

TRƯỜNG ĐẠI HỌC NGÂN HÀNG TP HỒ CHÍ MINH
KHOA HỆ THỐNG THÔNG TIN QUẢN LÝ
NGHIÊN CỨU KHOA HỌC

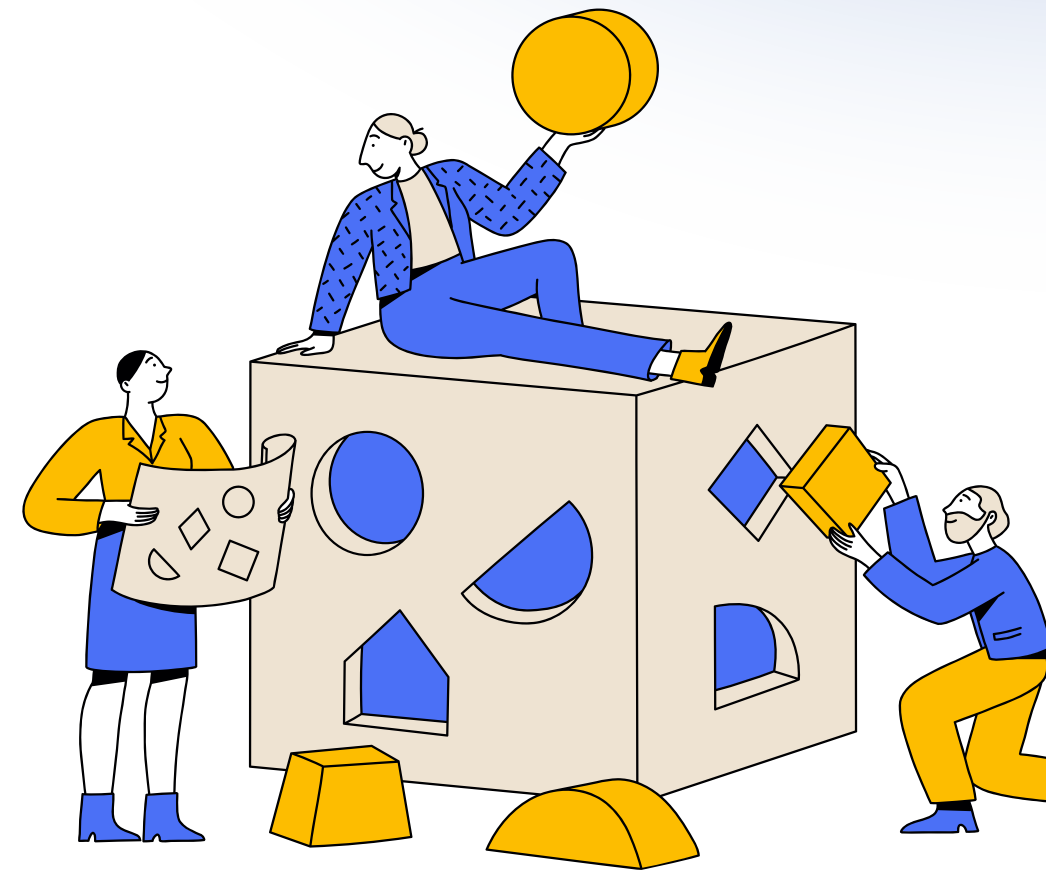
ĐỀ TÀI:

**PHÂN TÍCH CẢM XÚC CỦA KHÁCH HÀNG DỰA
TRÊN KHÍA CẠNH BÌNH LUẬN SỬ DỤNG MÔ
HÌNH HỌC SÂU**

**Giảng viên hướng dẫn:
ThS. Nguyễn Văn Kiên**

Các sinh viên thực hiện đề tài:

