



Ứng dụng phương pháp máy học trong đo lường sự hài lòng của khách hàng dựa trên các bình luận trực tuyến

THÁI KIM PHỤNG^{a,*}, NGUYỄN AN TẾ^a

^a Trường Đại học Kinh tế TP. Hồ Chí Minh

THÔNG TIN	TÓM TẮT
Ngày nhận: 24/09/2022 Ngày nhận lại: 31/03/2023 Duyệt đăng: 21/04/2023 Mã phân loại JEL: C61; C63; C67. Từ khóa: Phương pháp máy học; Đo lường sự hài lòng; Khai thác ý kiến. Keywords: Machine learning; Satisfaction measurement; Opinion mining.	Nghiên cứu này đề xuất ứng dụng phương pháp máy học để đo lường sự hài lòng của khách hàng dựa trên các bình luận trực tuyến. Dữ liệu nghiên cứu là các bình luận của khách du lịch trên trang Agoda.com về các khách sạn tại Việt Nam. Nghiên cứu tiến hành huấn luyện bằng các mô hình máy học: Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regression (LR), và Random Forest (RF) để tìm ra mô hình phân loại tốt nhất, sau đó ứng dụng mô hình này để dự báo sự hài lòng của khách hàng trên toàn bộ tập dữ liệu. Tiếp đến, nghiên cứu sử dụng phương pháp tổng hợp và xếp hạng mức độ hài lòng của khách hàng. Kết quả của nghiên cứu có giá trị tham khảo cho việc đánh giá sự hài lòng của khách hàng ở các lĩnh vực khác. Abstract This study proposes the application of a machine learning method to measure customer satisfaction based on online comments. Research data is the reviews of travelers on Agoda.com about hotels in Vietnam. The study conducts training using machine learning models: Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regression (LR), and Random Forest (RF) to find the best classification model, then apply this model to predict the customer satisfaction classification across the entire dataset. Next, the study uses the method of aggregating and ranking customer

* Tác giả liên hệ.

Email: phungthk@ueh.edu.vn (Thái Kim Phụng), tena@ueh.edu.vn (Nguyễn An Tế)

Trích dẫn bài viết: Thái Kim Phụng, & Nguyễn An Tế. (2023). Ứng dụng phương pháp máy học trong đo lường sự hài lòng của khách hàng dựa trên các bình luận trực tuyến. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 34(5), 46–57.

satisfaction. This study has reference value for measuring customer satisfaction in business fields.

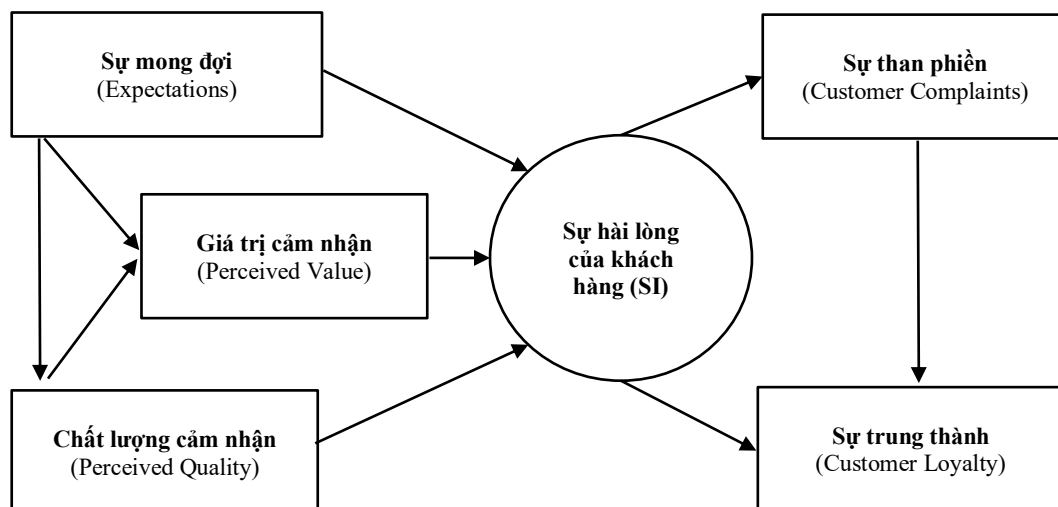
1. Giới thiệu

Với sự phát triển vượt bậc của công nghệ, các nền tảng trực tuyến đã ra đời và tăng trưởng theo cấp số nhân, kéo theo những thách thức về truyền thông do lượng thông tin quá lớn. Khách hàng ngoài việc mua sắm trực tuyến còn có được những thông tin có giá trị về một sản phẩm mà họ định mua thông qua nền tảng trực tuyến. Khách hàng chia sẻ trải nghiệm của họ bằng cách cung cấp phản hồi tạo ra một nguồn thông tin dạng văn bản và quá trình này liên tục tạo ra dữ liệu mỗi ngày. Thông tin do khách hàng cung cấp bao gồm cả văn bản chủ quan và ý nghĩa chứa đựng nhiều thông tin về hành vi, sở thích và sự hài lòng (SHL) đối với các sản phẩm hoặc dịch vụ. Hơn nữa, thông tin này có thể hữu ích cho những khách hàng chưa mua hoặc đang trong quá trình ra quyết định mua sắm (Hussain, 2019). SHL cũng có vai trò lớn trong việc kiểm soát chất lượng, tác động lớn đến các quyết định của doanh nghiệp (Becser & Zoltay-Paprika, 2003). Hiện nay, các doanh nghiệp đánh giá SHL của khách hàng thường dựa trên việc khảo sát, được tiến hành bằng những cách thức như: Phỏng vấn qua điện thoại, phỏng vấn qua thư, phỏng vấn trực tiếp, thảo luận nhóm, khảo sát trực tuyến (tiêu chuẩn ISO 10004). Tuy nhiên, theo Yussupova và cộng sự (2016), thách thức lớn của những cách thức này là nhiều công đoạn thực hiện thủ công, từ việc chuẩn bị bảng câu hỏi, dữ liệu người được phỏng vấn, thu thập dữ liệu, tiến hành phỏng vấn... đến việc trình bày các báo cáo. Những công việc này làm cho đợt khảo sát đo lường SHL trở nên tốn kém. Hơn nữa, việc tổng hợp, phân tích kết quả chỉ diễn ra tại một thời điểm và không cho thấy được xu hướng thay đổi của SHL. Ngoài ra, những cách thức này không thể đo lường SHL liên tục trong dài hạn (Yussupova và cộng sự, 2016). Do vậy, nghiên cứu này đề xuất ứng dụng phương pháp máy học trong việc đo lường SHL của khách hàng dựa trên các bình luận trực tuyến.

