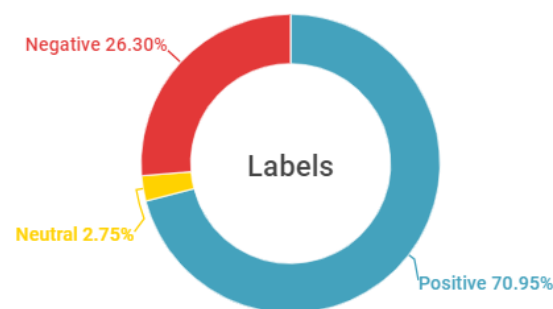
## 4.3 Thống kê tập dữ liệu

Bộ dữ liệu hoàn chỉnh chúng tôi gồm 8,009 điểm dữ liệu được phân chia vào 03 tập: training, development và test theo tỉ lệ 6:2:2. Thống kê sơ bộ về bộ dữ liệu của chúng tôi được trình bày tại bảng 2. Ngoài ra tỉ lệ phân phối nhãn của bộ dữ liệu như hình 5 ở các khía cạnh có nhiều nhãn nhất thì nhãn Positive chiếm số lượng nhiều hơn 02 nhãn còn lại, còn ở những khía cạnh có ít nhãn thì các nhãn Negative sẽ chiếm ưu thế hơn, cũng có thể thấy sự chênh lệch khá lớn số lượng nhãn giữa các khía cạnh.

Bảng 2: Thống kê trên bộ dữ liệu

	Training	Development	Test
<b>Review</b>	4805	1602	1602
<b>Avg review length</b>	34.5	35	34.8
<b>Max review length</b>	586	362	443
<b>Min review length</b>	2	2	3
<b>Total aspect</b>	22024	7377	7422

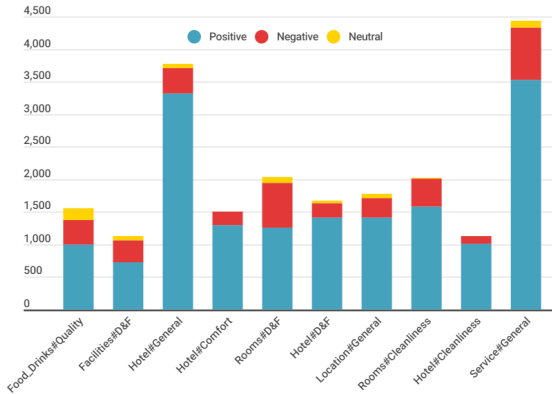
Ngoài ra, thông qua việc phân tích thống kê nhãn trong hình 7, chúng tôi đã nhận thấy tỉ lệ chênh lệch giữa các nhãn trong bộ dữ liệu này là khá lớn, đặc biệt là với nhãn cảm xúc “Neutral” chiếm tỉ lệ rất thấp.



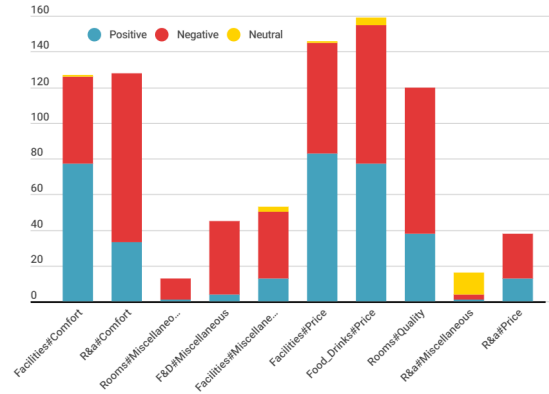
Hình 4: Thống kê nhãn cảm xúc của bộ dữ liệu

<sup>1</sup><https://www.traveloka.com/vi-vn/hotel>

<sup>2</sup><https://pypi.org/project/selenium/>



((a)) 10 khía cạnh có nhiều nhân nhiều nhất



((b)) 10 khía cạnh có nhiều nhân ít nhất

Hình 5: Thống kê nhân của các khía cạnh

Từ đó chúng tôi cũng thấy được thách thức của bộ dữ liệu là vừa mất cân bằng giữa các khía cạnh, vừa mất cân bằng giữa các nhân cảm xúc trong khía cạnh.