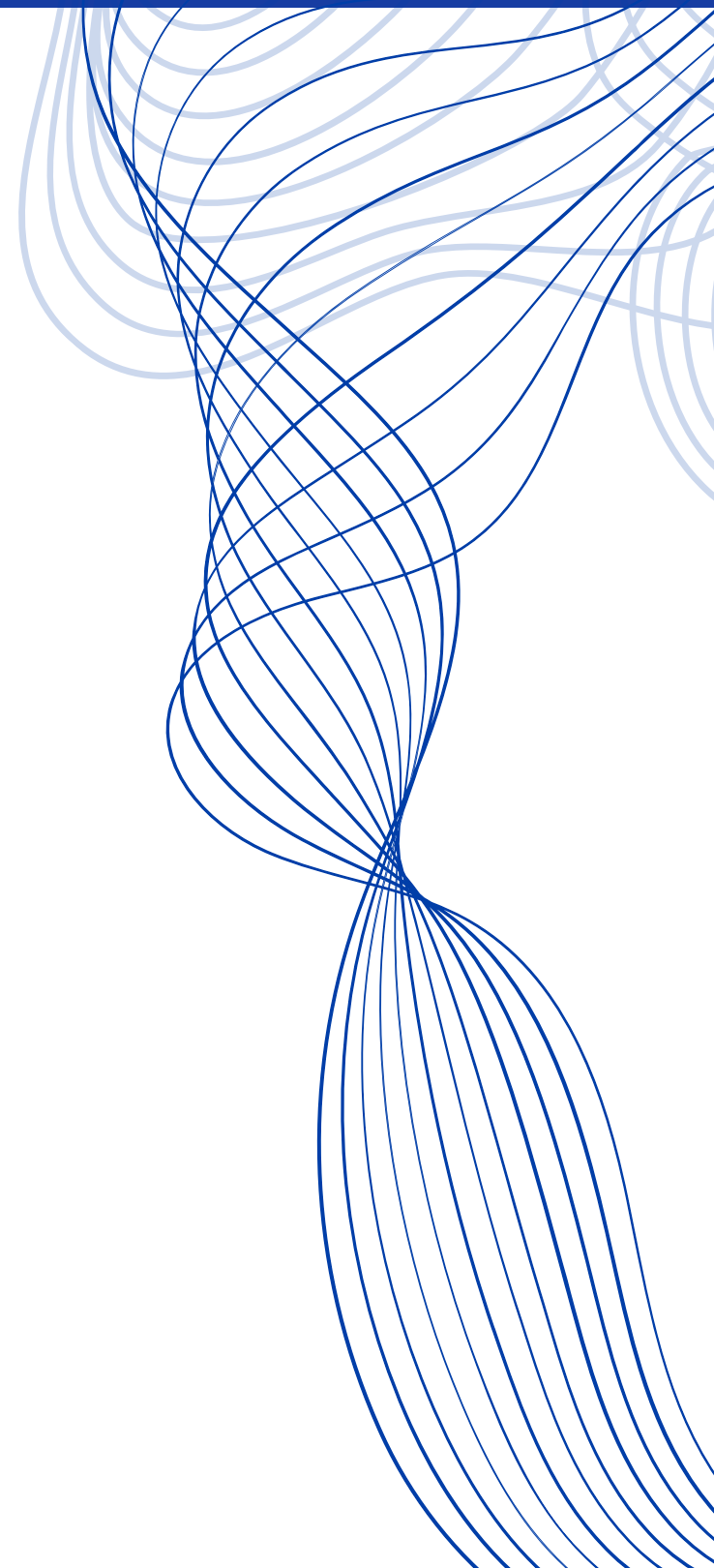
- **Bùi Mai Phương** **030238220195**
- **Lê Thị Thảo Hiền** **030238220059**
- **Lê Thị Thùy Linh** **030238220107**
- **Nguyễn Ngọc Hồng Hòa** **030238220067**
- **Lê Anh Vũ** **030238220300**



NỘI DUNG

I. GIỚI THIỆU ĐỀ TÀI

II. NỘI DUNG CHÍNH



I. PHẦN GIỚI THIỆU ĐỀ TÀI

I.1. Bối cảnh, tầm quan trọng của phân tích bình luận người dùng trong kỷ nguyên số



Thương mại điện tử đang bùng nổ với internet và công nghệ.

Phân tích bình luận người dùng là chìa khóa quan trọng

I. PHẦN GIỚI THIỆU ĐỀ TÀI

I.2. Lý do chọn đề tài:



Số lượng lớn bình luận trên các trang thương mại điện tử, cần phân tích cảm xúc khách hàng cần sự tự động hóa.



Thách thức trong việc xử lý ngôn ngữ tự nhiên, đặc biệt là tiếng Việt.



Quyết định thực hiện đề tài “Phân tích cảm xúc của khách hàng dựa trên khía cạnh bình luận sử dụng mô hình học sâu”.

I. PHẦN GIỚI THIỆU ĐỀ TÀI

I.3. Tính cấp thiết của đề tài



Là lĩnh vực nghiên cứu quan trọng, mang lại nhiều ứng dụng thực tiễn.