2. Cơ sở lý thuyết

2.1. Đo lường sự hài lòng khách hàng

Có nhiều lý thuyết và mô hình để đo lường SHL của khách hàng đã được đề xuất như: Lý thuyết xác nhận kỳ vọng (Expectancy Disconfirmation Theory), phân tích hiệu suất quan trọng (Importance Performance Analysis), lý thuyết công bằng (Equity Theory), lý thuyết phân bổ (Attribution Theory), lý thuyết nhận thức giá trị (Value-Percept Theory), và lý thuyết bất hòa (Dissonance Theory). Một đặc điểm chung của các lý thuyết đánh giá, đo lường này là xem xét SHL như là một khái niệm tương đối, liên quan đến những tiêu chuẩn như: Giá trị, đầu vào, trải nghiệm... ACSI (American Customer Satisfaction Index) là mô hình thường sử dụng nhất, theo đó, SHL của khách hàng được xác định bởi sự kết hợp giữa mong đợi của khách hàng, giá trị và chất lượng nhận được (Wang, 2016).



Hình 1. Mô hình chỉ số hài lòng khách hàng của Mỹ

Nguồn: Wang (2016).

Tại Việt Nam, việc đo lường SHL của khách hàng thường được thực hiện bằng nhiều cách: Phỏng vấn qua điện thoại, phỏng vấn qua thư, phỏng vấn trực tiếp, thảo luận nhóm, khảo sát trực tuyến (theo hướng dẫn của Tiêu chuẩn “Quản lý chất lượng, Sự thỏa mãn của khách hàng, Hướng dẫn theo dõi và đo lường” – ISO 10004).

Một đặc điểm chung của các phương thức trên là rất nhiều công đoạn thực hiện thủ công: Phác thảo bảng câu hỏi, lên danh sách người được phỏng vấn, thu thập dữ liệu, tiến hành phỏng vấn... đến việc trình bày các báo cáo kết quả. Những thủ tục này làm cho đợt khảo sát SHL của khách hàng khá tốn kém. Kết quả SHL của khách hàng được trình bày bằng các chỉ số khó hiểu và khó giải thích. Việc tổng hợp, phân tích kết quả chỉ diễn ra tại một thời điểm và không cho thấy được xu hướng thay đổi của SHL từ khách hàng. Bên cạnh đó, những cách thức này không thể theo dõi SHL của khách hàng liên tục trong dài hạn (Yussupova và cộng sự, 2016). Ngày nay, với sự xuất hiện của các phương tiện truyền thông mới, khách hàng dễ dàng trao đổi thông tin và bình luận về các sản phẩm trên một quy mô lớn. Theo Becser và Zoltay-Paprika (2003), “để khắc phục cách thức đánh giá SHL theo kiểu truyền thống, việc xây dựng một hệ thống thông tin hỗ trợ ra quyết định là cần thiết để xử lý và tổng hợp dữ liệu, sử dụng các mô hình phù hợp để phân tích, và chuẩn bị các thông tin kết quả giúp cho nhà quản trị ra quyết định”.