## 5 Phương Pháp Tiếp Cận

### 5.1 Tiền xử lý dữ liệu

Sau khi hoàn thành thu thập và gán nhãn, bước tiếp theo chúng tôi tiến hành tiền xử lý dữ liệu để làm sạch dữ liệu. Các kỹ thuật tiền xử lý được chúng tôi áp dụng thử nghiệm là: (1) Chuẩn hóa về dạng viết thường. *Ví Dụ:* “*NHân viên khó tíNH*” ⇒ “*nhân viên khó tính*”; (2) Chuẩn hóa Unicode về chuẩn UTF-8. (3) Chuẩn hóa cách gõ dấu tiếng Việt. *Ví Dụ:* “*Cứu hoả*” ⇒ “*Cứu hỏa*”; (4) Xóa các khoảng trắng và các kí tự dư thừa. (5) Chuẩn hóa teencode, từ viết tắt. *Ví Dụ:* “*Ks*”, “*Ksan*”, “*Hotel*” ⇒ “*Khách sạn*”; (6) Trong tiếng Việt một từ có thể tạo thành từ 02 từ đơn vì thế chúng tôi sử dụng thư viện Underthesea<sup>3</sup> để tách từ. *Ví Dụ:* “*Khách sạn tốt*” ⇒ “*Khách\_sạn\_tốt*”.

### 5.2 Vector Biểu Diễn Từ

Để máy có thể hiểu và học được từ dữ liệu chúng tôi cần đưa các điểm dữ liệu về dạng vector. Đối với các mô hình máy học chúng tôi thử nghiệm trên 02 phương pháp:

**Term Frequency – Inverse Document Frequency (TF-IDF):** là kỹ thuật sử dụng trong khai phá dữ liệu văn bản. Trọng số này được sử dụng để đánh giá tầm quan trọng của một từ trong một văn bản. Giá trị cao thể hiện độ quan trọng cao và nó phụ thuộc vào số lần từ xuất hiện trong văn bản

<sup>3</sup><https://github.com/undertheseanlp/underthesea>

nhưng bù lại bởi tần suất của từ đó trong tập dữ liệu.

**CountVectorizer (Cv):** là một kỹ thuật mã hóa được sử dụng để chuyển một văn bản nhất định thành một vectơ trên cơ sở tần suất của mỗi từ xuất hiện trong toàn bộ văn bản.

### 5.3 Word embedding

Để xây dựng lớp Word embedding cho các mô hình học sâu chúng tôi sử dụng fastText<sup>4</sup> để xây dựng lớp word embedding. FastText có khả năng biểu diễn tốt cho các từ ngữ, và các từ hiếm không xuất hiện trong quá trình huấn luyện, vì thế đây là một lựa chọn tốt cho việc Word embedding trên tiếng Việt.

### 5.4 Mô hình

Chúng tôi sử dụng nhiều mô hình học máy và học sâu để tìm ra mô hình áp dụng hiệu quả và phù hợp với bộ dữ liệu.

**Logistic Regression (LR):** Mô hình LR là một mô hình phân loại dựa trên dữ liệu có trước đó để huấn luyện mô hình. Với bài toán phân lớp các nhân dự đoán sẽ tương ứng với xác suất xảy ra và không xảy ra. Trong trường hợp này thuật toán sử dụng hàm sigmoid để dự đoán kết quả đầu ra.

**Multinomial Naive Bayes (Multi NB):** Mô hình Multi NB là mô hình phân loại theo xác suất được mô hình hóa từ định lý Bayes. Mô hình này thường đơn giản và có tốc độ nhanh với các bộ dữ liệu lớn.

**Support Vector Machine (SVM):** Mô hình (SVM) là một thuật toán học có giám sát có ứng dụng phổ biến cho cả 2 bài toán hồi quy và phân

<sup>4</sup><https://fasttext.cc/docs/en/crawl-vectors.html>



lớp. Với mục tiêu của mô hình là tìm ra siêu mặt phẳng sao cho phân chia dữ liệu một cách chính xác nhất.

**Long Short Term Memory (LSTM):** Mô hình LSTM (Hochreiter and Schmidhuber, 1997) lấy ý tưởng kết nối các thông tin phía trước để dự đoán thông tin hiện tại, vì vậy với các dữ liệu dạng chuỗi như các bình luận ta sẽ giữ được ngữ nghĩa của câu giúp cho việc dự đoán trên dữ liệu mới đạt kết quả tốt hơn vì ý nghĩa của các từ hay xuất hiện cùng nhau sẽ được mô hình ghi nhớ. LSTM ra đời giúp khắc phục các hạn chế của thuật toán trước đó.

