

Phân tích cảm xúc khách hàng về sản phẩm thời trang trực tuyến bằng một số mô hình mạng nơ-ron

Vương Thị Thảo Bình¹

Ngày nhận bài: 07/5/2026 | Ngày gửi phản biện: 11/5/2026 | Ngày duyệt đăng: 12/6/2026

Tóm tắt: Nghiên cứu này xem xét hiệu quả của ba mô hình mạng nơ-ron MLP, LSTM-RNN và CNN trong việc phân tích cảm xúc của người tiêu dùng trong ngành thời trang trực tuyến tại Việt Nam. Sử dụng hơn 19.000 đánh giá trên Shopee (giai đoạn tháng 1–2 năm 2023), dữ liệu được tiền xử lý bằng VnCoreNLP và mã hóa thông qua PhoBERT. Các mô hình được đánh giá dựa trên các chỉ số độ chính xác, độ chụm, độ thu hồi, F1-score và ROC-AUC. Cả ba mô hình đều đạt hiệu suất tương đương nhau: 88,49% (MLP), 87,87% (LSTM-RNN) và 87,74% (CNN), cho thấy tính khả thi của việc phân loại cảm xúc tiếng Việt tự động phục vụ cho các ứng dụng kinh doanh.

Từ khóa: Mạng nơ-ron, Perceptron đa lớp, LSTM-RNN, Mạng nơ-ron tích chập.

Customer Sentiment Analysis of Online Fashion Products Using Several Neural Network Models

Abstract: This study examines the effectiveness of three neural network models - MLP, LSTM-RNN, and CNN - in analyzing consumer sentiment in the Vietnamese online fashion industry. Using over 19,000 Shopee reviews (January–February 2023), preprocessed with VnCoreNLP and encoded via PhoBERT, models were evaluated on accuracy, precision, recall, F1-score, and ROC-AUC. All three achieved comparable performance: 88.49 percent (MLP), 87.87 percent (LSTM-RNN), and 87.74 percent (CNN), demonstrating the feasibility of automated Vietnamese sentiment classification for business applications.

Keywords: Neural network, Multilayer Perceptron, LSTM-RNN, Convolutional Neural Network.

1. Đặt vấn đề

Thị trường thương mại điện tử Việt Nam đã tăng trưởng mạnh mẽ, trong đó Shopee dẫn đầu với doanh thu khoảng 91 nghìn tỷ đồng (Metric, 2022), đồng thời tiếp tục được thúc đẩy bởi đại dịch COVID-19 (Euromonitor, 2021). Khối lượng lớn các đánh giá của khách hàng tạo thành một nguồn dữ liệu có giá trị cho doanh nghiệp - đặc biệt trong ngành thời trang, lĩnh vực cần phản ứng nhanh trước sự thay đổi liên tục của nhu cầu tiêu dùng (Cook và Yurchisin, 2017). Khi có tới 92% người tiêu dùng tin tưởng vào các khuyến nghị từ người dùng khác (Nielsen, 2012), và các đánh giá tiêu cực có thể làm lộ rõ những khiếm khuyết quan trọng của sản phẩm (Berger và cộng sự, 2010), việc phân loại cảm xúc tự động mang lại giá trị chiến lược đáng kể cho doanh nghiệp (Agarap, 2018).

Mặc dù phân tích cảm xúc nhận được nhiều sự quan tâm trên thế giới, các nghiên cứu về đánh giá thời trang bằng tiếng Việt vẫn còn hạn chế. Các nghiên cứu hiện tại chủ yếu tập trung vào dữ liệu tiếng Anh hoặc sử dụng các phương pháp học máy truyền thống mà chưa xử lý đầy đủ sự phức tạp ngôn ngữ tiếng Việt (Le và Huh, 2021). Đồng thời, rất ít nghiên cứu tiến hành so sánh có hệ thống các kiến trúc mạng nơ-ron trên dữ liệu thương mại điện tử tiếng Việt. Nghiên cứu này nhằm lấp đầy khoảng trống đó

¹ Trường Đại học Ngoại thương; Email: vuongthithaobinh.hcmc@ftu.edu.vn

bằng cách áp dụng các mô hình MLP, LSTM-RNN và CNN kết hợp với biểu diễn ngôn ngữ dựa trên PhoBERT để phân loại cảm xúc của người tiêu dùng trong ngành thời trang trực tuyến tại Việt Nam.