2.2. Ý kiến bình luận của khách hàng trực tuyến

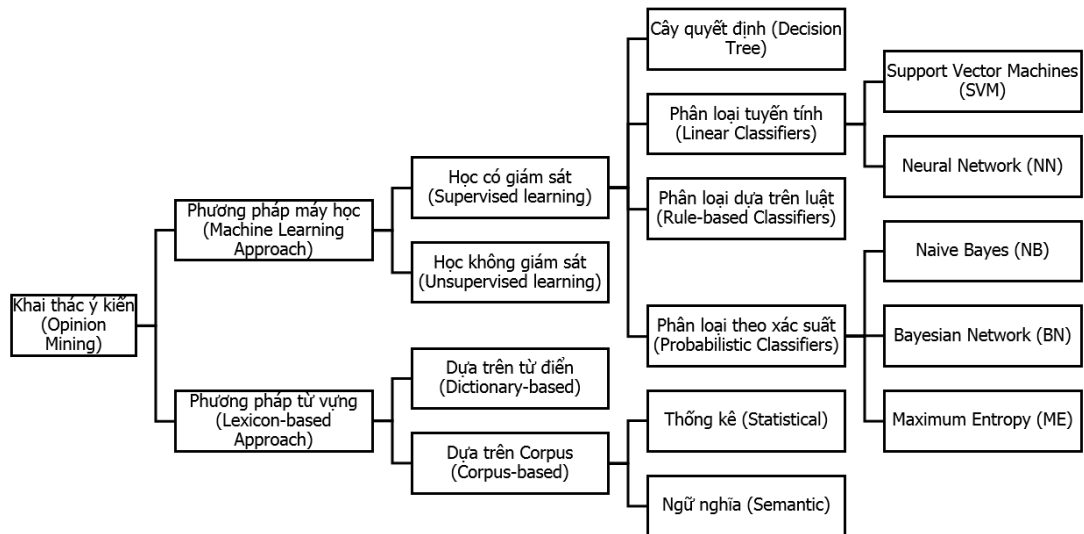
Sự phát triển của các phương tiện truyền thông mới đã khắc phục những trở ngại của phương tiện truyền miệng (Word-of-Mouth). Sự bùng nổ của thế hệ Web 2.0, và chuyển sang nền tảng di động, sau đó là sự ra đời của một số lượng lớn các nền tảng đánh giá sản phẩm trực tuyến (ví dụ: TripAdvisor, Yelp.com, Amazon...). Những nền tảng này cho phép khách hàng có thể đánh giá sản phẩm với nội dung dưới dạng điểm số (Rating) và các ý kiến về sản phẩm (Mudambi & Shuff, 2010). Theo Piccoli và Pigni (2013), “ý kiến từ mạng xã hội và các website không chỉ ảnh hưởng đến quyết định mua sắm của khách hàng mà còn hướng dẫn các nhà kinh doanh trong việc đưa ra các quyết định chiến lược”. Nhà quản trị có thể thấu hiểu phản ứng thị trường cung cấp hiện tại của các công ty và

sau đó đưa thông tin này vào phát triển sản phẩm và quy trình kiểm soát chất lượng dịch vụ (Dellarocas, 2003). Ý kiến bình luận của khách hàng trực tuyến (Online Customer Reviews) có thể được thu thập và khai thác bởi các công ty và các nhà nghiên cứu cho phép theo dõi SHL của khách hàng về dịch vụ. Với sự tăng trưởng và đa dạng của dữ liệu lớn, những ý kiến bình luận trực tuyến cần được thu thập và khai thác tự động, cho phép các nhà quản lý: (1) Xác định điểm mạnh, điểm yếu tương đối của sản phẩm hoặc dịch vụ; (2) phân tích các mối đe dọa từ đối thủ cạnh tranh; (3) quản lý rủi ro. Khách hàng cũng cần khai thác ý kiến đánh giá trực tuyến để đưa ra quyết định về việc mua sản phẩm, dịch vụ (Lee và cộng sự, 2011). Một ứng dụng khác của khai thác ý kiến là trong lĩnh vực chính quyền thông minh (Government Intelligence) (Pang & Lee, 2008), ở đó cho phép chính phủ có thể theo dõi ý kiến của người dân về các chính sách công và các ý kiến công chúng rất quan trọng trong các quyết định của chính phủ. Tương tự, chính phủ có thể dự đoán những gì công chúng đang nghĩ đến về các đề xuất chính sách và luật pháp (Stylios và cộng sự, 2010). Khai thác ý kiến cũng có ứng dụng tiềm năng trong đánh giá tin tức. Nó phân tích nội dung chứa cảm xúc trong tin tức và làm nổi bật những tin tức tương tự hoặc dư thừa (Wanner và cộng sự, 2009) hoặc xác định xu hướng thú vị và đặc thù trong các mục tin tức. Mặt khác, người xem có thể tìm thấy những bài viết phổ biến nhất, được thảo luận nhiều nhất, hoặc được trích dẫn nhiều nhất. Phân tích trích dẫn cũng là một ứng dụng khác mà việc khai thác ý kiến có thể hữu ích. Nó giúp xác định xem tác giả có trích dẫn một tác phẩm như là minh chứng, có quên trích dẫn hay không, hoặc theo dõi sự nổi danh trong văn học (Pang & Lee, 2008). Khai thác ý kiến cũng có thể được ứng dụng cho các hệ tư vấn thông tin (Recommender Systems) để đề xuất các danh mục được phản hồi tích cực (Pang & Lee, 2008). Người dùng có thể thấy được các nhận xét tích cực và tiêu cực về các phát hành gần đây, ví dụ các chương trình truyền hình phổ biến và phim điện ảnh bằng cách sử dụng các công cụ khai thác ý kiến hướng dẫn người dùng về các chương trình nào nên xem (Binali và cộng sự, 2009). Tương tự, nó cũng cải thiện hệ thống giáo dục dựa trên phân tích SHL của người học về các khóa học, cơ sở đào tạo và gia sư (Binali và cộng sự, 2009).

