

# PHÂN TÍCH CẢM XÚC CÁC BÌNH LUẬN CỦA KHÁCH HÀNG DỰA TRÊN CÁC KHÓA CẠNH

1<sup>st</sup> Mã Kim Phát

Khoa Khoa Học và Kỹ Thuật Thông Tin  
Trường Đại Học Công Nghệ Thông Tin DHQG.HCM  
HoChiMinh City, VietNam  
22521071@gm.uit.edu.vn

2<sup>nd</sup> Võ Tấn Trung

Khoa Khoa Học và Kỹ Thuật Thông Tin  
Trường Đại Học Công Nghệ Thông Tin DHQG.HCM  
HoChiMinh City, VietNam  
22521573@gm.uit.edu.vn

3<sup>rd</sup> Nguyễn Tiến Nam

Khoa Khoa Học và Kỹ Thuật Thông Tin  
Trường Đại Học Công Nghệ Thông Tin DHQG.HCM  
HoChiMinh City, VietNam  
22520920@gm.uit.edu.vn

4<sup>th</sup> Đặng Chí Nguyên

Khoa Khoa Học và Kỹ Thuật Thông Tin  
Trường Đại Học Công Nghệ Thông Tin DHQG.HCM  
HoChiMinh City, VietNam  
22520963@gm.uit.edu.vn

**Abstract**—Mua bán hàng online đang dần thay thế cho các hình thức bán hàng truyền thống. Nó mang lại nhiều tiện lợi, cũng như cũng cho phép khách hàng có thể bình luận về sản phẩm một cách thoải mái, điều mà mua hàng trực tiếp không làm được. Những bình luận về sản phẩm là nguồn dữ liệu quý giá, cho phép những người mua sau biết được chất lượng của sản phẩm, đồng thời cho phép bên bán hàng biết được những khuyết điểm để cải thiện sản phẩm và nâng cao chất lượng người dùng. Bằng phương pháp Aspect Based Sentiment Analysis (ABSA), ABSA cho phép ta phân tích cảm xúc đánh giá phản hồi của khách hàng thông qua từng khía cạnh của sản phẩm. Bài báo cáo sử dụng độ đo Accuracy và F1-score (macro) để đánh giá mức độ chính xác và hiệu quả của mô hình.

## I. INTRODUCTION

Trong bối cảnh cuộc cách mạng công nghiệp lần thứ tư đang diễn ra, thế giới đã và đang ghi nhận sự phát triển và trỗi dậy mạnh mẽ của công nghệ thông tin. Công nghệ thông tin dần len lỏi và ảnh hưởng tới mọi mặt của đời sống ngày nay, và việc mua sắm cũng không nằm ngoài tầm ảnh hưởng. Công nghệ thông tin đã dần thay đổi và định hình lại cách mà người dùng mua sắm, chi tiêu. Từ đó mở ra một thị trường đầy tiềm năng phát triển. Đó là thương mại điện tử. Trong thập kỷ vừa qua, đặc biệt là sau đại dịch COVID-19, ngành thương mại điện tử đã ghi nhận sự tăng trưởng đáng kinh ngạc (tổng giá trị giao dịch thương mại tăng 35% trong năm 2020 so với năm 2019 và dự kiến đạt 52 tỷ USD vào năm 2025).

Song song với việc phát triển mạnh mẽ thì các sàn thương mại điện tử, những công cụ hỗ trợ kết nối người dùng và nhà bán hàng, cũng ghi nhận số lượng người dùng tăng lên đáng kinh ngạc (số lượng người dùng hàng tháng năm 2022 tăng gấp đôi so với năm 2019). Việc số lượng người dùng tăng lên cũng đồng nghĩa với sự tăng lên của các bình luận, đánh giá của người dùng về sản phẩm mà họ mua. Đây là nguồn dữ liệu dồi dào, quý giá cho các nhà sản xuất nắm được cảm

nhận của khách hàng về sản phẩm mà họ làm ra một cách trực tiếp. Ngoài ra, những đánh giá này cũng giúp cho các đơn vị vận hành sàn thương mại điện tử nắm được về chất lượng dịch vụ mà họ cung cấp.