Hiểu rõ cảm xúc khách hàng giúp tối ưu hóa trải nghiệm mua sắm và tăng cường quan hệ khách hàng.



Giúp áp dụng các giải pháp kinh doanh hiệu quả và nắm bắt nhu cầu của khách hàng trong môi trường cạnh tranh hiện nay.

I. PHẦN GIỚI THIỆU ĐỀ TÀI

I.4. Mục tiêu nghiên cứu

- Thu thập và xử lý dữ liệu từ Tiki.vn để tạo bộ dữ liệu huấn luyện
- Sử dụng mô hình học sâu để phân tích cảm xúc theo khía cạnh trong bình luận của khách hàng
- Thực nghiệm và đánh giá hiệu suất mô hình bằng các độ đo chuẩn

I.5. Phạm vi nghiên cứu

- Dữ liệu thu thập từ Tiki.vn – một nền tảng thương mại điện tử phổ biến tại Việt Nam

II. Nội Dung Chính

Phân tích cảm xúc – Phân tích cảm xúc theo khía cạnh

Phân tích cảm xúc

Định nghĩa: Hay còn gọi là khai thác ý kiến, là lĩnh vực nghiên cứu phân tích ý kiến, cảm xúc, đánh giá, nhận định, thái độ của mọi người và cảm xúc đối với các thực thể như sản phẩm, dịch vụ, tổ chức, cá nhân, vấn đề, sự kiện, chủ đề và thuộc tính của họ.

(Bing Liu, 2012)

Ví dụ: Chất lượng sản phẩm tốt → cảm xúc tích cực
Thái độ nhân viên khó chịu → cảm xúc tiêu cực

II. Nội Dung Chính

Phân tích cảm xúc – Phân tích cảm xúc theo khía cạnh

Phân tích cảm xúc theo khía cạnh (ABSA)

Là quá trình **phân loại cảm xúc** của các thuật ngữ khía cạnh (aspect terms) được trích xuất từ bình luận, xác định liệu mỗi khía cạnh mang tính chất **tích cực, tiêu cực hay trung lập** về sản phẩm, dịch vụ dựa trên các bình luận đánh giá.

II. Nội Dung Chính

Phân tích cảm xúc – Phân tích cảm xúc theo khía cạnh

Phân tích cảm xúc theo khía cạnh (ABSA)

Tuy nhiên trong các bình luận có thể không có các khía cạnh (**Aspect Terms**) => dùng nhóm khía cạnh (**Aspect Category**) để thay thế

Ví dụ: Cũng được nhưng ồn
Nồi lớn hơn mình nghĩ, dùng ồn

II. Nội Dung Chính

Phân tích cảm xúc – Phân tích cảm xúc theo khía cạnh

Các nhóm khía cạnh phổ biến (aspect category)

- Chất lượng
- Giao hàng
- Hình thức
- Đóng gói
- Giá tiền
- Đánh giá chung
- Khía cạnh khác

Ví dụ

“Đôi giày này **tốt** nhưng **giao hàng quá chậm.**”

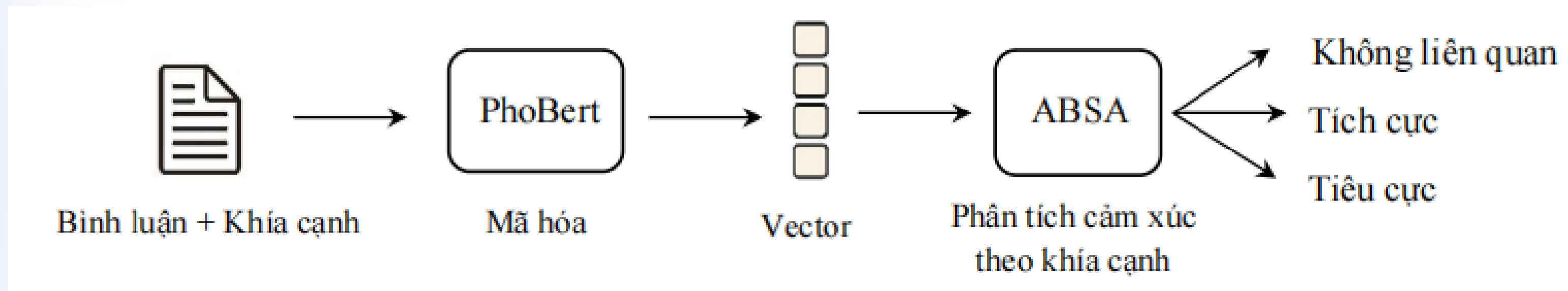
=> Chất lượng là cảm xúc tích cực (tốt)