2.3. Phương pháp khai thác ý kiến

Khai thác ý kiến (còn được gọi là phân tích cảm xúc) là một kỹ thuật xử lý ngôn ngữ tự nhiên. Các nghiên cứu thường dựa vào hai cách tiếp cận phổ biến: Phương pháp máy học và phương pháp từ vựng (Medhat và cộng sự, 2014; Dhokrat và cộng sự, 2015; Yadav, 2015; Sun và cộng sự, 2017). Phương pháp tiếp cận máy học đóng vai trò quan trọng trong khai thác ý kiến. Đối với khai thác ý kiến ở mức độ tài liệu và câu có thể được xem như một vấn đề phân loại để xác định một phát biểu là “tích cực” hoặc “tiêu cực”. Bộ phân loại (Classifiers) được huấn luyện để xác định cực của văn bản tiếp theo. Naive Bayes (NB), Maximum Entropy (ME), và Support Vector Machine (SVM) là các mô hình máy học có giám sát (Supervised Learning) được dùng phổ biến (Shawe-Taylor & Sun, 2011). Tuy nhiên, với những tình huống đa lĩnh vực (Cross-Domain) hoặc đa ngôn ngữ (Cross-Lingual), kết quả huấn luyện không thể đem áp dụng cho lĩnh vực khác hoặc ngôn ngữ khác. Latent Dirichlet Allocation (LDA) (Blei và cộng sự, 2003) là một mô hình học bán giám sát (Semi-Supervised Learning) được đề xuất sử dụng để giải phóng sự phụ thuộc vào ngôn ngữ. Phương pháp tiếp cận từ vựng xác định số điểm cảm xúc (Sentiment Score) của văn bản dựa trên tập từ vựng cảm xúc một cách không giám sát (Unsupervised). Tập từ vựng là một từ điển các từ và cụm từ cảm xúc cùng với tính phân cực của chúng. Đối với mỗi tài liệu hoặc câu, phân cực cảm xúc tương ứng được xác định bởi điểm số cảm xúc được tính toán bởi các từ hoặc cụm từ đã xảy ra và cực cảm xúc của chúng. Đối

với phương pháp từ vựng yêu cầu ít tài nguyên hơn. Ngoài ra, từ vựng cảm xúc có thể được kết hợp với các máy học để cải tiến chất lượng phân loại (Sun và cộng sự, 2017).



Hình 2. Kỹ thuật phân loại ý kiến

Nguồn: Sun và cộng sự (2017).

2.4. Các nghiên cứu liên quan

Tiếp cận dữ liệu lớn và các phương pháp máy học trong việc đo lường SHL khách hàng đang được quan tâm bởi các nhà nghiên cứu và cả các nhà quản trị doanh nghiệp. Farhadloo và cộng sự (2016) đã mô hình hóa SHL của khách hàng từ dữ liệu phi cấu trúc bằng phương pháp Bayes. Nghiên cứu chủ yếu tập trung vào các khía cạnh liên quan đến sản phẩm và đo lường mức độ hài lòng tổng thể của khách hàng từ văn bản được viết ở dạng tự do. Farhadloo và cộng sự (2016) đã chuyển đổi dữ liệu phi cấu trúc thành dạng dữ liệu bán cấu trúc bằng cách áp dụng phân tích cảm xúc dựa trên các khía cạnh. Dữ liệu được thu thập từ TripAdvisor và các phương pháp tiếp cận thực nghiệm được áp dụng trên dữ liệu đã cho kết quả hiệu quả với tỷ lệ chính xác 88,3%. Ting và cộng sự (2017) đã nghiên cứu trải nghiệm của khách hàng được đăng trên Yelp tác động đến ngành khách sạn bằng cách kết hợp các phương pháp lập trình và khai thác dữ liệu để phân tích bình luận của khách hàng Yelp. Mục đích là xác định lại trải nghiệm của khách hàng và mối liên hệ của nó đối với mức độ hài lòng. Kết quả nghiên cứu của Ting và cộng sự (2017) cho thấy có nhiều yếu tố quan trọng liên quan đến các đánh giá của khách hàng mang các trọng số khác nhau và tìm thấy thành phần ngữ nghĩa có ý nghĩa bên trong các bình luận của khách hàng. Hơn nữa, cách tiếp cận được lựa chọn có thể sử dụng phân tích dữ liệu lớn để tìm ra các khía cạnh khác nhau chưa được nghiên cứu trong ngành khách sạn. Zhao và cộng sự (2019) đã dự đoán mức độ hài lòng chung của khách hàng bằng cách sử dụng dữ liệu thu thập từ các bài bình luận trực tuyến về các khách sạn. Bằng cách sử dụng dữ liệu trực tuyến, Zhao và cộng sự (2019) đã nghiên cứu mức độ hài lòng tổng thể của khách hàng bằng cách sử dụng các thuộc tính kỹ thuật của các bình luận và kết luận rằng mức độ chú quan cao hơn và bình luận dài hơn dẫn đến sự không hài lòng của khách hàng. Hơn nữa, nghiên cứu cũng kết luận rằng sự tham gia bình luận của khách hàng trong cộng đồng có ảnh hưởng đến SHL chung của khách hàng. Sezgen và cộng sự (2019) đã tiến hành nghiên cứu để đo lường SHL của khách hàng bằng cách sử dụng các phương pháp