Phần còn lại của bài viết được tổ chức như sau. Phần II mô tả quy trình thu thập dữ liệu, quy trình tiền xử lý, phương pháp biểu diễn ngôn ngữ và cấu hình huấn luyện mô hình. Phần III trình bày và thảo luận kết quả thực nghiệm, bao gồm đánh giá so sánh giữa ba mô hình. Phần IV đưa ra kết luận, tóm tắt các phát hiện chính, hàm ý thực tiễn đối với doanh nghiệp thương mại điện tử Việt Nam, cũng như định hướng cho các nghiên cứu trong tương lai.

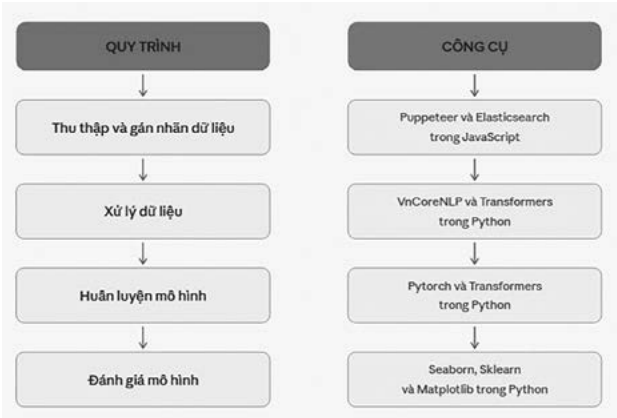
2. Phương pháp nghiên cứu

Nghiên cứu này sử dụng các đánh giá sản phẩm thời trang từ nền tảng thương mại điện tử Shopee, được thu thập trong khoảng thời gian từ ngày 10/01 đến ngày 26/02/2023. Dữ liệu gốc được gán nhãn theo thang điểm từ 1 đến 5 dựa trên cảm xúc, trong đó các mức điểm thấp thể hiện đánh giá tiêu cực và các mức điểm cao thể hiện đánh giá tích cực. Mặc dù thừa nhận rằng điểm đánh giá sao và cảm xúc trong văn bản là hai khái niệm khác nhau, các nghiên cứu trước đây đã chứng minh rằng điểm đánh giá có thể được sử dụng như một phương pháp gán nhãn thực tiễn trong các bộ dữ liệu thương mại điện tử quy mô lớn, nơi việc gán nhãn thủ công toàn bộ là không khả thi (Jagdale và cộng sự, 2019; Bui, 2021). Nguy cơ nhiễu nhãn chủ yếu tập trung ở các trường hợp trung gian - đặc biệt là các đánh giá 3 sao, vốn có thể phản ánh cảm xúc trung tính thay vì tích cực. Hạn chế tiềm ẩn này được thừa nhận trong nghiên cứu hiện tại và sẽ được thảo luận thêm ở Phần IV. Sau khi gán nhãn, dữ liệu được tiến hành tiền xử lý và biểu diễn ngôn ngữ.

Để tiền xử lý văn bản, nghiên cứu thực hiện ba bước: (1) chuyển đổi các ký tự ngoại lai sang ký tự tiếng Việt chuẩn bằng bảng ánh xạ được xây dựng thủ công; (2) loại bỏ ký tự đặc biệt, biểu tượng cảm xúc (emoji) và các ký hiệu Unicode bất thường nhằm loại bỏ các token không tồn tại trong từ vựng của PhoBERT; và (3) thực hiện tách từ tiếng Việt bằng công cụ VnCoreNLP. Do tiếng Việt là ngôn ngữ đơn lập, trong đó một từ có thể bao gồm nhiều âm tiết, việc tách từ là cần thiết để xác định ranh giới từ. Sau khi tách từ, các từ đa âm tiết như “chất lượng” và “sản phẩm” được gộp thành các token đơn - “chất_lượng” và “sản_phẩm” - nhằm bảo toàn ý nghĩa ngữ nghĩa trong quá trình mã hóa tiếp theo.