Với động lực ban đầu nhằm giúp các doanh nghiệp đang triển khai bán hàng trên các sàn thương mại điện tử, bài báo cáo này dựa trên các phương pháp máy học (Machine Learning) nhằm đưa ra các thông tin và dự đoán cảm xúc của khách hàng trên từng khía cạnh của sản phẩm và dịch vụ để từ đó có thể giúp các doanh nghiệp và sàn thương mại điện tử điều chỉnh các thiếu sót và phát triển hơn trong tương lai.

## II. FUNDAMENTAL

### A. Task Definition

Input: Tập dữ liệu các bình luận của khách hàng

Output: Xác định bình luận đó đang nói về khía cạnh nào và cảm xúc của khách hàng về khía cạnh đó là Tích cực (Positive) hay Tiêu cực (Negative) hoặc có thể là không đề cập về cảm xúc (Neutral)

### B. Related Works

Trong bài báo Product Aspect Identification: Analyzing Role of Different Classifiers, Xing Hu, Sukanya Manna và Brian N. Truong đã xác định các khía cạnh của câu được thu thập từ Amazon với độ chính xác lần lượt là 79% và 77% thông qua việc sử dụng kỹ thuật Local unigram cùng với SVM và Random Forest.

Trong bài báo Aspect-Based Sentiment Analysis of Amazon Reviews for Fitness Tracking Devices, Alaa Shafae đã cho ra độ chính xác 55% với mô hình SVM trong việc phân loại sentiment của các sản phẩm Fitness Tracking - bài báo nhắm chính vào 3 sản phẩm - dựa trên các khía cạnh được phân tích từ các sản phẩm đó. Bài báo cũng cho biết rằng các phương pháp học có giám sát được sử dụng để phân loại cảm xúc cho

câu thì thường cho ra kết quả vượt trội so với cách tiếp cận không giám sát. Thế nhưng, một điểm yếu của học có giám sát là nó cho ra kết quả khá tệ khi thực hiện ở các lĩnh vực khác nhau và yêu cầu một lượng lớn dữ liệu đã được gán nhãn. Dự án được cung cấp tại liên kết sau: DS108-project

### III. DATASET

#### A. Data Collection

AMAZON là sàn thương mại điện tử được thành lập vào năm 1994 với khởi đầu là một hiệu sách trực tuyến. Trải qua gần 30 năm hình thành và phát triển, amazon hiện nay đã có mặt tại hơn 100 quốc gia khác nhau với danh mục sản phẩm đồ sộ (2022 amazon đã thêm vào danh mục sản phẩm của họ hơn 200 triệu sản phẩm mới). Họ cũng ‘độc chiếm’ thị trường Hoa Kỳ với 37,8% thị phần tại quốc gia này (số liệu thống kê bởi Statista năm 2023), bỏ xa các đối thủ như Walmart(6,3%), eBay(3,5%), Apple(3,9%),.... Từ đó, ta có thể thấy mức độ ảnh hưởng của sàn thương mại điện tử này với thị trường Mỹ nói riêng và thị trường toàn cầu nói chung. Bên cạnh đó, Mỹ là đất nước có đa dạng các ngôn ngữ, chủng tộc và có một cộng đồng khách hàng đã được hình thành từ lâu. Vì vậy, cộng đồng những người mua hàng trên amazon tại Mỹ vô cùng lớn, đi kèm với số lượng người dùng lớn và danh mục sản phẩm đa dạng là nguồn dữ liệu dồi dào về các bình luận đánh giá chất lượng cũng như dịch vụ mà sản phẩm mang lại.

Từ các tính chất mà nhóm em đưa ra ở trên, nhóm chúng em quyết định chọn AMAZON là nơi để thu thập các phản hồi của khách hàng về làm Dataset cho bài báo bằng công cụ BeautifullSoup.

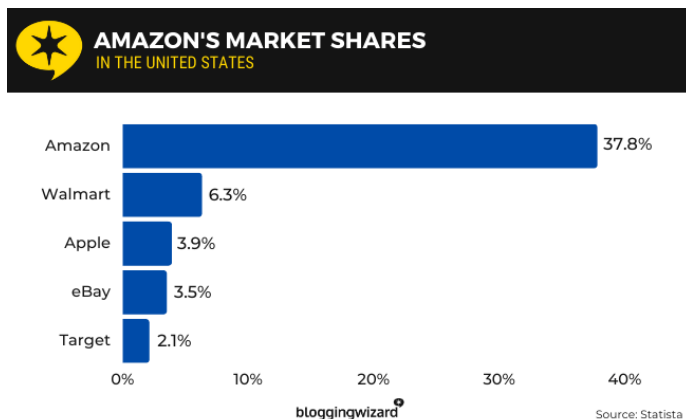


Fig. 1: Thị phần thị trường của AMAZON tại Mỹ  
Source: statista.com

#### B. Data Annotation

Dataset gồm 4 feature được nêu bên dưới:

- **Quality:** Đánh giá về chất lượng, nguyên liệu, thành phần,... liên quan đến sản phẩm
- **Price:** Liên quan đến giá cả, các chương trình khuyến mãi của sản phẩm.
- **Packaging:** Liên quan đến đóng gói, thiết kế và bao bì của nhà sản xuất.