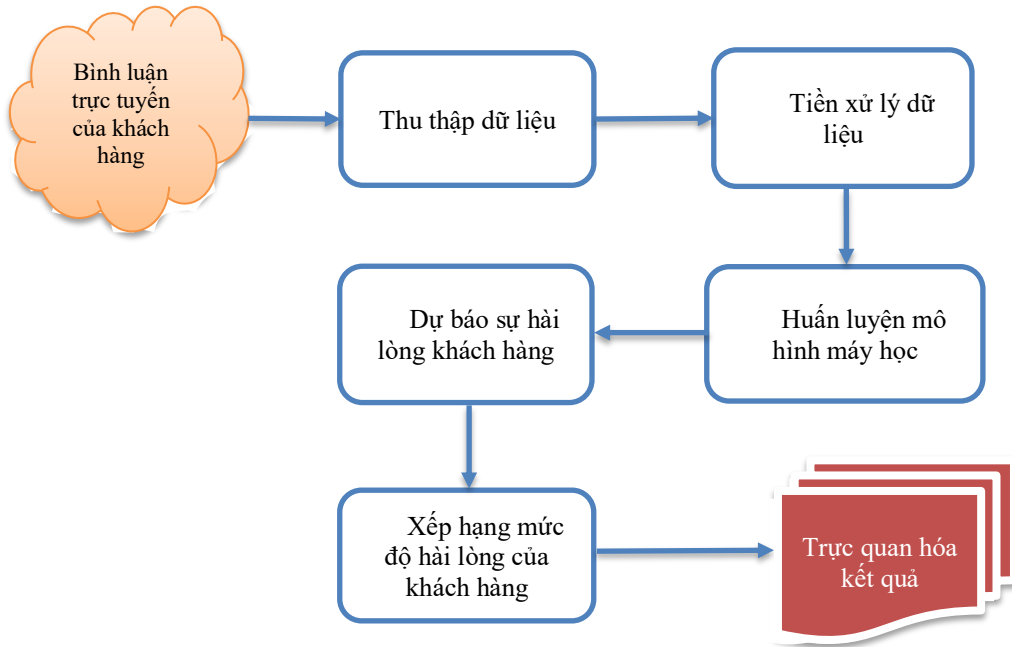
khai thác văn bản. Sezgen và cộng sự (2019) đã điều tra những nguyên nhân chính dẫn đến SHL và không hài lòng của khách hàng đối với hai nhà cung cấp dịch vụ khác biệt: Dịch vụ hạn chế và dịch vụ đầy đủ. Thu thập 500 đánh giá của hành khách trong đó 50 hãng hàng không khác nhau được đánh giá. Kết quả nghiên cứu của Sezgen và cộng sự (2019) cho rằng sự khác biệt cơ bản về mức độ hài lòng và không hài lòng của khách hàng là dựa trên mẫu và hạng cabin. Một yếu tố khác tạo nên SHL của khách hàng là sự thân thiện của tiếp viên trong khi thiếu yếu tố vệ sinh là yếu tố chính dẫn đến sự không hài lòng của khách hàng. Jung và Suh (2019) đã nghiên cứu ý kiến của nhân viên bằng cách triển khai các phương pháp khai thác văn bản để xác định và phân tích các yếu tố hài lòng trong công việc từ các bình luận trực tuyến của nhân viên. Jung và Suh (2019) đã thu thập 35.063 đánh giá trực tuyến của nhân viên từ trang Jobplanet để xác định mức độ hài lòng trong công việc bằng cách sử dụng phương pháp phân bổ Dirichlet tiềm ẩn (Latent Dirichlet Allocation – LDA) và tiến hành một loạt các phân tích dựa trên các yếu tố đang được nghiên cứu. Jung và Suh (2019) đã đo lường phân tích cảm xúc của nhân viên cũng như tầm quan trọng của từng yếu tố hài lòng trong công việc và kết luận rằng quản lý cấp cao là nhân tố quan trọng nhất đóng vai trò chi phối đối với SHL của nhân viên. Thu (2020) đã có nghiên cứu tập trung vào việc đo lường SHL của khách từ các đánh giá trực tuyến của khách sạn Việt Nam. Các khách đã có kinh nghiệm với các dịch vụ của khách sạn để lại đánh giá trên TripAdvisor, Booking, Hotel.com. Do có nhiều dữ liệu lớn, các kỹ thuật xử lý ngôn ngữ được áp dụng cho dữ liệu lớn từ TripAdvisor. Từ vựng cảm xúc cũng phát triển cho lĩnh vực khách sạn này để đạt được nhiều hơn độ chính xác của phép đo. Từ điển cảm xúc này giúp xác định và phân loại các bài đánh giá thành tích cực hoặc tiêu cực. Thu (2020) cũng đưa ra 6 tiêu chí được sử dụng để đo lường SHL của khách, kết quả thử nghiệm với dữ liệu thực tế cho thấy SHL của khách đối với các tiêu chí này là khác nhau, trong đó tiêu chí hài lòng nhất là nhân viên, tiêu chí không hài lòng khách là bữa ăn. Điều này có thể là giải thích là do thị hiếu của mỗi quốc gia khác nhau. Phung và cộng sự (2021) đã có nghiên cứu ứng dụng phương pháp máy học trong phân loại ý kiến khách hàng trực tuyến. Kết quả cho thấy các phương pháp Logistic Regression (LR), Support Vector Machines (SVM), và Neural Network (NN) có hiệu suất tốt nhất đối với khai thác ý kiến bằng ngôn ngữ tiếng Việt. Tuy nhiên, Phung và cộng sự (2021) chỉ tập trung vào việc phân loại ý kiến theo hai mức: Tích cực hoặc tiêu cực; chưa có đo lường SHL chung của khách hàng trên tổng thể tập dữ liệu. Oh và cộng sự (2022) đã ứng dụng mô hình học sâu dựa trên lý thuyết xác nhận kỳ vọng để dự đoán SHL của khách hàng đối với dịch vụ khách sạn, kết quả cho thấy mô hình được đề xuất đạt độ chính xác 83,54%. Almuqren và Cristea (2022) đã có nghiên cứu sử dụng khai thác Twitter để đo lường SHL của khách hàng đối với các dịch vụ viễn thông của các công ty ở Ả Rập Saudi, bên cạnh đó, phân tích bằng câu hỏi để đưa ra các khuyến nghị cho những công ty này, kết quả cho thấy mức độ hài lòng đối với từng dịch vụ nằm trong khoảng từ 31,50% đến 49,25%.