Đối với biểu diễn ngôn ngữ, mô hình tiền huấn luyện “vinai/phobert-base” từ thư viện Transformers được sử dụng để mã hóa mỗi đánh giá thành các vector số. Các đánh giá được chuẩn hóa về độ dài 256 token thông qua kỹ thuật padding trước khi mã hóa. Bộ tokenizer của PhoBERT tạo ra vector token_type_id và vector attention_mask, sau đó được xử lý bởi mô hình PhoBERT-base để tạo ra embedding có kích thước 768 chiều cho mỗi token - từ đó hình thành ma trận 256×768 cho mỗi đánh giá. Biểu diễn ngữ cảnh phong phú này được sử dụng làm đầu vào chung cho cả ba mô hình phân loại.

Hình 1. Tổng quan quy trình nghiên cứu



Ba kiến trúc mạng nơ-ron được áp dụng như các tầng phân loại phía trên bộ mã hóa PhoBERT dùng chung. MLP đóng vai trò là mô hình cơ sở, hoạt động trên biểu diễn token [CLS] đã được tổng hợp, xem

đầu vào đã mã hóa như một vector đặc trưng có độ dài cố định mà không mô hình hóa thứ tự tuần tự. LSTM xử lý toàn bộ chuỗi gồm 256 embedding token từ PhoBERT thông qua các cơ chế cổng - gồm cổng đầu vào, cổng quên và cổng đầu ra - nhằm nắm bắt các phụ thuộc dài hạn giữa các token trong đánh giá, điều này đặc biệt hữu ích với các đánh giá tiếng Việt dài có cảm xúc trải dài qua nhiều mệnh đề. CNN áp dụng các bộ lọc tích chập 1 chiều với nhiều kích thước kernel khác nhau (3, 4 và 5) lên các embedding token để phát hiện các mẫu n-gram cục bộ - chẳng hạn như cụm phủ định hoặc từ nhấn mạnh - bất kể vị trí của chúng trong chuỗi. Cả ba mô hình đều được huấn luyện trong 10 epoch bằng bộ tối ưu Adam với tốc độ học 2×10^{-5} và kích thước batch là 64.

3. Kết quả và thảo luận

Nghiên cứu đã thu thập tổng cộng 19.942 quan sát, được gán nhãn thành hai nhóm: nhóm tích cực bao gồm các đánh giá từ ba sao trở lên, và nhóm tiêu cực bao gồm các đánh giá dưới ba sao. Mô hình sử dụng 15.953 quan sát, trong đó tập huấn luyện gồm 11.964 quan sát (8.962 đánh giá tích cực và 3.002 đánh giá tiêu cực), còn tập kiểm tra gồm 3.989 quan sát (2.987 đánh giá tích cực và 1.002 đánh giá tiêu cực).

Bảng 1. Bảng so sánh kết quả của ba mô hình

Mô hình	Độ chính xác (Testing Accuracy)	Độ chuẩn xác (Testing Precision)	Độ phủ (Testing recall)	Điểm F1 (Testing F1 Score)	ROC-AUC
MLP	88,49%	93,68%	90,76%	92,20%	0,947
LSTM – RNN	87,87%	94,71%	88,75%	92,63%	0,947
CNN	87,74%	92,48%	91,03%	91,75%	0,945

Bảng 1 tóm tắt hiệu suất kiểm tra của các mô hình MLP, LSTM-RNN và CNN. Cả ba mô hình đều đạt hiệu quả cao, với độ chính xác trên tập kiểm tra (Testing Accuracy) dao động từ 87,74% (CNN) đến 88,49% (MLP), và chỉ số F1 nằm trong khoảng từ 91,75% (CNN) đến 92,63% (LSTM-RNN), cho thấy khả năng mạnh mẽ trong việc phân loại các đánh giá sản phẩm thời trang bằng tiếng Việt. Ngoài ra, các chỉ số ROC-AUC cũng đạt mức cao và tương đối đồng đều giữa ba mô hình (0,945–0,947), khẳng định khả năng phân biệt lớp rất tốt của các mô hình.