=> Giao hàng là cảm xúc tiêu cực (quá chậm)

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Xây dựng mô hình phân lớp



**Mô hình phân tích cảm xúc dựa trên
khía cạnh sử dụng mô hình PhoBERT để chuyển đổi văn bản thành vector**

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Thu thập dữ liệu:

- **Xác định và thu thập bình luận người dùng từ trang thương mại điện tử Tiki.vn**
- **Dữ liệu được thu thập đầy đủ đa dạng, khách quan và phản ánh ý kiến và cảm xúc của người dùng**
- **Thu thập dữ liệu tự động bằng Python, kết hợp thư viện Selenium (mô phỏng trình duyệt) và BeautifulSoup (phân tích HTML).**

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Tiền xử lý dữ liệu:

- Loại bỏ các ký tự đặc biệt, dấu câu không cần thiết
- Loại bỏ những bình luận không hợp lệ hoặc không có ý nghĩa
- Xử lý bình luận trùng lặp
- Chuyển đổi các từ thành chữ thường
- Tách từ (Tokenization) bằng thư viện Underthesea – xử lý tách từ tiếng Việt thay vì chỉ dựa vào khoảng trắng

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Tiền xử lý dữ liệu

Tách từ : sử dụng thư viện Underthesea để tách từ tiếng Việt thành các đơn vị từ có nghĩa

Ví dụ: Câu ban đầu: “Giày có thiết kế đẹp nhưng giá cao.”

Sau khi tách từ: “Giày có thiết_kế đẹp nhưng giá cao.”

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Gán nhãn:

- **Xác định các nhóm khía cạnh (aspect category)**
- **Đánh nhãn cảm xúc cho các nhóm khía cạnh liên quan dựa trên nội dung bình luận**
- **Gán nhãn cảm xúc theo khía cạnh sử dụng API ChatGPT theo tập khía cạnh xác định sẵn**

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Biến đổi vector:

- **Sử dụng Word Embeddings**

- + **Word2Vec**

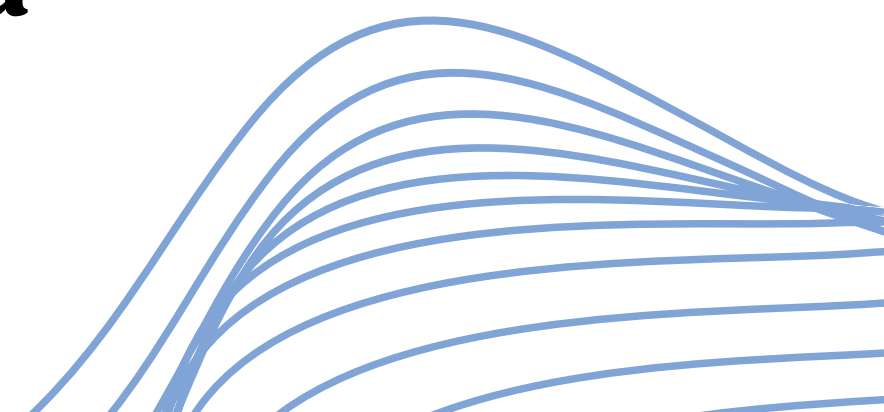
- + **Glove**

- + **FastText**

- + **ELMo**

- **Sử dụng BERT (Bidirectional Encoder Representations from Transformers)**

- + **Là một mô hình học sâu có thể biến đổi từng từ thành vector theo ngữ cảnh của câu**



II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Biến đổi từ thành vector : Pre-training model