Nhìn chung, những nghiên cứu trên tập trung vào hai hướng chính: (1) Phân loại ý kiến của khách hàng bằng các mô hình máy học; và (2) kết hợp phân loại ý kiến và các mô hình thống kê để kiểm định mối quan hệ của các yếu tố, các tiêu chí đối với SHL khách hàng. Gần như chưa có nhiều nghiên cứu kết hợp phân loại ý kiến bằng máy học và các phương pháp tổng hợp, xếp hạng mức độ hài lòng của khách hàng, đặc biệt là phân tích dựa trên những bình luận của khách hàng bằng ngôn ngữ tiếng Việt.

3. Mô hình ứng dụng máy học trong đo lường sự hài lòng khách hàng

3.1. Mô hình nghiên cứu đề xuất

Dựa trên cơ sở lý thuyết và các khoảng trống nghiên cứu, mô hình nghiên cứu được đề xuất như Hình 3.



Hình 3. Mô hình nghiên cứu tổng quát

3.2. Thiết kế thực nghiệm nghiên cứu

- *Thu thập và tiền xử lý dữ liệu*

Nghiên cứu này tiến hành thu thập dữ liệu bằng chương trình tự động, dữ liệu được lấy từ trang web Agoda.com (là một nền tảng đặt phòng có lượng khách hàng và bình luận nhiều nhất hiện tại ở Việt Nam, đặc biệt là các bình luận bằng tiếng Việt, phù hợp với đối tượng và phạm vi nghiên cứu). Tiếp đến, nghiên cứu thực hiện tiền xử lý dữ liệu, loại bỏ những dữ liệu khuyết, những bình luận không chứa đựng thông tin cần thiết để tiến hành bước xử lý tiếp theo.

- *Huấn luyện mô hình máy học*

Mục đích để xác định một ý kiến của khách hàng về dịch vụ là “hài lòng” hay “không hài lòng”. Nghiên cứu này dùng các kỹ thuật phân loại được cho là tốt nhất từ những nghiên cứu trước để tìm ra mô hình tốt nhất đối với tập dữ liệu đã được gán nhãn, sau đó tiến hành dự báo phân loại cho các dữ liệu chưa được gán nhãn hoặc các dữ liệu bình luận mới. Nghiên cứu thực hiện huấn luyện bằng 6 kỹ thuật, bao gồm: Neural Network (NN), Support Vector Machines (SVM), Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), và Random Forest (RF).

- *Dự báo sự hài lòng của khách hàng*

Tiếp theo, từ kết quả huấn luyện mô hình bên trên, việc lựa chọn mô hình tốt nhất thường dựa trên các chỉ số: Độ chính xác (Accuracy), Độ hội tụ (Precision), Độ phủ (Recall), và Giá trị trung bình điều hòa (F1).

Trong đó:

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Ở bước này, nghiên cứu lựa chọn một mô hình tốt nhất và tiến hành dự báo phân loại SHL cho toàn bộ tập dữ liệu chưa được phân loại.

- *Xếp hạng mức độ hài lòng của khách hàng*

Sau khi phân loại hài lòng bằng kỹ thuật máy học được xem là tốt nhất, kết quả có một tập dữ liệu thể hiện quan điểm “hài lòng” hay “không hài lòng” về sản phẩm, dịch vụ. Việc xếp hạng mức độ hài lòng dựa vào công thức được tham chiếu từ nghiên cứu của Bashir và cộng sự (2016) như sau:

$$H_i = \frac{TP_i - TN_i}{TR_i} \quad (5)$$

Trong đó:

H_i : Điểm mức độ SHL của khách hàng đối của dịch vụ thứ i , với $i = 1, 2, \dots, m$ (trong đó, m là tổng số dịch vụ);

TP_i : Số bình luận được đánh giá là “hài lòng” của dịch vụ thứ i ;

TN_i : Số bình luận được đánh giá là “không hài lòng” của dịch vụ thứ i ;

TR_i : Tổng số bình luận nhận được của dịch vụ thứ i .

Ta có thể chuẩn hóa thang đo theo công thức sau:

$$R'_i = a + \frac{(H_i - \min\{H_1, H_2, \dots, H_n\}) \times (b - a)}{\max\{H_1, H_2, \dots, H_n\} - \min\{H_1, H_2, \dots, H_n\}} \quad (6)$$

Trong đó,

R'_i : Điểm mức độ hài lòng của khách hàng đối với dịch vụ thứ i đã được chuẩn hóa;

a và b : Cận dưới và cận trên của thang đo.

4. Kết quả thực nghiệm nghiên cứu

4.1. Kết quả thu thập và tiền xử lý dữ liệu

Dữ liệu thu được là 40.000 bình luận về 1.190 khách sạn của 58 tỉnh thành trên cả nước (bao gồm tiếng Việt và tiếng Anh). Các bình luận bằng tiếng Anh bị loại bỏ, dữ liệu còn lại để thực nghiệm là 16.570 bình luận.