**Bidirectional-LSTM (Bi-LSTM):** Mô hình Bi-LSTM (Schuster and Paliwal, 1997) là phiên bản mở rộng của mô hình LSTM (Hochreiter and Schmidhuber, 1997) với mục đích cải thiện hiệu suất của mô hình. Ý tưởng của Bi-LSTM là sẽ chạy LSTM hai lần trên chuỗi đầu vào (một lượt đi và một lượt về) việc này giúp cung cấp thêm thông tin về ngữ nghĩa cho mô hình, nhờ đó mà hiệu suất của mô hình được tăng lên.

## 5.5 Độ đo đánh giá

Để đánh giá mô hình, chúng tôi sử dụng độ đo  $F1_{score}$  được định nghĩa cho phase B-ABSA tại VLSP 2018 (Nguyen et al., 2019). Với A là tập hợp nhãn do mô hình dự đoán (bao gồm khía cạnh và cảm xúc) và B là tập hợp các nhãn được chú thích thì Precision, Recall và  $F1_{score}$  được tính như sau:

$$Precision = \frac{|A \cap B|}{|A|}; Recall = \frac{|A \cap B|}{|B|};$$

$$F1_{score} = \frac{2 * Precision * Recall}{Precision + Recall}$$

## 6 Thử Nghiệm Và Thảo Luận

### 6.1 Thông Số Mô Hình

Chúng tôi thử nghiệm các phương pháp trên các tập dữ liệu được phân chia giống nhau nhằm đánh giá công bằng nhất hiệu suất của các mô hình. Tiến hành tinh chỉnh và tìm tham số tối ưu cho 05 mô hình phía trên: Logistic Regression, Multinomial Naive Bayes, SVM, LSTM, Bi-LSTM. Với các mô hình học máy chúng tôi sử dụng thư viện Optuna<sup>5</sup> để tối ưu tham số.

**Logistic Regression:** Cài đặt LR với các tham số  $C = 1.0$ , class weight = balanced, random state = 5, max iter = 1000.

<sup>5</sup><https://github.com/optuna>

**Multinomial Naive Bayes:** Sử dụng Multinomial NB với thông số alpha = 0.05.

**Support Vector Machine:** Sử dụng Linear SVC với các tham số  $C = 0.1$ , class weight = balanced, loss = hinge, random state = 5, max iter = 2000.

**LSTM và Bi-LSTM:** Chúng tôi cài đặt thử nghiệm với lớp Embedding, lớp SpatialDropout1D với rate = 0.35, Với LSTM chúng tôi cài lớp LSTM còn với Bi-LSTM chúng tôi cài lớp Bidirectional đều với 128 đơn vị ẩn với rate = 0.15, lớp Conv1D, lớp Concatenate để kết hợp đầu ra của GlobalAveragePooling1D, GlobalMaxPooling1D. Và output là các lớp Dense, mỗi lớp có 4 đơn vị ẩn và kích hoạt bằng hàm “softmax”. Ở cả 2 mô hình đều sử dụng fastText để xây dựng lớp Word embedding.

### 6.2 Kết Quả Thử Nghiệm

Trong phần này, chúng tôi trình bày kết quả của các mô hình trong quá trình thử nghiệm, chi tiết được trình bày tại bảng 3. Kết quả cho thấy mô hình Bi-LSTM + fastText đạt kết quả tốt nhất với  $F1_{score}$  đạt 71.21%. Mô hình LSTM cũng cho kết quả tốt hơn các mô hình học máy truyền thống.

Bảng 3: Kết quả đánh giá của các mô hình

Model	Evaluation metrics		
	Pre.	Recall	F1
LR + Tfidf	59.53	<b>72.74</b>	65.47
Multi NB + Tfidf	76.54	33.38	46.48
Linear SVC + Tfidf	76.41	38.07	50.82
LR + Cv	71.54	66.67	69.02
Multi NB + Cv	54.26	54.75	54.50
Linear SVC + Cv	76.95	64.13	69.96
LSTM + fastText	<b>76.98</b>	65.31	70.67
Bi-LSTM + fastText	75.20	67.61	<b>71.21</b>

Mô hình Multi NB tuy có thời gian training ngắn nhưng lại có kết quả thấp hơn các mô hình còn lại. Các mô hình học khác vẫn cho kết quả khá tốt như Linear SVC + Cv cho kết quả tốt nhất trong các mô hình máy học mà chúng tôi thử nghiệm với  $F1_{score}$  đạt 69.96%. Ngoài ra qua thử nghiệm 02 phương pháp rút trích đặt trưng Tfidf và CountVectorizer thấy được phương pháp CountVectorizer làm cải thiện được hiệu suất mô hình đặc biệt với Linear SVC tăng 19% so với sử dụng Tfidf. Tuy vậy nhưng với mô hình LR + Tfidf cho kết quả Recall vượt trội hơn các mô hình còn lại đạt 72.74%.

