

Dự Đoán Mức Độ Hải Lòng Của Khách Hàng Từ Các Bình Luận Tại Nhà Hàng

Võ Ngọc Anh Thy^{1,2,3}, Đinh Bảo Thy^{1,2,3}, Trần Đức Thiện^{1,2,3} và Đỗ Trọng Hợp^{1,2,4}

¹ Trường Đại học Công nghệ Thông tin, Thành phố Hồ Chí Minh, Việt Nam

² Đại học Quốc gia, Thành phố Hồ Chí Minh, Việt Nam

³{23521565, 23521563, 23521488}@gm.uit.edu.vn

⁴hopdt@uit.edu.vn

Tóm tắt. Bài báo này tập trung vào nhiệm vụ phân loại cảm xúc theo từng khía cạnh trong một bình luận. Hệ thống được xây dựng để nhận diện các khía cạnh liên quan và xác định thái độ của người dùng đối với từng khía cạnh, qua đó ước lượng mức độ hài lòng một cách chi tiết và chính xác hơn.

Keywords: Aspect Recognition, Aspect-Based Sentiment Analysis.

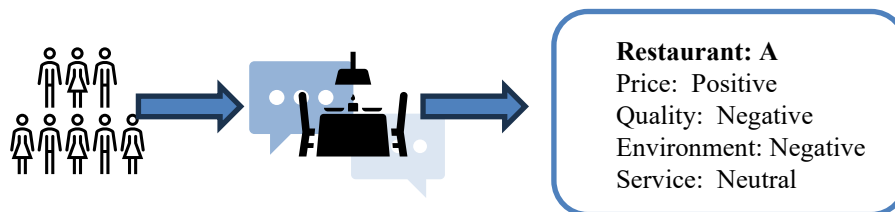
1 Giới thiệu

Với sự phát triển vượt bậc của thời đại số, các nền tảng mạng xã hội và thương mại điện tử ngày càng phổ biến, kéo theo là sự gia tăng nhanh chóng của các đánh giá và phản hồi từ khách hàng. Nếu các chủ doanh nghiệp, chuỗi nhà hàng có thể hiểu rõ được mong muốn, thị hiếu và mức độ hài lòng của khách hàng, họ sẽ có thể cải thiện chất lượng nhằm mang lại nhiều lợi ích và củng cố vị thế của doanh nghiệp.

Tuy nhiên, việc xử lý thủ công một khối lượng lớn dữ liệu phi cấu trúc – bình luận dưới dạng văn bản – là bất khả thi. Do đó, các kỹ thuật xử lý ngôn ngữ tự nhiên – Natural Language Processing (NLP) và phân tích cảm xúc – Sentiment Analysis đã trở thành công cụ hữu dụng trong việc khai thác thông tin.

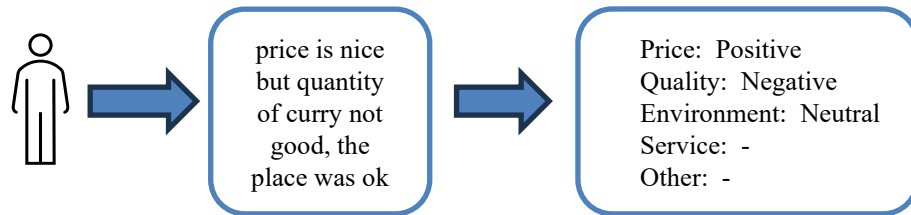
Hệ thống mà nhóm nghiên cứu có khả năng xác định các khía cạnh được đề cập trong bình luận bao gồm **giá cả, chất lượng, không gian, dịch vụ** và **các yếu tố khác**; từ đó dự đoán cảm xúc tương ứng cho từng khía cạnh. Kết quả tổng hợp là cái nhìn toàn diện của người dùng đối với doanh nghiệp. Để hiện thực hóa mục tiêu trên, nhóm cần giải quyết hai bài toán:

- *Bài toán lớn (Generalized Problem):* Phân Tích Phản Hồi Khách Hàng Theo Khía Cạnh Cho Nhà Hàng



Hình 1. Mô hình bài toán lớn

- *Bài toán nhỏ (Specialized Problem)*: Phân Tích Cảm xúc Bình Luận của Khách Hàng Dựa Trên Khía Cạnh, với hai nhiệm vụ chính:
 - Nhiệm vụ 1: Phát hiện khía cạnh
 - Nhiệm vụ 2: Phân loại cảm xúc dựa trên khía cạnh



Hình 2. Mô hình bài toán nhỏ

2 Bộ dữ liệu

2.1 Giới thiệu

Bộ dữ liệu mà nhóm sử dụng được trích xuất từ tập dữ liệu *10000 Restaurants Reviews* từ Kaggle bao gồm các thuộc tính:

- Restaurants: Tên nhà hàng
- Reviewer: Tên người nhận xét
- Review: Nội dung nhận xét
- Rating: Số sao của nhận xét
- Time (00:00:00 00/00/0000): Thời gian đăng nhận xét
- Pictures: Số lượng hình ảnh
- Metadata: thông tin về người đăng nhận xét
- 7514: Unknown

Với đặc thù của bài toán, thuộc tính chính cần được khai thác là nội dung nhận xét – “Review”. Nhằm đảm bảo tính khách quan trong quá trình huấn luyện mô hình, nhóm tiến hành trích xuất các bình luận từ nguồn dữ liệu gốc theo từng mức Rating, sao cho số lượng bình luận ở mỗi mức đánh giá được phân bố tương đối đồng đều.