4.2. Kết quả huấn luyện mô hình máy học

Để tiến hành huấn luyện bằng các phương pháp máy học, cần phải có tập dữ liệu đã được gán nhãn (đã phân loại) đủ lớn. Thông thường, với các nghiên cứu áp dụng kỹ thuật máy học, tập dữ liệu này sẽ được gán nhãn thủ công. Nghiên cứu lấy mẫu ngẫu nhiên 50 quan sát từ dữ liệu thu thập được, sau đó xem xét nội dung các ý kiến bình luận và tiến hành thảo luận nhóm 10 du khách để xác định ngưỡng đánh giá. Kết quả phát hiện rằng những bình luận có điểm đánh giá (dựa vào cột Rating trong tập dữ liệu) nhỏ hơn 7,0 điểm mang ý nghĩa tiêu cực (không hài lòng) và ngược lại, các bình luận có điểm đánh giá lớn hơn 7,0 điểm mang ý nghĩa tích cực (hài lòng). Từ việc xác định ngưỡng điểm này, nghiên cứu tiến hành gán nhãn tự động bằng phần mềm xử lý dữ liệu Python. Từ đó, tập dữ liệu huấn luyện được xây dựng gồm có 16.570 bình luận, trong đó: 6.860 bình luận là “hài lòng” (được gán nhãn 0) và 9.710 bình luận là “không hài lòng” (được gán nhãn 1). Nghiên cứu này thực hiện huấn luyện bằng 6 kỹ thuật, bao gồm: Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), Support Vector Machines (SVM), Logistic regression (LR), Random Forest (RF).

Bảng 1.

Kết quả huấn luyện bằng các kỹ thuật phân loại

STT	Tên phương pháp	Accuracy	Precision	Recall	F1	Thời gian (giây)
1	Neural Network (NN)	0,79	0,79	0,79	0,77	133,50
2	Decision Tree (DT)	0,71	0,71	0,71	0,71	46,37
3	Naïve Bayes (NB)	0,48	0,66	0,49	0,48	2,87
4	Support Vector Machines (SVM)	0,78	0,78	0,78	0,78	0,86
5	Logistic Regression (LR)	0,79	0,79	0,79	0,78	0,74
6	Random Forest (RF)	0,68	0,47	0,68	0,55	0,88

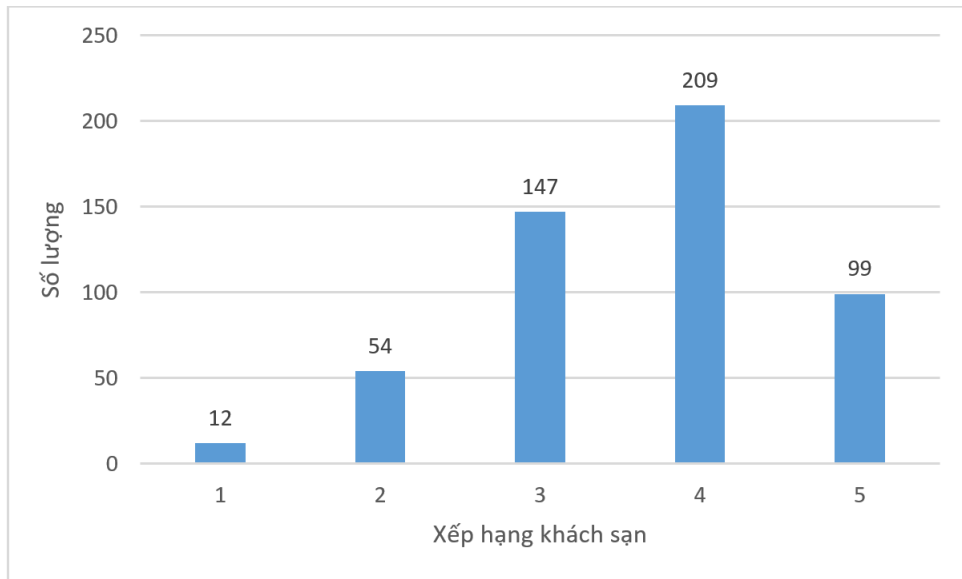
4.3. Kết quả dự báo sự hài lòng của khách hàng

Kết quả huấn luyện cho thấy các mô hình LR, SVM có độ chính xác khá cao ($F1 = 0,78$). Nghĩa là các mô hình này tương đối phù hợp với tập dữ liệu huấn luyện. Tuy nhiên, nếu xét thêm chỉ số Accuracy và yếu tố thời gian thì mô hình LR là tốt nhất. Kết quả này giúp xác định phương pháp và công cụ phân loại ý kiến phù hợp. Đây được xem là bước quan trọng nhất của quy trình khai thác ý kiến, làm cơ sở cho việc ứng dụng khai thác ý kiến trong nhiều lĩnh vực. Do vậy, nghiên cứu chọn phương pháp LR và tiến hành dự báo phân loại SHL cho toàn bộ tập dữ liệu chưa được phân loại.

4.4. Kết quả xếp hạng sự hài lòng của khách hàng

Sau khi thực hiện phân loại SHL bằng phương pháp máy học được xem là tốt nhất trong bước huấn luyện mô hình (nghiên cứu này đã lựa chọn phương pháp LR), kết quả có một tập dữ liệu các bình luận thể hiện SHL của khách hàng về các khách sạn. Việc xếp hạng khách sạn theo SHL của

khách hàng được thực hiện như trong các công thức (5) và (6). Nghiên cứu này sử dụng thang điểm từ 1 đến 5 và thực hiện làm tròn điểm số. Kết quả cho thấy số lượng khách sạn được đánh giá hài lòng ở mức độ 4 (4 điểm) là cao nhất (209 khách sạn) và số lượng khách sạn được khách hàng đánh giá hài lòng ở mức độ 1 (1 điểm) là thấp nhất (12 khách sạn).