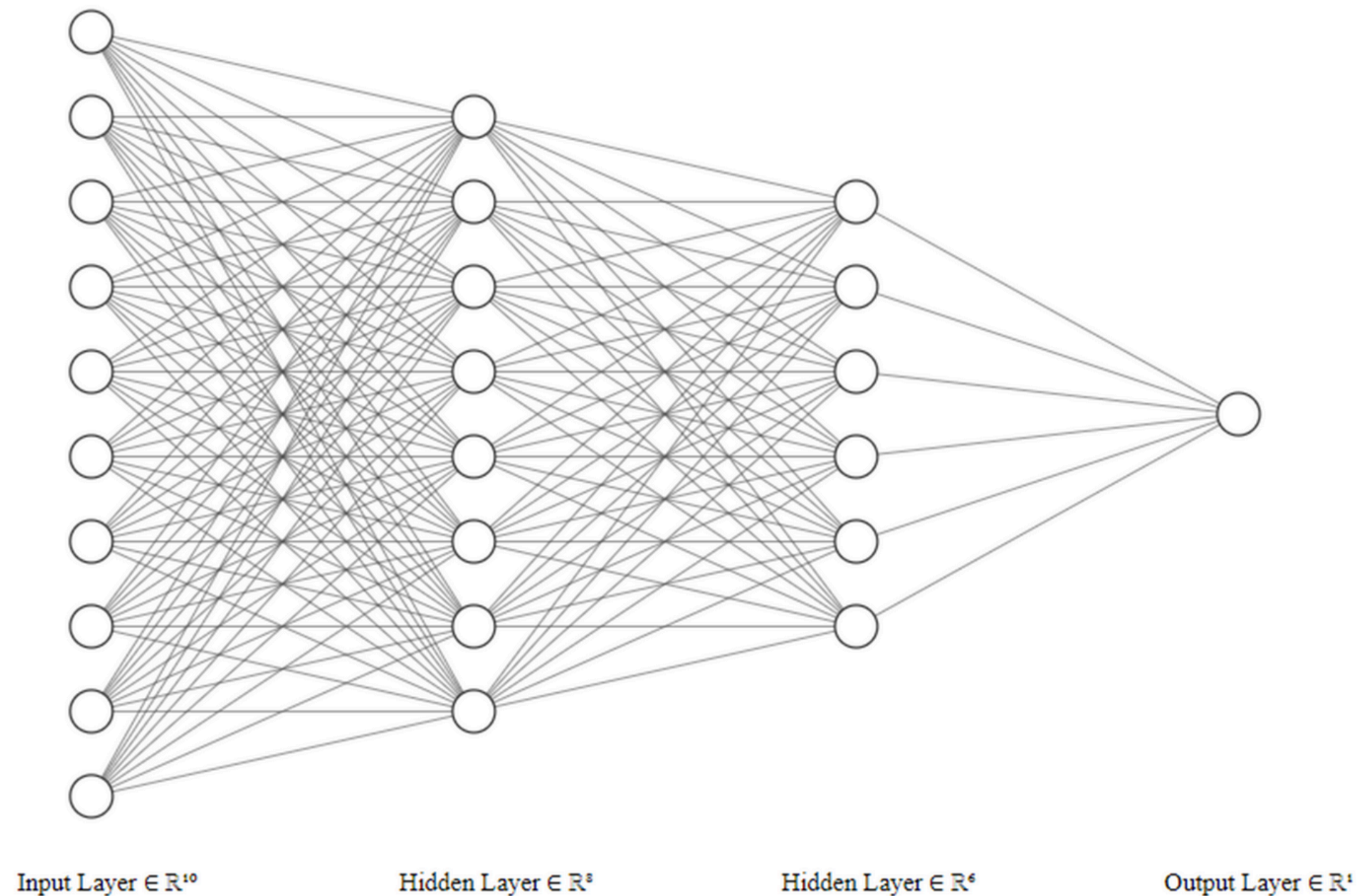
PhoBert

- Là mô hình ngôn ngữ tiền huấn luyện dành riêng cho tiếng Việt, dùng để biến đổi văn bản thành các vector đặc trưng.
- Giúp cải thiện hiệu suất cho tác vụ phân tích cảm xúc.

II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

- **Mô hình phân lớp**



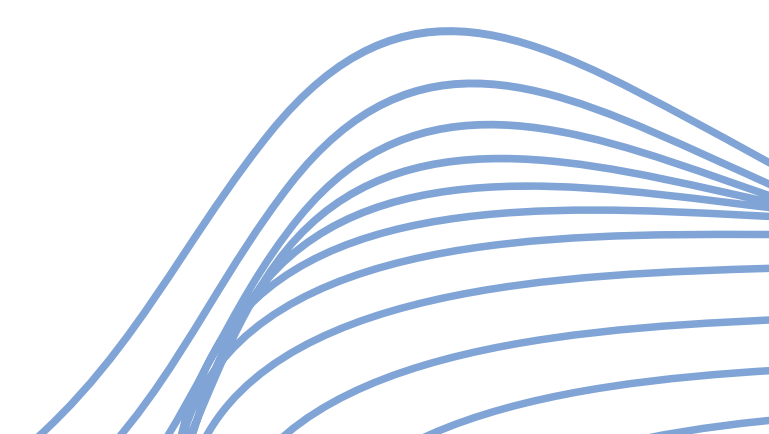
II. Nội Dung Chính

Các bước thực hiện đề tài nghiên cứu

Xây dựng mô hình phân lớp

Phân chia dữ liệu Training và Testing

Cross-validation (Stratified K-Fold)