Từ bảng 4 có thể thấy được tuy mô hình đạt kết quả tương đối tốt nhưng điểm  $F1_{score}$  của từng

Bảng 4: Kết quả các khía cạnh của mô hình Bi-LSTM

Khía Cạnh	F1	Khía Cạnh	F1
Service#General	<b>88.64</b>	Facilities#Quality	39.44
Rooms#Clean	86.9	F&D#Misc	37.50
Hotel#Clean	83.72	R&a#Clean	33.96
Location#General	81.69	Facilities#Clean	30.30
Hotel#General	80.03	F&D#Price	25.45
F&D#Sty&Opt	73.32	R&a#Quality	25.00
F&D#Quality	71.5	Hotel#Quality	20.69
Rooms#D&F	71.1	Facilities#General	20.59
Hotel#D&F	71.03	Hotel#Misc	19.42
Hotel#Comfort	69.07	Rooms#Quality	19.35
Rooms#General	64.77	Facilities#Price	12.12
R&a#General	62.86	R&a#Comfort	6.67
Hotel#Price	59.33	Facilities#Comfort	4.76
Rooms#Comfort	56.35	R&a#Price	0.00
Facilities#D&F	52.29	Facilities#Misc	0.00
R&a#D&F	45.52	R&a#Misc	0.00
Rooms#Price	44.66	Rooms#Misc	0.00

khía cạnh khá chênh lệch. Các khía cạnh có tần suất hiện cao ở hình 5a đạt kết quả tương đối tốt có khía cạnh đạt  $F1_{score}$  88.64%, còn các khía cạnh có số lượng nhãn ít ở hình 5b thì kết quả trên tập test rất thấp, có khía cạnh không được mô hình đoán đúng.

### 6.3 Phân tích lỗi

Để phân tích lỗi và tìm hướng phát triển cho bài toán chúng tôi tiến hành lấy ra 200 mẫu các kết quả dự đoán trên tập development bởi 02 mô hình cho kết quả tốt nhất là Bi-LSTM và LSTM.

Thông qua phân tích kết quả ở trên chúng tôi thấy được ở mô hình LSTM dự đoán đúng tất cả các khía cạnh và cảm xúc của điểm dữ liệu là 36.5%, còn Bi-LSTM là 34.5%. Sau đó, chúng tôi phân tích lỗi các lỗi thường xuất hiện và xác định ra 05 lỗi chính được thống kê ở hình 6.

(1) **Nhầm lẫn giữa các khía cạnh** là trường hợp lỗi phổ biến nhất, thường nhầm lẫn giữa các khía cạnh lẫn nhau, dự đoán thiếu khía cạnh, “*giá tiền trên app quá đắt so với phòng như vậy*” mô hình dự đoán là “*Hotel#Price*” thay vì “*Rooms#Price*”.

(2) **Xuất hiện các từ lạ, hiếm xuất hiện** “*Khách sạn nằm ngay trên đường hoa Nguyễn Huệ*” mô hình không phát hiện được nhãn “*Location#General*”.

(3) **Chứa nhiều thông tin, cấu trúc phức tạp** qua thống kê cho thấy cả 02 mô hình dự đoán khá tốt với câu có ít thông tin, tuy nhiên với câu có cấu trúc phức tạp, yêu cầu suy luận thì mô hình chưa thực sự tốt.

(4) **Nhầm lẫn nhãn cảm xúc** trong từng khía cạnh thì khách hàng có thể có nhiều cảm xúc cho

khía cạnh đó như “*vị trí gần sân bay nhưng xa trung tâm*” nhắc đến khía cạnh “*Location#General*” vừa có tích cực là “*gần sân bay*” và tiêu cực là “*xa trung tâm*” cho vì thế giải pháp hiện tại của chúng tôi là gộp tích cực và tiêu cực thành nhãn trung tính tuy vậy mô hình chưa phát hiện tốt. Các cảm xúc *Neutral* thường bị nhầm lẫn do các từ làm tăng giảm mức độ như: *tạm, căn bản, hơi, nhưng, etc.*

(5) **Lỗi do người gán nhãn** Tuy đã xây dựng một quy trình đảm bảo cho bộ dữ liệu chất lượng cao nhưng vẫn xảy ra một vài trường hợp gán lỗi sai thường xảy ra ở các câu phức tạp.