Hình 4. Kết quả xếp hạng mức độ hài lòng của khách hàng

5. Kết luận

Trong nghiên cứu này, chúng tôi đã đề xuất ứng dụng phương pháp máy học trong đo lường SHL của khách hàng dựa trên các bình luận trực tuyến. Các kết quả thực nghiệm đạt được là một bản tham chiếu giúp cho các doanh nghiệp, các nhà quản trị có thể áp dụng phương pháp này để đo lường SHL của khách hàng trong bối cảnh sự bùng nổ của dữ liệu lớn, đó là các ý kiến của khách hàng trên các nền tảng trực tuyến.

Nghiên cứu này cũng có nhiều điểm thiếu sót, có thể làm khoảng trống tri thức cho những nghiên cứu tiếp theo. Thứ nhất, dữ liệu nghiên cứu là các ý kiến bình luận trên trang Agoda.com, có thể mở rộng thu thập thêm các bình luận về các sản phẩm trên các trang web bán hàng trực tuyến hoặc các nền tảng mạng xã hội. Thứ hai, nghiên cứu này phân loại theo hai mức: Hài lòng (Positive) và không hài lòng (Negative). Các nghiên cứu tiếp theo có thể phân loại theo nhiều mức hơn. Bên cạnh đó, nghiên cứu này chỉ áp dụng các kỹ thuật máy học cơ bản, nếu có điều kiện thử nghiệm với kỹ thuật học sâu (Deep Learning) có thể sẽ thu được kết quả tốt hơn.

Chú thích

Bài báo này được trích một phần trong kết quả nghiên cứu từ luận án tiến sĩ của nghiên cứu sinh Thái Kim Phụng.

Tài liệu tham khảo

- Almuqren, L., & Cristea, A. I. (2022). Predicting STC customers' satisfaction using Twitter. *IEEE Transactions on Computational Social Systems*, 10(1), 204–210.
- Bashir, S., Afzal, W., & Baig, A. R. (2016). Opinion-based entity ranking using learning to rank. *Applied Soft Computing*, 38, 151–163.
- Becser, N., & Zoltay-Paprika, Z. (2003). *A Decision Support Model for Improving Service Quality, SQI-DSS: A New Approach* (Working Paper). Budapest University of Economics and Public Administration, Department of Business Economics.
- Binali, H., Potdar, V., & Wu, C. (2009). *A state of the art opinion mining and its application domains*. In Industrial Technology, 2009. ICIT 2009. IEEE International Conference on. IEEE.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407–1424.
- Dhokrat, A., Khillare, S., & Mahender, C. N. (2015). Review on techniques and tools used for opinion mining. *International Journal of Computer Applications Technology and Research*, 4(6), 419–424.
- Farhadloo, M., Patterson, R. A., & Rolland, E. (2016). Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems*, 90, 1–11.
- Hussain, Y. (2019). *Predicting customer satisfaction with product reviews: A comparative study of some machine learning approaches* (Master's thesis). Faculty of Information Technology and Communication Sciences, Tampere University.
- Jung, Y., & Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decision Support Systems*, 123, 113074.
- Lee, J., Park, D.-H., & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: An advertising perspective. *Internet Research*, 21(2), 187–206.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185–200.
- Oh, S., Ji, H., Kim, J., Park, E., & del Pobil, A. P. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. *Information Technology & Tourism*, 24(1), 109–126.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

- Phung, T. K., Te, N. A., & Ha, T. T. T. (2021). *A machine learning approach for opinion mining online customer reviews*. In 2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), (pp. 243–246).
- Piccoli, G., & Pigni, F. (2013). Harvesting external data: The potential of digital data streams. *MIS Quarterly Executive*, 12(1), 143–154.
- Sezgen, E., Mason, K. J., & Mayer, R. (2019). Voice of airline passenger: A text mining approach to understand customer satisfaction. *Journal of Air Transport Management*, 77, 65–74.
- Shawe-Taylor, J., & Sun, S. (2011). A review of optimization methodologies in support vector machines. *Neurocomputing*, 74(17), 3609–3618.
- Stylios, G., Christodoulakis, D., Besharat, J., Vonitsanou, M., Kotrotsos, I., Koumpouri, A., & Stamou, S. (2010). Public opinion mining for governmental decisions. *Electronic Journal of e-Government*, 8(2), 203–214.
- Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 36, 10–25.
- Thu, H. N. T. (2020). Measuring guest satisfaction from online reviews: Evidence in Vietnam. *Cogent Social Sciences*, 6(1), 1801117.
- Ting, P. J. L., Chen, S. L., Chen, H., & Fang, W. C. (2017). Using big data and text analytics to understand how customer experiences posted on yelp.com impact the hospitality industry. *Contemporary Management Research*, 13(2). doi: 10.7903/cmr.17730
- Wang, Y. (2016). *More important than ever: Measuring tourist satisfaction*. Griffith Institute for Tourism Research Report Series Report No. 10. Griffith University
- Wanner, F., Rohrdantz, C., Mansmann, F., Oelke, D., & Keim, D. A. (2009). *Visual sentiment analysis of RSS news feeds featuring the us presidential election in 2008*. Retrieved from <https://kops.uni-konstanz.de/server/api/core/bitstreams/29e30b9f-8714-45d5-9f7e-46eb9811381f/content>
- Yadav, S. K. (2015). Sentiment analysis and classification: A survey. *International Journal of Advance Research in Computer Science and Management Studies*, 3(3), 113–121.
- Yussupova, N., Kovács, G., Boyko, M., & Bogdanova, D. (2016). Models and methods for quality management based on artificial intelligence applications. *Acta Polytechnica Hungarica*, 13(3), 45–60.
- Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111–121.