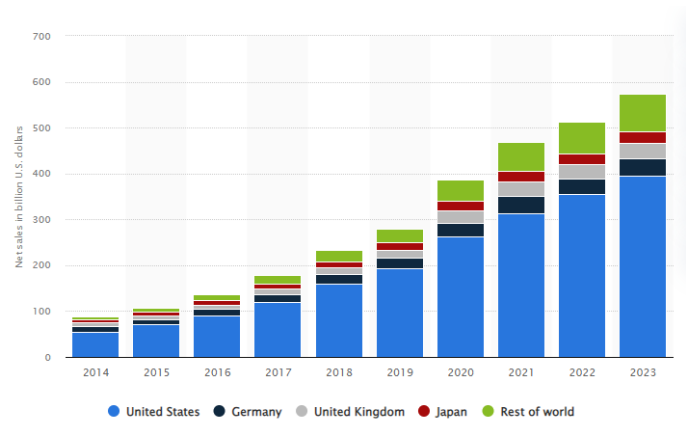


Fig. 2: Doanh số bán hàng của AMAZON trên toàn thế giới  
Source: statista.com

- **Service:** Liên quan đến giao hàng nhanh, tình trạng sản phẩm trong quá trình giao hàng và dịch vụ hỗ trợ, chăm sóc khách hàng.
- **Sentiment:** Thể hiện cảm xúc của câu (Tích cực, Tiêu cực, Trung tính).

Trong đó các feature được gán nhãn thủ công bởi 2 annotators với quy định lần lượt như sau:

- **Quality, Price, Packaging, Service:** nếu câu đó có ý nhắc đến feature nào thì đánh nhãn 1 vào feature đó, ngược lại thì đánh 0.
- **Sentiment:** được gán nhãn 0, 1, 2 với ý nghĩa lần lượt là Negative, Neutral, Positive

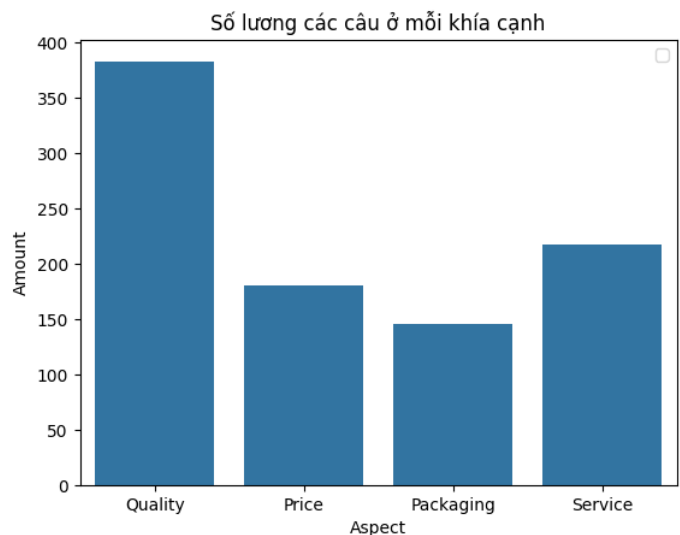


Fig. 3: Số câu ở mỗi khía cạnh

Sau khi đã gán thủ công các label, nhóm tính độ đồng thuận Cohen's kappa bằng thư viện sklearn. Theo đó, thu được giá trị Cohen's kappa:

- **Quality:** 75.107%
- **Price:** 87.808%

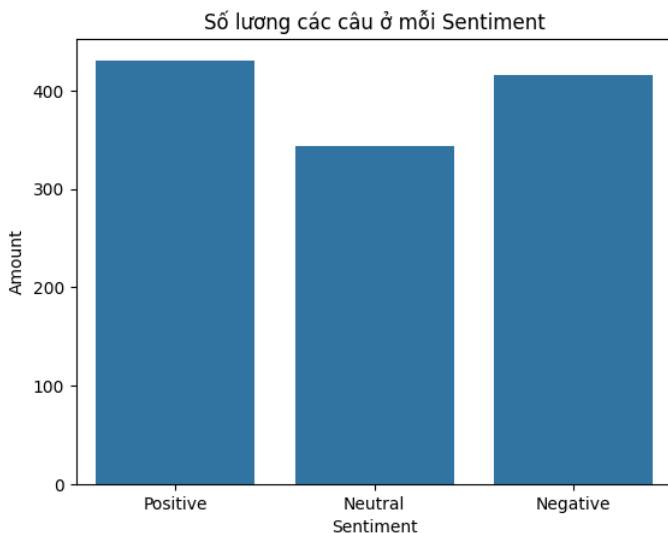


Fig. 4: Số câu ở mỗi sentiment

- **Packaging:** 75.148%
- **Service:** 71.3023%
- **Sentiment:** 80.187%

### C. Pre-processing

Trước hết, làm sạch bộ dữ liệu bằng cách loại bỏ các dữ liệu bị null, bị lỗi phong (nếu sửa được thì sẽ sửa bằng tay). Sau đó là tới các bước tiền xử lý từng câu, có 4 bước tiền xử lý dữ liệu.

- **Bước 1 : Loại bỏ các icon, kí tự lạ và chuyển từ viết hoa thành viết thường**  
Loại bỏ bằng cách sử dụng bảng mã ascii, vì các từ trong tiếng anh đều giới hạn trong khoảng [65,90] và [97,122]. Nên chỉ cần bỏ đi những từ có giá trị nằm ngoài khoảng này là được và chỉ giữ lại 1 số dấu câu có giá trị. Sau đó thì dùng hàm lower() để chuyển các từ viết hoa thành viết thường.
- **Bước 2 : Loại bỏ stop word**  
Nhóm chúng em sử dụng module stopwords của package nltk.corpus trong ngôn ngữ tiếng Anh để lấy về corpus, một điểm đặc biệt là chúng em không bỏ đi các stopwords trong câu mang hàm ý phủ định, ví dụ: wasn't, haven't.
- **Bước 3 : Word stemming**  
Nhóm em dùng module LancasterStemmer của package nltk.stem để đưa các từ về dạng ban đầu của nó, nhằm mục đích giúp cho bộ từ điển chính xác và trực quan hơn.  
Ví dụ:  
  - running -> run
  - happily -> happy
  - foxes -> fox

### D. Data vectorizing

1) *Sử dụng Tokenizer của thư viện tensorflow*: Sau khi pre-processing bộ dữ liệu, chúng em sử dụng tf.keras.preprocessing.text.Tokenizer để chuyển đổi text thành

các vector để phục vụ cho giai đoạn huấn luyện mô hình. Độ dài của vector ban đầu được chọn từ độ dài phổ biến nhất của các câu trong bộ dữ liệu. Khi đó các câu bị quá độ dài sẽ bị cắt bớt và các câu không đủ độ dài thì sẽ được thêm giá trị 0 vào cuối vector cho tới khi độ dài vector bằng với độ dài quy định ban đầu.

2) *Sử dụng phương pháp Count Vector*: Kỹ thuật CountVectorizer được áp dụng lên bộ dữ liệu đã được tiền xử lý. Thu được bộ từ vựng Vocabulary chứa tất cả những từ xuất hiện trong dữ liệu, mỗi từ trong bộ từ vựng đều có 1 con số định danh riêng. Dữ liệu ban đầu sẽ được mã hóa theo bộ từ vựng và có kích thước :

$$\text{lenght}(\text{dataset}) * \text{lenght}(\text{Vocabulary})$$

3) *Phương án của nhóm đề xuất: Ý tưởng của thuật toán* : dựa trên ý tưởng của **TF-IDF** nhưng thay vì tìm mức độ quan trọng của từ đó thì tìm mức độ của từ đó trên từng khía cạnh

-Sau các bước tiền xử lý bộ dữ liệu, nhóm chúng em tiếp tục như sau:

- **Create Key Word Dict** Có 2 bước để tạo 1 dòng trong Key Word Dict

- **Bước 1** : Đếm số lần từ đó xuất hiện trong các label.
- **Bước 2** : Sau đó chuyển đổi giá trị lại theo tỉ lệ phần trăm trong sentiment hoặc aspect. Trong sentiment, thì nếu sự chênh lệch giữa 3 nhãn Neg, Neu, Pos càng thấp thì giá trị của từ đó càng cao

sau khi chạy xong, một giá trị từ sẽ thành 1 vector 8 chiều theo thứ tự là :

[Pos, Neg, Neu, None, Qual, Pri, Pack, Ser]

Ví Dụ: great -> gre

**Trước Khi Scaler :**

gre = ( 1, 3, 55, 3, 14, 26, 6, 0 )

**Sau Khi Scaler :**

gre = ( 0.009, 0.0277, 0.509, 0.061, 0.286, 0.531, 0.122, 0 )

- **Sen To Vec**

Chuyển đổi các câu thành vector theo cách là cộng tất cả vector của từ đó (trong bước trên) lại thành một vector cho câu đó.

ví dụ :

**Câu ban đầu :**

- i recently purchased the pepsi cola soda 12oz cans, 24 pack from amazon, and i must say, it was an excellent choice!

**Câu sau khi tiền xử lý :**

- rec purchas peps col sod 12oz cans, 24 pack amazon, must say, excel choice!

**Câu sau khi vec to sen :**

(7.578, 6.131, 10.4292, 1.643, 9.107, 0.926, 2.324, 0.0)

- **Chạy Mô Hình**

Để dự đoán được nhãn của 1 câu thì cần phải chạy qua 2 mô hình

- **Mô hình 1** : dùng 5 feature cuối để dự đoán khía cạnh được nhắc đến trong câu

- **Mô hình 2** : gồm nhiều mô hình nhỏ, mỗi mô hình dùng 3 feature đầu để dự đoán sentiment cho mỗi khía cạnh

Ví dụ :

Cho câu < A >

vec <- sen2vec(A)

aspect <- model\_1.predict(vec[-5:])

sentiment <- model\_2[aspect].predict(vec[:3])

Vậy suy ra câu < A > có nhãn là [sentiment,aspect]

Độ dài mỗi vector trong dữ liệu mới thu được là như nhau.

#### E. Model Building

Nhóm em sử dụng 3 mô hình học máy để dự đoán các khía cạnh và sentiment cho câu

- Mô hình KNN:
  - + Hoạt động dựa trên nguyên tắc: Gần nhau thì giống nhau
  - + Để dự đoán nhãn cho dữ liệu mới, ta cần xác định điểm dữ liệu mới này, sau đó tiến hành tính toán khoảng cách từ điểm dữ liệu mới đến các điểm dữ liệu có sẵn. Sau đó chọn ra K điểm gần nhất. Nhãn của điểm dữ liệu mới phụ thuộc vào nhãn phổ biến nhất trong K điểm dữ liệu lân cận.
  - + Trong bài toán này thì chúng em chọn K=3
  - + model1 = KNeighborsClassifier(n\_neighbors=3)
- Mô hình SVM:
  - + Hoạt động dựa trên nguyên tắc tìm ra mặt phẳng (hoặc siêu phẳng) trong không gian đa chiều để phân chia các điểm dữ liệu vào các lớp khác nhau.
  - + Mỗi điểm dữ liệu trong không gian đa chiều sẽ được biểu diễn bằng một vector. SVM sẽ tìm ra mặt phẳng (siêu phẳng) phân chia các vector này sao cho khoảng cách từ điểm gần nhất của mỗi class tới mặt phẳng phân chia là như nhau và khoảng cách này của cả 2 class là lớn nhất. Đồng thời, số lượng các điểm nằm trên mặt phẳng này phải là nhỏ nhất.
  - + Việc dự đoán nhãn cho dữ liệu mới phụ thuộc vào vị trí của vector dữ liệu đó so với mặt phẳng phân chia.
  - + model2 = svm.SVC(kernel = 'linear', gamma = 0.1, C = 100)
- Mô hình Logistic Regression:
  - + Hoạt động bằng cách mô hình hóa sự liên quan giữa các biến độc lập (các biến đầu vào) và biến phụ thuộc (các biến đầu ra). Sau đó dự đoán xác suất điểm dữ liệu thuộc về 1 trong 2 lớp
  - + Sử dụng thuật toán tối ưu nhằm tìm ra tham số tối ưu cho hàm số logistic sigmoid. Mục tiêu là để tối ưu hàm mất mát ( entropy chéo hoặc log loss)
  - + Trong bài toán này thì chúng em sử dụng hàm tối ưu là 'lbfgs' (Limited-memory BFGS) với số lần lặp tối đa là 1000
  - + model3 = linear\_model.LogisticRegression(C=1e5, solver = 'lbfgs', max\_iter=1000)