Ngoài ra việc mất cân bằng giữa các khía cạnh, tần suất của các cảm xúc làm thiếu ngữ cảnh để huấn luyện khía cạnh đó. Đồng thời dữ liệu bị nhiễu khá nhiều do viết tắt, sai chính tả, lỗi gõ phím làm khó khăn trong việc hiểu ngữ cảnh.

## 7 Kết Luận Và Hướng Phát Triển

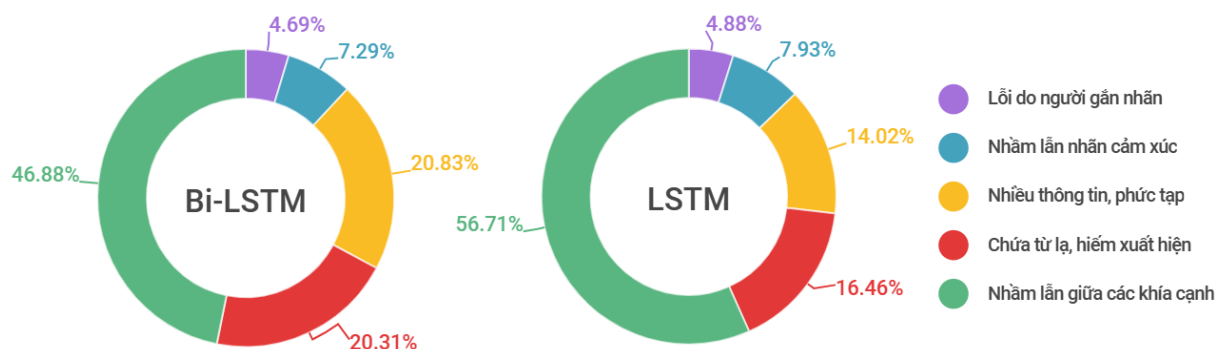
Trong nghiên cứu này chúng tôi đã xây dựng thành công được một bộ dữ liệu ABSA trên lĩnh vực khách sạn với hơn 8,000 bình luận và 36,000 cặp khía cạnh và cảm xúc. Từ kết quả thực nghiệm cho thấy được các mô hình học máy cho kết quả khá tốt, các mô hình học sâu cũng cho thấy được tiềm năng rất lớn đặc biệt với Bi-LSTM đạt  $F1_{score}$  71.21% vì thế trong tương lai chúng tôi thử nghiệm với các mô hình khác như: XLM-R, BiGRU, GRU, etc. Cũng như thử nghiệm thêm nhiều phương pháp Word embedding như GloVe, Word2vec.

Ngoài ra thông qua khảo sát ở mục A có thể thấy được thách thức và đặc điểm trong dữ liệu có rất nhiều trường hợp viết tắt, sai chưa tiền xử lý được, việc mất cân bằng các khía cạnh và tần suất của các cảm xúc. Nên trong tương lai chúng tôi sẽ cải tiến phương pháp tiền xử lý để có thể xử lý tốt hơn các trường hợp trên.

Chúng tôi cũng hướng tới việc xây dựng một quy trình tự động để thu thập các bài đánh giá dự đoán khía cạnh và cảm của các bài đánh giá cũng như trực quan hóa kết quả để mang lại cái nhìn bao quát hơn để người quản lý hiểu rõ hơn về chất lượng khách sạn của mình cũng như khách hàng lựa chọn được khách sạn phù hợp với mình.

## Acknowledgements

Cảm ơn thầy Nguyễn Gia Tuấn Anh và thầy Trần Quốc Khánh – người trực tiếp hướng dẫn, hỗ trợ nhiệt tình trong quá trình hoàn thành đồ án.