Kết quả là tập dữ liệu sau xử lý bao gồm 1.244 bình luận, được lưu trữ trong một cột duy nhất là “Review”.

2.2 Khảo sát dữ liệu

Khảo sát dữ liệu thô thu thập được sẽ qua bước tiền xử lý cơ bản, bao gồm: loại bỏ trùng lặp, loại các dòng thiếu giá trị và đưa các thuộc tính về đúng kiểu dữ liệu trước khi khai phá dữ liệu sâu:



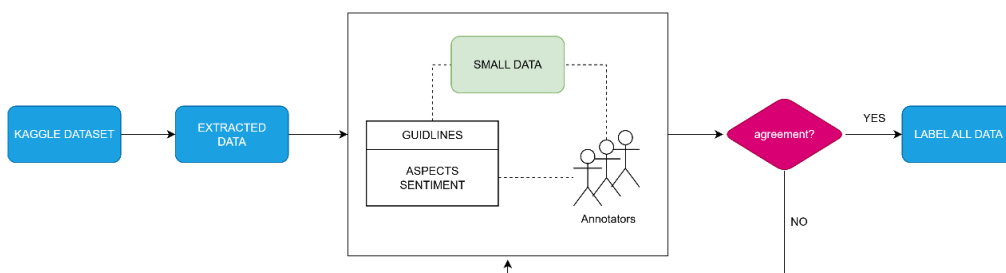
Hình 3. WordCloud của những bình luận ≥ 4 Rating, ≤ 2 Rating và 3 Rating

Qua hình 3, WordCloud thể hiện các tính từ thường thấy ở các tập Rating khác nhau cũng cho ta cái nhìn tổng quan hơn về cảm xúc của người dùng.

Mỗi bình luận sẽ được tách thành các câu riêng biệt, sau khi tách thành từng câu, nhóm có 5,288 câu từ 1,244 bình luận ban đầu đã được tiền xử lý và sẵn sàng cho việc gán nhãn.

2.3 Gán nhãn

Quy trình gán nhãn được thể hiện trong hình 4 và được giải thích theo từng giai đoạn sau:



Hình 4. Quá trình gán nhãn

Giai đoạn 1: Xây dựng guidelines – hướng dẫn gán nhãn

- Các khía cạnh (Aspects) (Bảng 1):

Tên	Mô tả
Price	Nhận xét liên quan đến giá cả, chi phí
Quality	Nhận xét về chất lượng món ăn, hương vị, số lượng
Environment	Không gian, nội thất, vị trí, không khí
Service	Thái độ phục vụ, nhân viên, tốc độ phục vụ
Other	Trường hợp chung, không thuộc các nhóm trên

Bảng 1. Mô tả các khía cạnh cần gán nhãn

- Cảm xúc (Sentiment):

- Positive: Tích cực
- Negative: Tiêu cực
- Neutral: Trung lập

- Công cụ *gán nhãn*:

- Môi trường làm việc: VS Code
- Công cụ gán nhãn: Label-Studio

- Tiêu chí gán nhãn cảm xúc:

Positive (Tích cực): Gán khi nhận xét thể hiện sự hài lòng, khen ngợi, đánh giá cao về một khía cạnh.

- Dấu hiệu: “delicious”, “worth it”, “amazing”, “great”, “affordable”
- “The biryani here is delicious.” → Quality-Positive

Negative (Tiêu cực): Gán khi câu thể hiện phàn nàn, chê bai, không hài lòng với một khía cạnh.

- Dấu hiệu: “bad”, “overpriced”, “tasteless”, “slow”, “rude”, “poor”
- “Service was slow.” → Service-Negative

Neutral (Trung lập): Gán khi câu không rõ cảm xúc, chỉ mô tả khách quan, thể hiện cảm xúc ở mức vừa phải, không rõ ràng, hoặc có cả tích cực lẫn tiêu cực.

- Dấu hiệu: “fine”, “ok”, hoặc chỉ liệt kê thông tin món ăn.
- “Food was fine, nothing special.” → Quality-Neutral

Trường hợp đặc biệt:

- Khi câu có cả nhận định tốt và xấu về một khía cạnh thì nhãn trả về sẽ là Negative.

Giai đoạn 2: Gán nhãn chéo nhằm kiểm định chất lượng gán nhãn

Để đảm bảo độ tin cậy của quá trình gán nhãn thủ công, nhóm tiến hành đo độ đồng thuận giữa các người gán nhãn trên một tập con gồm 1,123 câu, được trích từ 291 bình luận. Mỗi câu được gán nhãn độc lập bởi 3 người, với mục tiêu đánh giá mức độ nhất quán giữa các annotator.

Nhóm sử dụng 2 thước đo để tính độ đồng thuận:

- Độ đồng thuận tuyệt đối:

$$Agreement Rate = \frac{\text{số dòng có từ 2 annotator đồng thuận}}{\text{tổng số dòng}}$$

- Fleiss’ Kappa: chỉ số thống kê dùng để đo mức độ đồng thuận giữa nhiều annotator khi gán nhãn phân loại rời rạc, đồng thời hiệu chỉnh cho sự đồng thuận xảy ra do ngẫu nhiên.

Aspects	Total samples	Agree samples	Agreement_rate	Fleiss' Kappa
Price	1123	1090	0.9706	0.7583
Quality	1123	890	0.7925	0.7522
Environment	1123	1035	0.9216	0.8020
Service	1123	1020	0.9083	0.8215
Other	1123	946	0.8424	0.5785

Bảng 2. Độ đồng thuận khi gán nhãn chéo (lần 1)

Từ *bảng 2*, tỷ lệ đồng thuận (agreement rate) cao ở hầu hết các khía cạnh, dao động từ 79% đến 97%.