Kiểm định ANOVA đã được thực hiện để đánh giá sự khác biệt về hiệu suất giữa các mô hình dựa trên các chỉ số như Training Loss, Validation Loss, Accuracy, Precision, Recall và F1 Score. Kết quả cho thấy, tất cả các giá trị p-value đều lớn hơn 0,05, cho thấy không tồn tại sự khác biệt có ý nghĩa thống kê giữa các mô hình. Điều này hàm ý rằng ba mô hình có hiệu suất tương đương nhau và việc lựa chọn mô hình có thể dựa trên các yếu tố khác như độ phức tạp của mô hình hoặc tài nguyên tính toán.

Kết quả của nghiên cứu này cũng tương đồng với nghiên cứu của Agarap (2018), trong đó mô hình BRNN được áp dụng cho dữ liệu thời trang tiếng Anh và đạt độ chính xác 88%. Phụng (2019) cũng đã so sánh một số mô hình học máy trên các bình luận du lịch bằng tiếng Việt và báo cáo rằng mô hình Neural Network đạt độ chính xác 79%. So với các nghiên cứu này, các mô hình mạng nơ-ron trong nghiên cứu của chúng tôi cho thấy hiệu suất cạnh tranh hoặc vượt trội trên bộ dữ liệu bình luận thời trang tiếng Việt từ Shopee. Đặc biệt, việc tích hợp các mô hình mạng nơ-ron tiên tiến với quy trình xử lý ngôn ngữ tự nhiên tiếng Việt chuyên sâu dựa trên PhoBERT đã chứng minh hiệu quả trong việc nắm bắt các sắc thái cảm xúc phức tạp của ngôn ngữ trong bối cảnh thương mại điện tử. Mặc dù các nghiên cứu của Jagdale

và cộng sự (2019) và Bui (2021) đạt độ chính xác rất cao (trên 90%), các nghiên cứu trước đó sử dụng dữ liệu tiếng Anh hoặc các nền tảng/sản phẩm khác, qua đó làm nổi bật đóng góp của nghiên cứu này trong việc đạt được hiệu suất tương tự trên dữ liệu tiếng Việt có độ phức tạp cao. Kết quả nghiên cứu cũng tương đồng với nghiên cứu của Hoàng và cộng sự (2023) có thực hiện phân tích cảm xúc cho mặt hàng điện tử.

Tóm lại, các mô hình MLP, LSTM-RNN và CNN đều phân loại hiệu quả cảm xúc trong các đánh giá thời trang tiếng Việt, với độ chính xác (Accuracy) trên 87% và chỉ số F1-Score trên 91%. Mặc dù tồn tại một số khác biệt nhỏ về các chỉ số đánh giá, hiệu suất của các mô hình là tương đương nhau ở mức cao, và việc lựa chọn mô hình tối ưu sẽ phụ thuộc vào tiêu chí ưu tiên cụ thể từng bài toán kinh doanh.