- Chia dữ liệu thành k phần ($k = 5$).
 - Huấn luyện mô hình 5 lần, mỗi lần dùng 1 phần để kiểm tra, 4 phần còn lại để huấn luyện.
 - Trung bình kết quả các lần kiểm tra để có đánh giá ổn định, khách quan và hạn chế overfitting.
- 

II. Nội Dung Chính

Số lượng mẫu dữ liệu

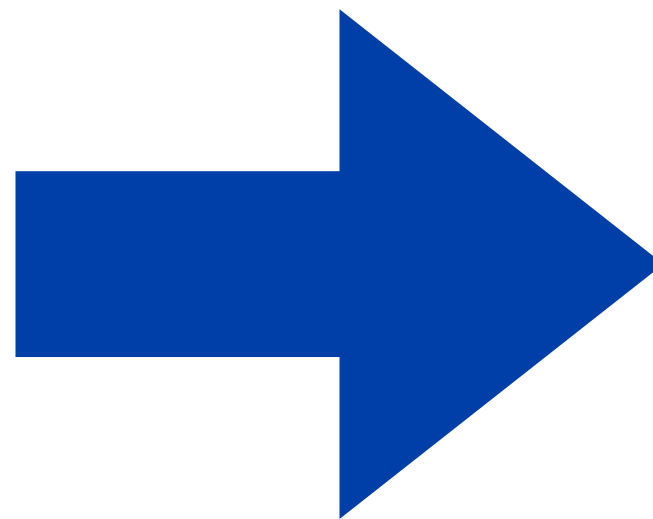
Khía cạnh	Số lượng mẫu	
	Tích cực	Tiêu cực
CHẤT LƯỢNG	3659	3106
GIAO HÀNG	2241	1137
HÌNH THỨC	1989	471
ĐÓNG GÓI	1302	431
GIÁ TIỀN	1076	254
ĐÁNH GIÁ CHUNG	2186	1065
KHÍA CẠNH KHÁC	607	2194

II. Nội Dung Chính

Độ đo đánh giá

- Thực hiện đánh giá mô hình phân tích cảm xúc thông qua các độ đo dưới đây :

1. Accuracy
2. Precision
3. Recall
4. F1-score



Các thước đo này giúp đánh giá toàn diện hiệu suất của mô hình

II. Nội Dung Chính

Kết quả

Khía cạnh	Nhãn lớp	Accuracy	Precision	Recall	F1-score
CHẤT LƯỢNG	Không liên quan	0.8837 ± 0.0068	0.8111 ± 0.0198	0.8197 ± 0.0142	0.8152 ± 0.0084
	Tích cực		0.9150 ± 0.0082	0.9235 ± 0.0124	0.9192 ± 0.0078
	Tiêu cực		0.9126 ± 0.0115	0.8934 ± 0.0136	0.9029 ± 0.0098
GIAO HÀNG	Không liên quan	0.9472 ± 0.0033	0.9624 ± 0.0035	0.9634 ± 0.0046	0.9629 ± 0.0028
	Tích cực		0.9582 0.0058	0.9607 ± 0.0099	0.9594 ± 0.0031
	Tiêu cực		0.8377 0.0227	0.8276 ± 0.0174	0.8324 ± 0.0141
HÌNH THỨC	Không liên quan	0.8904 ± 0.0080	0.9336 0.0057	0.9249 ± 0.0085	0.9292 ± 0.0054
	Tích cực		0.7974 ± 0.0126	0.8527 ± 0.0170	0.8241 ± 0.0142
	Tiêu cực		0.5973 ± 0.0771	0.5032 ± 0.0148	0.5436 ± 0.0241
ĐÓNG GÓI	Không liên quan	0.9625 ± 0.0013	0.9815 ± 0.0021	0.9749 ± 0.0015	0.9782 ± 0.0005
	Tích cực		0.9346 ± 0.0173	0.9616 ± 0.0105	0.9478 ± 0.0105
	Tiêu cực		0.7029 ± 0.0447	0.7286 ± 0.0366	0.7147 ± 0.0305
GIÁ TIỀN	Không liên quan	0.9784 ± 0.0069	0.9918 ± 0.0030	0.9866 ± 0.0050	0.9892 ± 0.0039
	Tích cực		0.9253 ± 0.0271	0.9526 ± 0.0246	0.9387 ± 0.0252
	Tiêu cực		0.7727 ± 0.0625	0.8108 ± 0.0575	0.7908 ± 0.0549
ĐÁNH GIÁ CHUNG	Không liên quan	0.7843 ± 0.0078	0.8353 ± 0.0077	0.8449 ± 0.0180	0.8399 ± 0.0074
	Tích cực		0.7193 ± 0.0266	0.7356 ± 0.0378	0.7266 ± 0.0202
	Tiêu cực		0.5837 ± 0.0166	0.5136 ± 0.0520	0.5457 ± 0.0357
KHÓA CẠNH KHÁC	Không liên quan	0.8298 ± 0.0075	0.8696 ± 0.0101	0.9012 ± 0.0121	0.8850 ± 0.0065
	Tích cực		0.5928 ± 0.0116	0.4993 ± 0.0473	0.5412 ± 0.0300
	Tiêu cực		0.7541 ± 0.0246	0.7010 ± 0.0345	0.7258 ± 0.0149