Chỉ số Fleiss' Kappa, đo lường mức độ đồng thuận loại trừ yếu tố ngẫu nhiên, cũng cho thấy mức độ nhất quán đáng tin cậy ở các khía cạnh Service (0.82), Environment (0.80) và Price (0.76). Duy nhất khía cạnh Other có độ đồng thuận thấp hơn (Kappa = 0.58), vì thế nhóm thực hiện gán nhãn lần 2 sau khi làm rõ các lỗi sai trong lần 1. Kết quả độ đồng thuận của lần 2 đạt yêu cầu (Fleiss' Kappa > 0.7).

Kết quả cho thấy rằng bộ hướng dẫn gán nhãn có tính rõ ràng và quy trình gán nhãn là đáng tin cậy. Bên cạnh đó, tập dữ liệu của người gán nhãn có độ tin cậy cao nhất (trong các mẫu được dùng để tính độ đồng thuận) sẽ được đưa vào bộ dữ liệu chạy mô hình.

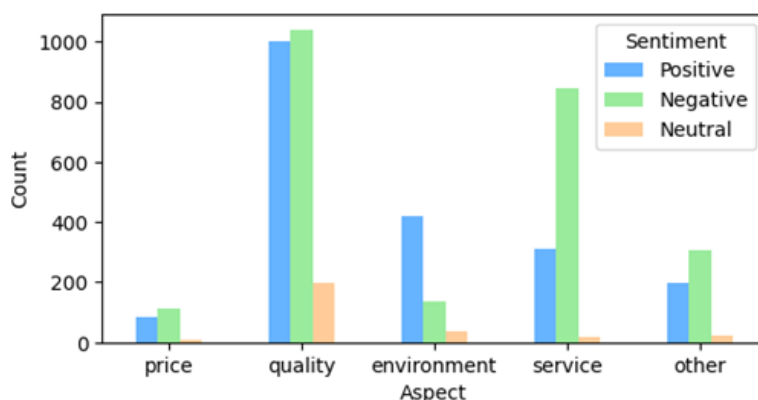
Giai đoạn 3: Gán nhãn toàn bộ dữ liệu

Sau khi gán nhãn ở giai đoạn 2, bộ dữ liệu còn 4,165 mẫu cần gán nhãn. Số mẫu này sẽ được chia đều cho 3 annotator nhằm đẩy nhanh tốc độ chuẩn bị dữ liệu.

id	sentence	price	quality	environment	service	other
0 386	i personally liked the bagel with cheese and c...	NaN	Positive	NaN	NaN	NaN
1 386	there was some renovation work happening outdo...	NaN	NaN	NaN	NaN	NaN
2 386	we were disturbed by the noise and hence could...	NaN	NaN	NaN	Negative	NaN
3 386	good for groups	NaN	NaN	NaN	NaN	Positive
4 387	good place to go around ambience is good and c...	NaN	Neutral	Positive	Positive	NaN

Hình 5. Dữ liệu mẫu

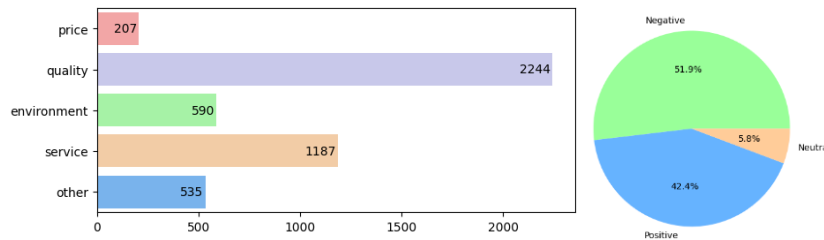
Sau cùng, bộ dữ liệu được gán nhãn bao gồm 5,288 mẫu và được thống kê: như *hình 6*.



Hình 6. Thống kê các khía cạnh và cảm xúc tương ứng

Hình 7 cho thấy dữ liệu chi tiết hơn về sự phân phối của các khía cạnh, “quality” là khía cạnh được nhắc đến nhiều nhất và “price” là ít nhất, cho thấy người dùng rất quan tâm đến chất lượng món ăn. Bên cạnh đó, “other” với 535 đánh giá đại diện cho cảm xúc của khách hàng mặc dù họ không nhắc tới bất kì khía cạnh cụ thể nào.

Biểu đồ tròn ở hình 7 cũng cho thấy nhãn “Negative” và “Positive” khá cân bằng nhau, chênh lệch khoảng 10%, trong khi đó, số lượng nhãn “Neutral” lại tương đối thấp với số lượng nhãn chưa chiếm tới 6% dữ liệu. Đây là một dấu hiệu rõ ràng của mất cân bằng nhãn (label imbalance) trong tập dữ liệu.