Hình 6: Thống kê lỗi của mô hình LSTM và Bi-LSTM

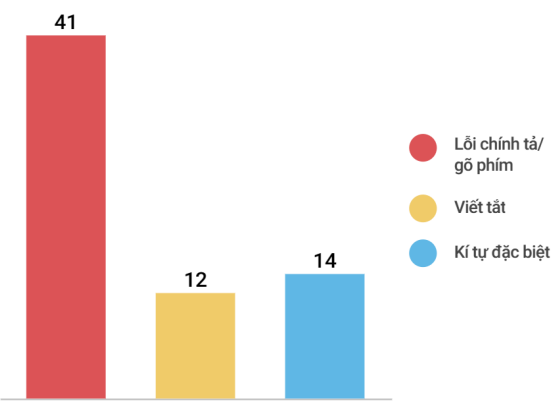
## References

- Anita Blomberg-Nygard and Chris K. Anderson. 2016. [United nations world tourism organization study on online guest reviews and hotel classification systems: An integrated approach](#). *Service Science*, 8(2):139–151.
- Hoang-Quan Dang, Trong-Hop Do, et al. 2022. Multi-task solution for aspect category sentiment analysis on vietnamese datasets. In *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pages 404–409. IEEE.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long short-term memory](#). *Neural Comput.*, 9(8):1735–1780.
- Luong Luc Phan, Phuc Huynh Pham, Kim Thi Thanh Nguyen, Sieu Khai Huynh, Tham Thi Nguyen, Luan Thanh Nguyen, Tin Van Huynh, and Kiet Van Nguyen. 2021. Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. In *Knowledge Science, Engineering and Management*, pages 647–658, Cham. Springer International Publishing.
- Huyen T M Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. 2019. [Vlsp shared task: Sentiment analysis](#). *Journal of Computer Science and Cybernetics*, 34(4):295–310.
- Kiet Van Nguyen, Vu Duc Nguyen, Phu X. V. Nguyen, Tham T. H. Truong, and Ngan Luu-Thuy Nguyen. 2018. [Uit-vsfc: Vietnamese students’ feedback corpus for sentiment analysis](#). In *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, pages 19–24.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. [SemEval-2016 task 5: Aspect based sentiment analysis](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. [SemEval-2015 task 12: Aspect based sentiment analysis](#). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. [SemEval-2014 task 4: Aspect based sentiment analysis](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
- M. Schuster and K.K. Paliwal. 1997. [Bidirectional recurrent neural networks](#). *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
- Kim Nguyen Thi Thanh, Sieu Huynh Khai, Phuc Pham Huynh, Luong Phan Luc, Duc-Vu Nguyen, and Kiet Nguyen Van. 2021. [Span detection for aspect-based sentiment analysis in vietnamese](#). In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pages 318–328, Shanghai, China. Association for Computational Linguistics.
- Quang-Linh Tran, Phan Thanh Dat Le, and Trong-Hop Do. 2022. [Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites](#). In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pages 767–776, Manila, Philippines. De La Salle University.

## A Khảo sát bộ dữ liệu

Do đây là những bình luận được thu thập trên nền tảng trực tuyến ở Việt Nam nên nó cũng mang các đặc điểm của dữ liệu văn bản tiếng Việt vì thế chúng tôi tiến hành khảo sát trên 5% dữ liệu của tập Train để có thể hiểu sâu hơn về bộ dữ liệu. Các

trường hợp thường gặp trong dữ liệu như: Lỗi chính tả / gõ phím; Viết tắt; Ký tự đặc biệt / icon.



Hình 7: Khảo sát đặc điểm bộ dữ liệu

Qua khảo sát, thu được kết quả như hình 7 thấy được đặc điểm lỗi chính tả và gõ phím gặp khá nhiều cao hơn nhiều so với 02 đặc điểm còn lại, lỗi nay xảy ra phổ biến là do một số lý do như sau.

**Nhầm lẫn đầu do phát âm của tùy vùng miền:** dui dễ, cũ kỹ, bài trí, xử lý, nghi ngơi, sạch sẽ , etc.

**Lỗi do gõ phím:** Từ *khách sạn* có thể bị gõ nhầm thành *khách sạnh*, *khâch sạn*, *kháchsạn* hoặc một từ cũng phổ biến như *sạch sẽ* sẽ bị gõ nhầm thành *sạch se* , *sạch sẽ*.

**Sai chính tả các từ mượn nước ngoài** (tiếng Anh) phổ biến nhất trong bộ dữ liệu là *buffet* sẽ bị sai thành *buffetr* , *buffe* hoặc *toilet* bị sai thành *totlet*, *toalet*.

B Chi tiết công việc

Bảng 5: Phân công công việc chi tiết

Nội dung	Người thực hiện
Định nghĩa và xây bài toán	Cả hai
Crawl dữ liệu	Q.Duy
Xây dựng guideline và gắn nhãn	Cả hai
Khảo sát thống kê dữ liệu	M.Duyên
Thử nghiệm các mô hình	Cả hai
Phân tích lỗi, hướng phát triển	Cả hai
Viết báo cáo	Q.Duy