4. Kết luận

Nghiên cứu này đã áp dụng ba mô hình mạng nơ-ron - MLP, LSTM-RNN và CNN - để phân loại các đánh giá thời trang trực tuyến bằng tiếng Việt thành nhóm tích cực hoặc tiêu cực, đạt độ chính xác trên tập kiểm tra lần lượt là 88,49%, 87,87% và 87,74%. Kết hợp với quy trình xử lý ngôn ngữ tự nhiên toàn diện dựa trên PhoBERT, các kết quả này cho thấy gần 9 trên 10 bình luận của khách hàng có thể được phân loại tự động với độ tin cậy cao (Jagdale và cộng sự, 2019), mang lại giá trị trực tiếp cho nhiều ứng dụng kinh doanh như xếp hạng sản phẩm, cải thiện chất lượng, hệ thống gợi ý và nghiên cứu thị trường. Việc triển khai các mô hình này mang lại nhiều lợi ích rõ rệt cho các doanh nghiệp thương mại điện tử tại Việt Nam, bao gồm giảm chi phí đánh giá thủ công, phát hiện nhanh hơn các vấn đề về chất lượng sản phẩm và khai thác thông tin người tiêu dùng kịp thời hơn. Sau khi được triển khai, các mô hình có khả năng mở rộng hiệu quả theo quy mô dữ liệu - đây là một lợi thế quan trọng đối với các nền tảng thời trang có lưu lượng truy cập cao, nơi hàng nghìn đánh giá được tạo ra mỗi ngày.

Về mặt ứng dụng thực tiễn, kết quả nghiên cứu có thể được khai thác theo bốn hướng chính trong hoạt động kinh doanh thương mại điện tử. *Thứ nhất*, các mô hình có thể hỗ trợ xây dựng hệ thống xếp hạng sản phẩm và đơn vị kinh doanh dựa trên cảm xúc thực tế của người mua. Điểm xếp hạng có thể được tính theo công thức tổng hợp giữa số lượng đánh giá tích cực và tiêu cực trên tổng số đánh giá nhận được, thay vì chỉ dựa vào điểm sao trung bình như hiện nay. Việc trực quan hóa chỉ số này giúp khách hàng mới có cơ sở tin tưởng hơn khi đưa ra quyết định mua sắm, đồng thời cải thiện khả năng hiển thị của sản phẩm trên các công cụ tìm kiếm. *Thứ hai*, mô hình có thể phục vụ trực tiếp cho đánh giá và cải thiện chất lượng sản phẩm. Bằng cách trích xuất và phân tích các đánh giá tiêu cực theo từ khóa lặp lại, doanh nghiệp có thể xác định chính xác các thuộc tính sản phẩm đang bị phản hồi xấu - chẳng hạn chất liệu vải, kích cỡ, hay quy cách đóng gói - từ đó đưa ra điều chỉnh kịp thời trong vòng đời phát triển sản phẩm. Đặc biệt trong ngành thời trang, nơi xu hướng tiêu dùng thay đổi nhanh, việc cập nhật mô hình định kỳ giúp doanh nghiệp bắt kịp phản hồi của thị trường mà không cần tốn chi phí khảo sát thủ công. *Thứ ba*, phân loại cảm xúc có thể được tích hợp vào hệ thống gợi ý sản phẩm cá nhân hóa. Thay vì chỉ dựa vào lịch sử mua hàng, mô hình cho phép hệ thống nhận diện những nhóm sản phẩm mà từng khách hàng thể hiện thái độ tích cực qua đánh giá, từ đó gợi ý các sản phẩm tương tự phù hợp hơn với sở thích thực sự của họ. Trên các nền tảng có lưu lượng lớn như Shopee, cách tiếp cận này có tiềm năng gia tăng tỷ lệ chuyển đổi và giảm hiện tượng "đuôi dài" - tức những sản phẩm phù hợp với nhu cầu cụ thể nhưng ít được người mua tìm thấy. *Thứ tư*, trong bối cảnh nghiên cứu thị trường, mô hình có thể xử lý khối lượng lớn phản hồi dạng văn bản từ các chiến dịch khảo sát, phân loại tự động thái độ của từng nhóm khách hàng mục tiêu. Điều này giúp doanh nghiệp không chỉ tiết kiệm thời gian tổng hợp dữ liệu định tính mà còn có thể phân khúc khách hàng theo cảm xúc, xác định nhóm đang có thiện cảm hay bất mãn với thương hiệu để có chiến lược truyền thông phù hợp. Tóm lại, giá trị của các mô hình được đề xuất không dừng lại ở độ chính xác phân loại, mà nằm ở khả năng mở rộng sang nhiều bài toán kinh

doanh khác nhau sau khi mô hình đã được huấn luyện mà không cần tiến hành huấn luyện lại cho từng ứng dụng cụ thể.