Hình 7. Thống kê các khía cạnh (hình bên trái) và thống kê các cảm xúc (hình bên phải)

3 Phương pháp máy học

3.1 Tiền xử lý

Dữ liệu bao gồm các dấu câu, biểu tượng cảm xúc không cần thiết và có nguy cơ gây nhiễu đến mô hình cần được loại bỏ ngay. Các bước tiền xử lý bao gồm:

- **Loại bỏ dấu câu (Punctuation Removal):** Tất cả các dấu câu không cần thiết như dấu chấm (.), dấu phẩy (,), dấu hỏi (?), dấu chấm than (!), dấu nháy đơn ('), dấu ngoặc kép ("), dấu gạch nối (-), v.v., sẽ được loại bỏ khỏi văn bản.
- **Chuyển đổi chữ hoa thành chữ thường (Case Conversion):** Tất cả các ký tự trong văn bản sẽ được chuyển thành chữ thường để đảm bảo tính nhất quán và tránh việc các từ giống nhau nhưng khác cách viết bị coi là khác nhau.
- **Chuẩn hóa từ ngữ (Normalization):** Văn bản được chuẩn hóa bằng cách thay thế các từ viết tắt (contractions) và từ lóng (slang) bằng hình thức đầy đủ hoặc chuẩn hơn. Các từ như "can't" sẽ được chuyển thành "cannot", và từ lóng như "u" sẽ được chuyển thành "you", dựa trên từ điển xây dựng sẵn.
- **Sửa lỗi chính tả (Typo Correction):** Các lỗi chính tả phổ biến trong văn bản sẽ được phát hiện và hiệu chỉnh về dạng chính xác bằng công cụ SymSpell nhằm tăng chất lượng dữ liệu đầu vào cho mô hình học máy.

3.2 Rút trích đặc trưng

Sau khi dữ liệu được tiền xử lý, bước tiếp theo là trích xuất đặc trưng quan trọng, tức là các thông tin nổi bật có ý nghĩa đối với phân tích cảm xúc.

Việc trích xuất đúng đặc trưng từ là vô cùng quan trọng, bởi vì hiệu quả của các thuật toán học máy phụ thuộc rất nhiều vào chất lượng và mức độ phù hợp của các đặc trưng này.

Để lựa chọn được mô hình phù hợp, nhóm thực hiện thực nghiệm thống kê trên 3 phương pháp phổ biến với 3 cách tiếp cận từ ngữ khác nhau:

- **TF-IDF [1]:**

- Phương pháp này đi theo hướng toán học – thống kê, xác định mức độ quan trọng của một từ trong toàn bộ tập văn bản – đánh giá các từ có ảnh hưởng lớn nhất dựa trên tần suất xuất hiện của từ trong mỗi văn bản.
- Mô hình TF-IDF là sự kết hợp của hai thành phần:

- *Term Frequency (TF)*: Tần suất xuất hiện của từ (t) trong tài liệu (d):

$$TF(t, d) = \frac{f(t, d)}{\sum f(t', d)}$$

- *Inverse Document Frequency (IDF)*: Mức độ hiếm của từ trong toàn bộ tập văn bản

$$IDF(t) = \log\left(\frac{N}{n_t}\right)$$

- *TF-IDF*: tích của hai đại lượng kể trên

- Nhược điểm: xem các từ là như nhau, không hiểu từ tích cực, tiêu cực, vì vậy ta sẽ loại bỏ các từ không cần thiết để tránh gây nhiễu.
 - **Lọc stopwords (Filtering stopwords)**: Stopwords là những từ không mang nhiều ý nghĩa cảm xúc ('a', 'is', 'the', 'with') và cần được loại bỏ. Nhóm sử dụng thư viện `ENGLISH_STOP_WORDS` cho công việc này. Tuy nhiên, trong bài toán phân loại cảm xúc, các từ như "cannot", "cant", "couldnt", "hasnt", "nor" là rất quan trọng để phân loại câu có cảm xúc Tiêu cực/Tích cực. Vì vậy, tập stopwords sẽ phải loại trừ các kí tự này trước khi áp dụng lên dữ liệu.

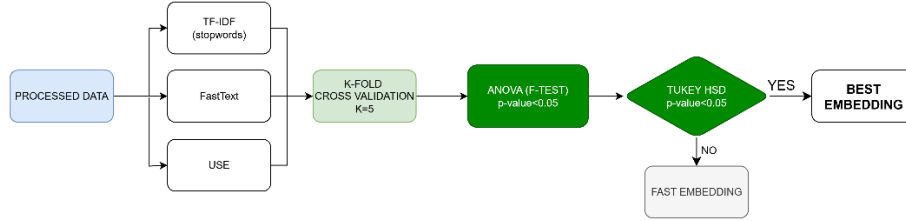
- **FastText [2]:**

- FastText mở rộng mô hình Word2Vec bằng cách xem mỗi từ là tập các n-gram ký tự, giúp mô hình hiểu được cấu trúc bên trong từ, ngữ nghĩa của các từ ngữ.
- Với FastText, vector của một từ là trung bình của các vector của các n-gram trong từ. Điều này cải thiện khả năng tổng quát hóa của mô hình.
- Nhược điểm: gặp khó khăn với câu đa nghĩa, có 2 vế đối lập.

- **USE [3]:**

- Phương pháp này sử dụng mô hình học sâu để mã hóa câu hoàn chỉnh thành vector có kích thước cố định (512 chiều), chứa thông tin ngữ nghĩa toàn câu.
- Vector đầu ra giữ được ngữ cảnh và ý nghĩa toàn diện của câu, nên mô hình hiểu đầy đủ cảm xúc.
- Nhược điểm: tốn nhiều tài nguyên tính toán, tốc độ xử lý chậm.

Các mô hình rút trích đặc trưng kể trên sẽ được thực hiện trong bài toán nhỏ (Specialized Problem) như *hình 8*.