#### F. Độ đo để đánh giá mô hình

Confusion matrix để ước tính hiệu suất của mô hình phân loại trong đó chỉ ra số dự đoán đúng và dự đoán không đúng bằng cách xem xét giá trị đúng đã biết trước. Dựa trên positive and negative classes, ta có được các giá trị True Positive (TP), True Negative (TN), False Positive (FP) và False Negative (FN).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

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

#### IV. EXPERIMENT AND RESULT

Nhóm sử dụng phương pháp GridSearch để tinh chỉnh các tham số và tìm ra tham số tối ưu cho mô hình từ đó sử dụng tham số đó vào mô hình để dự đoán.

Dưới đây lần lượt là kết quả thực nghiệm ở cả ba phương pháp vector hóa câu

Classifier	Accuracy	Precision	recall	F1_score
Sentiment				
SVM	0.4017	0.3879	0.3872	0.3734
KNN	0.3766	0.3592	0.3581	0.3550
Logistic	0.3808	0.3718	0.3697	0.3593
Quality				
SVM	0.6778	0.5287	0.6331	0.4774
KNN	0.6736	0.5576	0.6071	0.5452
Logistic	0.6485	0.5228	0.5474	0.4976
Price				
SVM	0.8452	0.5000	0.4226	0.4580
KNN	0.8410	0.4975	0.4223	0.4568
Logistic	0.8452	0.5000	0.4226	0.4580
Packaging				
SVM	0.8787	0.5000	0.4393	0.4677
KNN	0.8745	0.4976	0.4391	0.4665
Logistic	0.8787	0.5000	0.4393	0.4677
Service				
SVM	0.8326	0.5000	0.4163	0.4543
KNN	0.8243	0.4950	0.4156	0.4518
Logistic	0.8201	0.4925	0.4153	0.4506

Fig. 5: Phương Pháp Tokenizer Của Tensorflow

#### V. LIMITATION AND FUTURE WORK

##### A. Limitation

Với kết quả mà phương pháp nhóm chúng em đề xuất, nhóm nhận thấy rằng vẫn còn một vài trường hợp trong từ điển từ cho ra chỉ số của từ còn nhầm lẫn dẫn đến kết quả dự đoán câu còn hạn chế.

**Ví dụ:**

taste bad -> tast bad [negative, quality]

taste not bad -> tast not bad [negative, quality]

Classifier	Accuracy	Precision	recall	F1_score
Sentiment				
SVM	0,82438	0,41471	0,74740	0,43586
KNN	0,82645	0,44486	0,66793	0,47425
Logistic	0,81749	0,56938	0,60243	0,58182
Quality				
SVM	0,75207	0,57474	0,38243	0,38318
KNN	0,74931	0,49034	0,36879	0,35928
Logistic	0,75482	0,55336	0,5219	0,5336
Price				
SVM	0,90634	0,75759	0,4979	0,55678
KNN	0,90634	0,79968	0,52463	0,595
Logistic	0,90358	0,66344	0,64947	0,6509
Packaging				
SVM	0,84022	0,89077	0,41299	0,44555
KNN	0,83747	0,62828	0,45848	0,48094
Logistic	0,80992	0,58533	0,54291	0,5604
Service				
SVM	0,7989	0,76648	0,36551	0,35793
KNN	0,81267	0,75341	0,42755	0,46179
Logistic	0,80165	0,6076	0,56324	0,58238