Tuy nhiên, nghiên cứu vẫn tồn tại một số hạn chế. Phân loại nhị phân có thể đơn giản hóa quá mức quan điểm của người tiêu dùng; việc sử dụng điểm đánh giá sao làm nhãn thay thế có thể gây nhiễu nhãn đối với các đánh giá 3 sao ở vùng ranh giới; và các mô hình vẫn gặp khó khăn trong việc xử lý sự mỉa mai hoặc châm biếm trong ngôn ngữ. Hiện tượng quá khớp (overfitting) cũng vẫn tồn tại mặc dù đã áp dụng kỹ thuật oversampling và các kiến trúc tiên tiến. Các nghiên cứu trong tương lai nên xem xét phân tích cảm xúc theo khía cạnh (aspect-based sentiment analysis), dữ liệu đa phương thức, các kỹ thuật điều chuẩn (regularization) mạnh hơn, cũng như gán nhãn thủ công độc lập để xây dựng bộ dữ liệu chuẩn đáng tin cậy hơn. Nhìn chung, nghiên cứu này cho thấy việc kết hợp học sâu tiên tiến với các công cụ xử lý ngôn ngữ tự nhiên chuyên biệt cho tiếng Việt có thể tạo ra một khuôn khổ mạnh mẽ và có khả năng mở rộng cho phân tích cảm xúc, với tiềm năng ứng dụng rộng lớn trong lĩnh vực thương mại điện tử đang phát triển nhanh chóng tại Việt Nam.

Lời cảm ơn: Tác giả xin cảm ơn nhóm nghiên cứu thuộc Đề tài NCKH cấp cơ sở Trường Đại học Ngoại thương, mã số NTCS2022-10, vì những hỗ trợ kỹ thuật và góp ý trong quá trình thực hiện nghiên cứu. Mọi nội dung và kết luận trong bài viết thuộc trách nhiệm của tác giả.

Tài liệu tham khảo

1. Agarap, A. F. (2018). *Statistical analysis on e-commerce reviews, with sentiment classification using bidirectional recurrent neural network (RNN)*. arXiv. <https://arxiv.org/abs/1805.03687>
2. Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29(5), 815–827.
3. Bui, M. H. (2021). Phân tích ý kiến khách hàng trực tuyến dựa theo phương pháp học máy. *Tạp chí Kinh tế và Phát triển*, 291.
4. Cook, S. C., & Yurchisin, J. (2017). Fast fashion environments: Consumer's heaven or retailer's nightmare? *International Journal of Retail & Distribution Management*, 45(2), 143–157.
5. Euromonitor. (2021). *Consumers in 2021: Market research report*. Euromonitor International.
6. Hoàng, N. T. T., Nguyễn, T. C., Bùi, Đ. M., & Vương, T. T. B. (2023). Ứng dụng mô hình Neural Network trong việc phân loại cảm xúc khách hàng tiêu dùng trực tuyến. *Tạp chí Kinh tế Châu Á - Thái Bình Dương*, 649, 28–30.
7. Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. *In Cognitive informatics and soft computing: Proceedings of CISC 2017*. Springer.
8. Le, N.-B.-V., & Huh, J.-H. (2021). Applying sentiment product reviews and visualization for BI systems in Vietnamese e-commerce website: Focusing on Vietnamese context. *Electronics*, 10(20), 2481.
9. Metric. (2022). *Báo cáo tổng quan thị trường thương mại điện tử 2022*. Metric.
10. Nielsen. (2012). *Global trust in advertising and brand messages*. Nielsen.
11. Phung, T. K. (2019). Tiếp cận phương pháp máy học trong khai thác ý kiến khách hàng trực tuyến. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*.