Hình 8. Quy trình thực nghiệm thống kê các phương pháp embeddings

Bước 1: Dữ liệu được chia thành 5-fold theo K-fold cross-validation nhằm đảm bảo đánh giá mô hình một cách khách quan và ổn định.

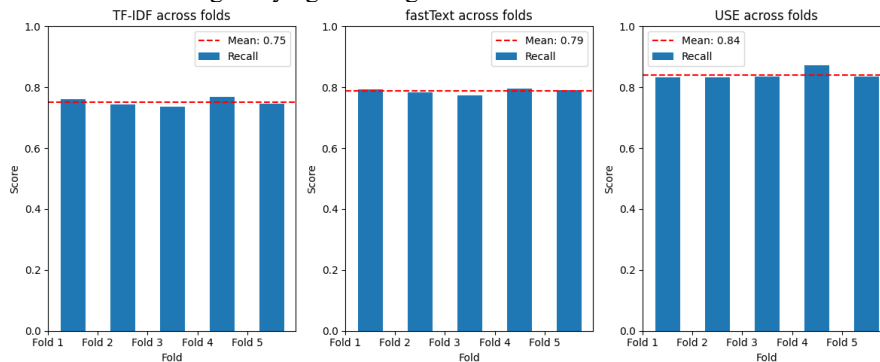
Bước 2: Trong mỗi fold, dữ liệu chia thành tập huấn luyện và kiểm tra theo tỉ lệ 4:1, giúp kiểm tra độ tổng quát hóa của mô hình trên nhiều phân phối khác nhau của tập dữ liệu. Mô hình được sử dụng là Logistic Regression, nhờ vào ưu điểm tính toán nhanh, dễ triển khai.

- Lưu ý: Với mỗi phương pháp embedding, quá trình chia K-fold được thực hiện độc lập, nhằm đảm bảo rằng các giá trị kiểm nghiệm là ngẫu nhiên và không phụ thuộc lẫn nhau, từ đó tăng độ tin cậy cho kết quả phân tích thống kê.
- Sau khi huấn luyện và dự đoán ở từng fold, các chỉ số đánh giá sẽ được ghi nhận để phân tích hiệu quả của từng phương pháp rút trích đặc trưng.

Bước 3: Phép kiểm định thống kê ANOVA (F-test) [4] nhằm đánh giá ý nghĩa thống kê của sự khác biệt giữa các phương pháp.

Với nhiệm vụ 1 “Phát hiện khía cạnh”:

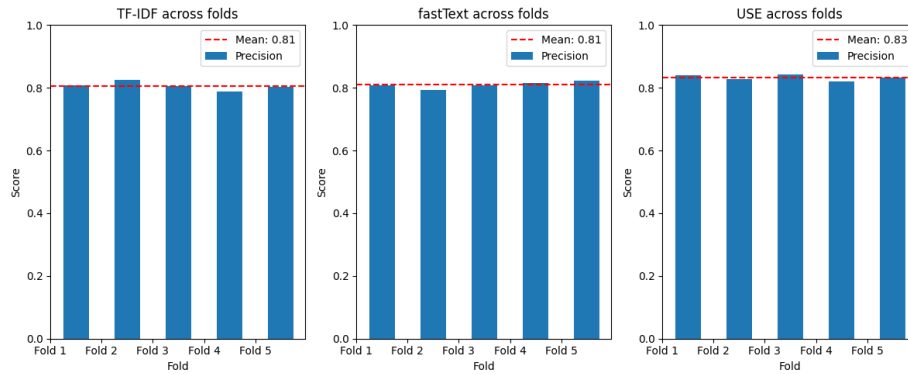
- Chỉ số Recall được ưu tiên vì cần đảm bảo không bỏ sót các khía cạnh có trong câu (*hình 9*).
- Do $p\text{-value} < 0.05$ ($p\text{-value} = 0.0000000898$) ta bác bỏ giả thuyết không (null hypothesis) và kết luận rằng sự khác biệt giữa các phương pháp embeddings có ý nghĩa thống kê.



Hình 9. Kết quả độ đo của các phương pháp embeddings của nhiệm vụ 1

Với nhiệm vụ 2 “Phân loại cảm xúc dựa trên khía cạnh”:

- Chỉ số Precision được ưu tiên vì mục tiêu là đảm bảo dự đoán đúng cảm xúc khi đã xác định một khía cạnh (hình 10).
- Do $p\text{-value} < 0.05$ ($p\text{-value} = 0.00347$) nên sự khác biệt giữa các phương pháp embeddings có ý nghĩa thống kê.



Hình 10. Kết quả độ đo với các phương pháp embeddings của nhiệm vụ 2

Bước 5: Thực hiện phép kiểm định hậu nghiệm Tukey HSD (Honestly Significant Difference) [4] để xác định cặp phương pháp nào khác biệt rõ rệt, từ đó tìm ra phương pháp embeddings cho kết quả tốt nhất.

Tukey HSD kết quả:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
FAST_TEXT	TF-IDF	-0.0398	0.0003	-0.0589	-0.0207	True
FAST_TEXT	USE	0.0526	0.0	0.0335	0.0717	True
TF-IDF	USE	0.0924	0.0	0.0733	0.1115	True

Bảng 3. Kết quả Tukey HSD trên nhiệm vụ 1

Sau khi tiến hành kiểm định hậu nghiệm Tukey HSD ở bảng 3, ta thấy sự khác biệt giữa các cặp phương pháp đều mang ý nghĩa thống kê ($p < 0.05$). Trong đó, phương pháp USE đạt recall trung bình cao nhất (0.8388), cao hơn đáng kể so với hai phương pháp fastText và TF-IDF.