Fig. 6: Phương Pháp CountVector

Classifier	Accuracy	Precision	recall	F1_score
Sentiment				
SVM	0.9453	0.9203	0.9187	0.9186
KNN	0.9127	0.8653	0.8659	0.8628
Logistic	0.9394	0.9109	0.9106	0.9105
None				
SVM	0.9349	0.8491	0.8491	0.8491
KNN	0.9431	0.8197	0.9433	0.8772
Logistic	0.9309	0.8214	0.8679	0.8440
Quality				
SVM	0.9268	0.8902	0.8902	0.8902
KNN	0.9308	0.8824	0.9146	0.8982
Logistic	0.9268	0.8809	0.9024	0.8916
Price				
SVM	0.9837	0.9444	0.9444	0.9444
KNN	0.9715	0.9143	0.8889	0.9014
Logistic	0.9797	0.9429	0.9167	0.9296
Packaging				
SVM	0.9552	0.7352	0.9259	0.8197
KNN	0.9512	0.7419	0.8519	0.7931
Logistic	0.9553	0.75	0.8889	0.8136
Service				
SVM	0.9309	0.8780	0.75	0.8089
KNN	0.9268	0.9412	0.6667	0.7805
Logistic	0.9228	0.8718	0.7083	0.7816

Fig. 7: Phương Pháp nhóm chúng em đề xuất

**Lý giải:** Nhóm nhận thấy rằng chỉ số của từ "not" chưa đủ để biểu thị trong trường hợp này.

Mặc dù dữ liệu được lấy từ AMAZON, nơi mà bất kì đất nước hay vùng lãnh thổ nào cũng có thể tham gia. Do đó ngôn ngữ được sử dụng bởi các khách hàng là vô cùng đa dạng. Thế nhưng, nhóm chúng em vẫn chưa thể xây dựng mô hình có thể sử dụng cho đa dạng các loại ngôn ngữ mà chỉ khai thác một ngôn ngữ duy nhất là Tiếng Anh.

Bên cạnh đó, do các từ được gán các trọng số ở từng khía cạnh, điều này còn một số hạn chế ở những câu mặc dù dùng nhiều từ ngữ positive từ đó cho ra kết quả là positive nhưng lại mang nghĩa mỉa mai về người hay một điều gì đó.

**Ví dụ:** Sometimes I just agree with you, so you will stop talking.

Cuối cùng do phải tạo một bộ dữ liệu về các từ vựng chứa đựng trọng số thế nên đòi hỏi một lượng tài nguyên lưu trữ rất lớn khi phát triển tiếp tục tập từ vựng này.

## B. Future work

Từ đó, trong tương lai chúng em đề xuất đến việc xây dựng bộ dữ liệu rộng hơn, để từ đó có được bộ từ điển các từ chính xác hơn và đa dạng hơn. Bên cạnh đó, nhóm chúng em cũng sẽ phát triển mô hình ở những bộ dữ liệu ở ngôn ngữ và các loại sản phẩm đa dạng khác nhau để tăng tính chính xác cho mô hình.

Cuối cùng, để tận dụng lượng dữ liệu lớn mà bọn em khai thác trong tương lai, chúng em cũng đề xuất các mô hình học sâu (Deep Learning) để huấn luyện dữ liệu.

## VI. CONCLUSION

Tóm lại, trong bài báo cáo này nhóm đã đề xuất thêm 1 phương pháp để phân tích cảm xúc các bình luận của khách hàng dựa trên các khía cạnh. Dựa vào kết quả được đo bằng các chỉ số Accuracy, Precision, Recall, F1-score, cho thấy kết quả của nhóm chúng em là khá tốt khi so sánh với các bài nghiên cứu được nêu trên phần Related Works, cũng như là các phương pháp sử dụng thư viện để xử lý như trong bảng kết quả ở phần EXPERIMENT AND RESULT.

Trong tương lai, để cải thiện độ chính xác của phương pháp này, chúng em đề xuất việc xây dựng bộ dữ liệu lớn hơn để có được bộ từ điển chính xác và đa dạng hơn. Bên cạnh đó, chúng em cũng thấy được tiềm năng của việc áp dụng mô hình Deep Learning trong phương pháp này.

## REFERENCES

- [1] Xing Hu, Sukanya Manna, and Brian N Truong. "Product aspect identification: Analyzing role of different classifiers". In: *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE. 2014, pp. 202–209.
- [2] Md Shahriare Satu et al. "TClustVID: A novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets". In: *Knowledge-Based Systems* 226 (2021), p. 107126.
- [3] Alaa Shafae et al. "Aspect-based sentiment analysis of amazon reviews for fitness tracking devices". In: *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2014 International Workshops: DANTH, BDM, MobiSocial, BigEC, CloudSD, MSMV-MBI, SDA, DMDA-Health, ALSIP, SocNet, DMBIH, BigPMA, Tainan, Taiwan, May 13-16, 2014. Revised Selected Papers 18*. Springer. 2014, pp. 50–61.