II. Nội Dung Chính

So sánh với các công trình liên quan và đề xuất cải tiến

- Ý nghĩa khoa học và ứng dụng của đề tài nghiên cứu.

Kết quả thực nghiệm của đề tài:

- Đánh giá sản phẩm cho thấy mô hình đạt hiệu suất phân lớp khả quan, đặc biệt đối với các khía cạnh phổ biến như giao hàng, giá tiền và đóng gói.
- Độ chính xác và F1-score cao ở những khía cạnh này phản ánh khả năng học của mô hình khi xử lý ngữ liệu tiếng Việt trong bối cảnh thực tế.
- Những kết quả tích cực này cho thấy tiềm năng ứng dụng của mô hình trong các hệ thống giám sát tự động cảm xúc khách hàng, góp phần hỗ trợ doanh nghiệp theo dõi trải nghiệm người dùng theo thời gian thực, phát hiện kịp thời các vấn đề phát sinh và đưa ra quyết định chiến lược hiệu quả hơn.

II. Nội Dung Chính

So sánh với các công trình liên quan và đề xuất cải tiến

- **Hướng phát triển tiềm năng trong tương lai**
 - **Mở rộng và cân bằng dữ liệu huấn luyện để nâng cao hiệu suất mô hình.**
 - **Tích hợp trích xuất khía cạnh tự động**
 - **Mở rộng phạm vi ứng dụng nhằm kiểm chứng và gia tăng giá trị thực tiễn.**

TÀI LIỆU THAM KHẢO

Dang, C. N., Moreno-García, M. N., De la Prieta, F., et al. (2023). Sentiment analysis for vietnamese–based hybrid deep learning models. International Conference on Hybrid Artificial Intelligence Systems.

Devlin, J., Chang, M.-W., Lee, K., et al. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers).

Ho, T., Bui, H. M., & Phung, T. K. (2023). A hybrid model for aspect-based sentiment analysis on customer feedback: research on the mobile commerce sector in Vietnam. International Journal of Advances in Intelligent Informatics, 9(2).

Joulin, A., Grave, E., Bojanowski, P., et al. (2016). Bag of Tricks for Efficient Text Classification. arXiv preprint arXiv:.01759.

TÀI LIỆU THAM KHẢO

- Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463): Springer.**
- Mao, Y., Liu, Q., Zhang, Y., et al. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University-Computer, 36(4), 102048.**
- Mikolov, T., Chen, K., Corrado, G., et al. (2013). Efficient estimation of word representations in vector space. arXiv.**
- Nguyen, D. Q., & Nguyen, A. T. (2020). PhoBERT: Pre-trained language models for Vietnamese. arXiv preprint arXiv:.00744.**
- Tran, Q.-L., Le, P. T. D., & Do, T.-H. (2022). Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites. Proceedings of the 36th Pacific Asia conference on language, information and computation**



**CẢM ƠN HỘI ĐỒNG ĐÃ
LẮNG NGHE**