Do đó, **USE** được chọn là phương pháp embedding tối ưu cho mô hình trong bài toán nhận diện khía cạnh.

Tukey HSD kết quả:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
FAST_TEXT	TF-IDF	-0.0057	0.72	-0.0253	0.0138	False
FAST_TEXT	USE	0.0243	0.016	0.0047	0.0438	True
TF-IDF	USE	0.03	0.004	0.0104	0.0496	True

Bảng 4. Kết quả Tukey HSD trên nhiệm vụ 2

Kết quả kiểm định hậu nghiệm Tukey HSD trên nhiệm vụ 2 ở bảng 4 cho thấy sự khác biệt giữa USE và hai phương pháp embedding còn lại có ý nghĩa thống kê.

- USE và TF-IDF: $p = 0.004$
- USE và FastText: $p = 0.016$

Trong khi đó, không có sự khác biệt rõ rệt giữa fastText và TF-IDF ($p = 0.72$).

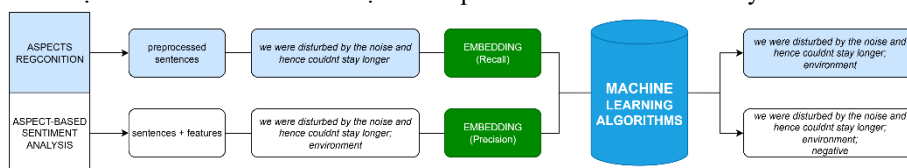
Với kết quả precision trung bình đạt 0.8333, **USE** được xem là phương pháp embedding hiệu quả nhất về mặt precision cho bài toán hiện tại.

Kết luận, cả hai nhiệm vụ đều chỉ ra rằng **USE (Universal Sentence Encoder)** là phương pháp biểu diễn văn bản cho hiệu suất mô hình tốt nhất, với sự khác biệt có ý nghĩa thống kê so với các phương pháp còn lại.

3.3 Huấn luyện mô hình

Nhóm nghiên cứu đề xuất phương pháp dựa trên học máy (Machine Learning) để phân tích cảm xúc trong bình luận về nhà hàng. Quy trình thực hiện được minh họa trong hình 11 với hai nhiệm vụ chính:

- Nhận diện khía cạnh – Aspect Recognition
- Dự đoán cảm xúc của khía cạnh – Aspect-based Sentiment Analysis



Hình 11. Quy trình huấn luyện mô hình

Cả hai nhiệm vụ đều được huấn luyện và đánh giá trên cùng tập các mô hình học máy gồm: **Logistic Regression**, **Random Forest**, **SVM** và **XGBoost**. Việc lựa chọn các mô hình này dựa trên khả năng xử lý tốt dữ liệu văn bản đã được biến đổi về dạng đặc trưng (feature vector) như USE, đồng thời cân nhắc đến hiệu năng, tính giải thích và độ ổn định của mô hình trong các bài toán phân loại đa lớp hoặc nhị phân.

Lưu ý: Với nhiệm vụ 1, do mỗi câu có thể liên quan đến nhiều khía cạnh cùng lúc nên đây là một bài toán gán đa nhãn. Vì vậy, ta sử dụng MultiOutputClassifier để mở rộng các mô hình phân loại sao cho có thể học và dự đoán nhiều nhãn đồng thời cho từng câu.

Logistic Regression (Hồi quy Logistic) [5]

Là một thuật toán phân loại tuyến tính được sử dụng để ước lượng xác suất của các lớp phân loại. Trong các bài toán phân loại nhị phân hoặc đa lớp, mô hình sử dụng hàm sigmoid (hoặc softmax) để chuyển đầu ra thành xác suất, từ đó dự đoán nhãn lớp phù hợp.

Đây là mô hình nhẹ, nhanh và hoạt động tốt trong các thiết lập cơ sở.

Random Forest (Rừng ngẫu nhiên) [6]

Là một mô hình học máy theo phương pháp tập hợp (ensemble), kết hợp nhiều cây quyết định (decision trees) để đưa ra dự đoán cuối cùng. Mỗi cây được huấn luyện trên một tập con dữ liệu khác nhau và mô hình chọn kết quả theo đa số (classification) hoặc trung bình (regression).

Với khả năng các mối quan hệ phi tuyến giữa các đặc trưng, Random Forest phù hợp để nhận diện khía cạnh.

Support Vector Machine (SVM) [7]

Là thuật toán học có giám sát, hoạt động bằng cách tìm siêu phẳng (hyperplane) tối ưu để phân tách các điểm dữ liệu thuộc các lớp khác nhau. Với khả năng sử dụng các hàm kernel, SVM có thể mở rộng cho các bài toán phân loại phi tuyến.

SVM làm việc hiệu quả trên các tập dữ liệu thưa (sparse data). Với các hàm kernel, SVM có thể mở rộng cho bài toán phân loại phi tuyến, phù hợp cho cả hai nhiệm vụ của bài toán.

XGBoost [8]

Là thuật toán tăng cường độ dốc (gradient boosting) được tối ưu hóa về tốc độ và hiệu suất. Mô hình xây dựng dần dần qua nhiều cây quyết định, trong đó mỗi cây mới học từ sai sót của các cây trước. XGBoost tích hợp cơ chế regularization để tránh overfitting, đồng thời hỗ trợ xử lý dữ liệu không cân bằng như tập dữ liệu hiện tại có nhãn Neutral rất ít.

Với khả năng xử lý thiếu dữ liệu, chống overfitting tốt và hỗ trợ tính song song, XGBoost là mô hình được ưu tiên trong nhiệm vụ phân loại cảm xúc theo khía cạnh.

Cuối cùng, việc kết hợp và đánh giá nhiều mô hình trên cả hai nhiệm vụ của bài toán giúp lựa chọn mô hình phù hợp nhất theo từng mục tiêu, đồng thời đảm bảo kết quả phân tích đạt được tính khách quan và độ tin cậy cao.

3.4 Độ đo đánh giá

Với mỗi nhiệm vụ riêng biệt, nhóm sử dụng độ đo khác nhau để đánh giá hiệu suất sao cho phù hợp với yêu cầu đầu ra:

Nhiệm vụ 1 - Nhận diện khía cạnh:

- Đây là bài toán đa nhãn, yêu cầu phát hiện đầy đủ các khía cạnh có mặt trong câu. Do đó, nhóm ưu tiên Recall để giảm thiểu việc bỏ sót các khía cạnh thực sự tồn tại.
- **Recall (Tỷ lệ phát hiện dương tính):** Còn gọi là độ nhạy (sensitivity) hoặc tỷ lệ dương tính đúng, đo lường tỷ lệ các trường hợp dương tính thật được mô hình phát hiện đúng.

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

- *TP (True Positive):* Số trường hợp dương tính được dự đoán đúng
- *FN (False Negative):* Số trường hợp dương tính bị dự đoán nhầm là âm tính

Nhiệm vụ 2 – Phân loại cảm xúc theo khía cạnh:

- Trong bài toán này, việc gán nhãn cảm xúc chính xác cho từng khía cạnh là quan trọng. Do đó, nhóm ưu tiên Precision để tránh dự đoán sai lệch, gây hiểu lầm trong phân tích cảm xúc.
- **Precision (Độ chính xác):** Tỷ lệ các dự đoán dương tính đúng trong tổng số dự đoán dương tính mà mô hình đưa ra.

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

- *TP (True Positive):* Số trường hợp dương tính được dự đoán đúng
- *FP (False Positive):* Số trường hợp âm tính bị dự đoán nhầm là dương tính

3.5 Kết quả

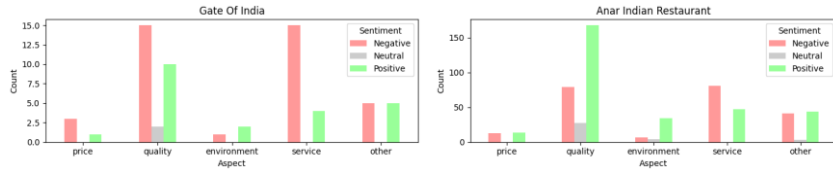
	Nhiệm vụ 1 - AR (Recall)	Nhiệm vụ 2 - ABSA (Precision)
Logistic Regression	0.831414	0.860294
Random Forest	0.524607	0.813208
SVC	0.693194	0.841046
XGBoost	0.649215	0.827158

Bảng 5. Kết quả khi chạy các mô hình cho 2 nhiệm vụ của bài toán 1

Từ kết quả chạy mô hình ở *Bảng 5*, ta thấy trong cả hai nhiệm vụ mô hình hồi quy Logistic Regression cho ra kết quả cao nhất, chứng tỏ nó phù hợp với yêu cầu của bài toán. Vì vậy, mô hình này được sử dụng cho bài toán lớn.

4 Thực nghiệm

Để hiện thực hóa bài toán lớn (Generalized Problem), nhóm tiến hành thu thập dữ liệu đầu vào là các nhận xét của khách hàng đối với cửa hàng cần nghiên cứu, được lấy từ nền tảng Google Maps. Sau khi trải qua các bước tiền xử lý và đưa vào mô hình phân tích, hệ thống sẽ cho ra kết quả đầu ra như minh họa tại *hình 12*. Với 2 nhà hàng thực nghiệm là “*Gate Of India*” gồm 24 bình luận và “*Anar Indian Restaurant*” với 181 bình luận.



Hình 12. Kết quả phân tích cảm xúc trên khía cạnh

Để xác định cảm xúc chung cho từng khía cạnh, nhóm áp dụng ngưỡng so sánh cho tỷ lệ giữa hai nhãn Positive và Negative như sau:

$$r_{aspect} = \frac{\text{số lượng nhãn Positive}}{\text{số lượng nhãn Negative}}$$

- $r_{aspect} > 1$: gán cảm xúc Negative cho khía cạnh tương ứng
- $1 \leq r_{aspect} \leq 1.5$: gán cảm xúc Neutral cho khía cạnh tương ứng
- $r_{aspect} > 1.5$: gán cảm xúc Positive cho khía cạnh tương ứng

Với cách đặt ngưỡng như trên, nhóm mong muốn giảm thiểu ảnh hưởng từ ý kiến chủ quan và sự khác biệt trong cảm nhận giữa các khách hàng, giúp kết quả phản ánh trung thực xu hướng cảm xúc chung.

Bảng 4 là kết quả cảm xúc của các khía cạnh của 2 mẫu nhà hàng.

	Gate Of India	Anar Indian Restaurant
Price	Negative	Neutral
Quality	Negative	Positive
Environment	Positive	Positive
Service	Negative	Negative
Other	Neutral	Neutral

Bảng 6. Kết quả của 2 mẫu thực nghiệm

5 Phân tích lỗi và hướng phát triển

5.1 Phân tích lỗi

Dù kết quả chạy mô hình khá tốt, nhóm nhận thấy phương pháp còn nhiều lỗi sai.

sentence	price	quality	environment	service	other
we ordered egg bryan and we were delivered chicken bryan with chicken replaced with eggs	0	1	0	0	0
but the place turned out to be a cheap place where people are allowed to smoke everywhere	1	0	0	1	0
yes the offer free ice cream if you write good reviews for them	0	1	0	0	0
worst quality we found spider at glass of beer and we informed to them but there was no response from their side	0	0	0	1	0

Hình 13. Lỗi sai trong nhiệm vụ 1

	sentence	y_pred	y_test
	so please give the food tasty [ASP] quality	Positive	Negative
	main course could have more variety and some taste [ASP] quality	Positive	Negative
	overall it is a bit expensive but if you have some free time to spend do visit this place [ASP] price	Positive	Negative
	weird however the food was served quickly [ASP] service	Negative	Positive

Hình 14. Lỗi sai trong nhiệm vụ 2

Trong hình 13, mô hình gặp khó khăn trong việc nhận diện khía cạnh gián tiếp, mơ hồ. Ví dụ trong câu “chicken replaced with eggs” người bình luận đang muốn nhắc tới việc dịch vụ giao hàng và chuẩn bị thực phẩm có sai lệch, và không có ý kiến về thức ăn – chất lượng thực phẩm. Mặc dù vậy, mô hình nhận diện “quality” và “service” là hai khía cạnh trong câu.

Hay trong hình 14, mô hình mắc lỗi khi các câu đang có ý nghĩa phủ định nhưng sử dụng các từ mang hướng tích cực nhằm đánh giá, so sánh.

Nguyên nhân có thể:

- Câu có từ tích cực nhưng ý nghĩa tiêu cực: câu có các từ ‘tasty, ‘taste nhưng mang mục đích so sánh với sự “Tiêu cực” ở hiện tại
- Ngữ cảnh gián tiếp, đa tầng: mặc dù khách hàng nói “the food tasty” nhưng đi kèm ngữ cảnh “please give the food tasty” mang ý nghĩa phủ định việc thức ăn ngon miệng. Điều này yêu cầu mô hình hiểu toàn bộ quan hệ giữa kỳ vọng và thực tế - điều mà mô hình chưa hoàn thiện được.

5.2 Hướng phát triển

Để cải thiện hiệu quả mô hình và mở rộng ứng dụng, một số hướng phát triển tiềm năng bao gồm:

- **Áp dụng mô hình học sâu (Deep Learning):** như BERT, RoBERTa, BiLSTM-attention... giúp hiểu ngữ cảnh tốt hơn.
- **Sử dụng embedding ngữ nghĩa nâng cao:** khi sử dụng các mô hình học sâu thì việc dùng TF-IDF, FastText hay USE sẽ không còn phù hợp, cần chuyển qua các mô hình ngữ cảnh như BERT hoặc Sentence-BERT và cần bổ sung xử lý phủ định ngữ nghĩa.
- **Tăng cường dữ liệu (Data Augmentation):** Sinh thêm dữ liệu bằng cách hoán đổi cấu trúc câu, paraphrase, hoặc dịch ngược để giảm tình trạng mất cân bằng và cải thiện khả năng tổng quát, đặc biệt là với nhãn “Neutral”.
- **Ứng dụng vào hệ thống thực tế:** Tích hợp mô hình vào chatbot, hệ thống đánh giá nhà hàng/ứng dụng thương mại điện tử để kiểm tra hiệu quả trên dữ liệu thực và thu thập phản hồi người dùng.

6 Kết luận

Việc phân tách rõ ràng hai nhiệm vụ và áp dụng nhiều mô hình học máy là một hướng đi hiệu quả trong bài toán ***Dự Đoán Mức Độ HÀi Lòng Của Khách Hàng Từ Các Bình Luận Tại Nhà Hàng***. Tuy nhiên, để đạt được độ chính xác cao hơn trong thực tế, cần tiếp tục cải tiến về mặt biểu diễn đặc trưng, mô hình học sâu, và dữ liệu huấn luyện. Những hướng phát triển nêu trên sẽ là cơ sở để nâng cao chất lượng hệ thống trong các nghiên cứu tiếp theo.

Tài liệu tham khảo

- [1] G. S. a. C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, pp. 513-523, 1988.
- [2] P. a. G. E. a. J. A. a. M. T. Bojanowski, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, 2017.
- [3] Kurzweil, Daniel Cer and Yinfei Yang and Sheng-yi Kong and Nan Hua and Nicole Limtiaco and Rhomni St. John and Noah Constant and Mario Guajardo-Cespedes and Steve Yuan and Chris Tar and Yun-Hsuan Sung and Brian Strope and Ray, "Universal Sentence Encoder," *CoRR*, vol. abs/1803.11175, 2018.
- [4] D. C. Montgomery, "Design and Analysis of Experiments," *John Wiley & Sons*, vol. 8th ed., 2013.
- [5] K. a. Y. S. Kumari, "Linear regression analysis study," *Journal of the Practice of Cardiovascular Sciences*, vol. 4, p. 33, 2018.
- [6] A. a. C. D. a. S. J. Cutler, "Random Forests," *Machine Learning*, vol. 45, pp. 157-176, 2011.
- [7] T. a. P. M. Evgeniou, "Support Vector Machines: Theory and Applications," vol. 2049, pp. 249-257, 2001.
- [8] T. a. G. C. Chen, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 785–794, 2016.