

HỆ THỐNG HỖ TRỢ ĐÁNH GIÁ VÀ KHUYẾN NGHỊ DỊCH VỤ DU LỊCH DỰA TRÊN KHAI THÁC Ý KIẾN KHÁCH HÀNG TRỰC TUYẾN

THÁI KIM PHỤNG¹, NGUYỄN AN TÊ², TRẦN THỊ THU HÀ³

^{1,2} Khoa Công nghệ thông tin kinh doanh, Trường Đại học Kinh tế TP. Hồ Chí Minh,

³ Viện Công nghệ thông tin và Kinh tế số, Trường Đại học Kinh tế Quốc dân;

phungthk@ueh.edu.vn, tena@ueh.edu.vn, thuha.tim@gmail.com

Tóm tắt: Nghiên cứu này được tiến hành nhằm đề xuất mô hình kiến trúc hệ thống cùng với các giải pháp hỗ trợ đánh giá và khuyến nghị dịch vụ du lịch dựa trên phương pháp khai thác ý kiến. Nghiên cứu này tiếp cận theo phương pháp khai phá dữ liệu theo chuẩn công nghiệp (CRISP-DM). Dữ liệu thực nghiệm nghiên cứu là những bình luận của du khách về các khách sạn tại các tỉnh và thành phố lớn tại Việt Nam, được thu thập tự động trên trang web Agoda. Trên cơ sở các kết quả thực nghiệm, nghiên cứu đưa ra một số khuyến nghị để có thể triển khai hệ thống này trong thực tiễn ngành du lịch. Nghiên cứu này có giá trị tham chiếu cho các nhà nghiên cứu không chỉ trong lĩnh vực du lịch mà còn trong các lĩnh vực kinh doanh và quản lý.

Từ khóa: Khai thác ý kiến, Đánh giá và khuyến nghị dịch vụ du lịch, Hệ thống hỗ trợ ra quyết định

A SUPPORT SYSTEM FOR TOURISM SERVICES ASSESSMENT AND RECOMMENDATION BASED ON OPINION MINING ONLINE CUSTOMER REVIEWS

Abstract: This research was conducted to propose the a architecture model with solutions to support tourism services assessment and recommendations based on the opinion mining methods. This research approach is based on an industry standard data mining method (CRISP-DM). Data for the research are reviews of tourist about hotels in major provinces and cities in Vietnam, which automatically collected on Agoda. Based on empirical results, the research has some recommendations to be able to implement this system in the tourism industry. This research is valuable as a reference for researchers not only in the field of tourism but also in the fields of business and management.

Keywords: Opinion mining, Tourism service assessment and recommendation, Decision support system

1 GIỚI THIỆU

Ngày nay, những tiến bộ của công nghệ đã làm thay đổi cách thức truyền thông giúp cho khách hàng dễ dàng truy cập thông tin và trao đổi ý kiến về sản phẩm và dịch vụ trên một quy mô lớn trong thời gian thực. Sự ra đời của mạng xã hội và các website đánh giá trực tuyến cho phép khách hàng có cơ hội đưa ra ý kiến của mình thông qua các bài bình luận về sản phẩm, dịch vụ [35]. Với sự bùng nổ của dữ liệu lớn (big data) và công nghệ Internet kết nối vạn vật - IoTs (Internet of Things), các ý kiến đánh giá trực tuyến của khách hàng cần được thu thập, khai thác và tổng hợp một cách tự động bằng các hệ thống máy tính, cho phép các nhà kinh doanh có thể dễ dàng theo dõi hành vi mua sắm, phát hiện sở thích và đánh giá sự hài lòng của khách hàng về chất lượng sản phẩm, dịch vụ. Đồng thời, khách hàng cũng cần thông tin tổng hợp ý kiến đánh giá của cộng đồng để có những quyết định mua sắm của mình. Chính vì thế, khai thác ý kiến (opinion mining) một cách tự động vì thế đã trở thành tiêu điểm của rất nhiều nghiên cứu trong các lĩnh vực: nghiên cứu thị trường (market research), kinh doanh điện tử (e-business), thăm dò chính sách (political polls) [34]. Du lịch là một ngành công nghiệp phát triển năng động và đóng vai trò quan trọng ở các quốc gia và khu vực trên thế giới [26]. Theo báo cáo nghiên cứu hàng năm của Hội đồng du lịch và lữ hành thế giới (World Tourism and Travel Council – WTTC), ngành du lịch là một trong những ngành đóng góp chính cho GDP thế giới và đã tăng trưởng trong sáu năm liên tiếp cho đến năm 2015 và tăng lên 9.8% GDP thế giới (7.2 nghìn tỷ USD) [19]. Sự phát triển của Internet đã cung cấp nhiều phương thức mới để kinh doanh dịch vụ du lịch và quảng bá các sản phẩm điểm đến cho du khách. Tại Việt Nam, xu hướng ứng dụng Công nghệ thông tin trong ngành du lịch đã bắt đầu phát triển mạnh mẽ với sự xuất hiện các công ty kinh doanh dịch vụ du lịch trực tuyến, các ứng dụng đặt tour trên nền tảng di động và nhiều website, diễn đàn, cổng thông

tin xã hội, v.v... nhằm trao đổi, chia sẻ các điểm đến du lịch và cũng đã thu hút một lượng lớn du khách trong và ngoài nước tham gia (Chudu24.vn, TripAdvisor.com.vn, Agoda.com, Booking.com,...). Tuy vậy, hiện nay việc đánh giá chất lượng dịch vụ du lịch vẫn được thực hiện bằng những cách thức truyền thống. Những phương pháp này không thể giám sát sự hài lòng của khách hàng một cách liên tục, và không có khả năng theo dõi xu hướng chất lượng dịch vụ trong dài hạn [30]. Khi chuẩn bị các quyết định liên quan đến việc đánh giá chất lượng dịch vụ, dữ liệu cần phải được thể được thu thập, phân tích và đề xuất các phương án để sẵn sàng cho người ra quyết định. Để đảm bảo tất cả các nhiệm vụ này, một hệ thống hỗ trợ ra quyết định (Decision support system - DSS) là cần thiết để xử lý và phân loại dữ liệu, sử dụng các mô hình phù hợp để phân tích, và chuẩn bị các kết quả để cung cấp cho nhà quản lý ra quyết định [28]. Chính vì vậy, mục tiêu của nghiên cứu này là đề xuất một mô hình kiến trúc hệ thống cùng với các giải pháp hỗ trợ đánh giá và khuyến nghị dịch vụ du lịch (trong đó tập trung vào dịch vụ khách sạn) dựa trên phương pháp khai thác ý kiến khách hàng trực tuyến.

2 CƠ SỞ LÝ THUYẾT

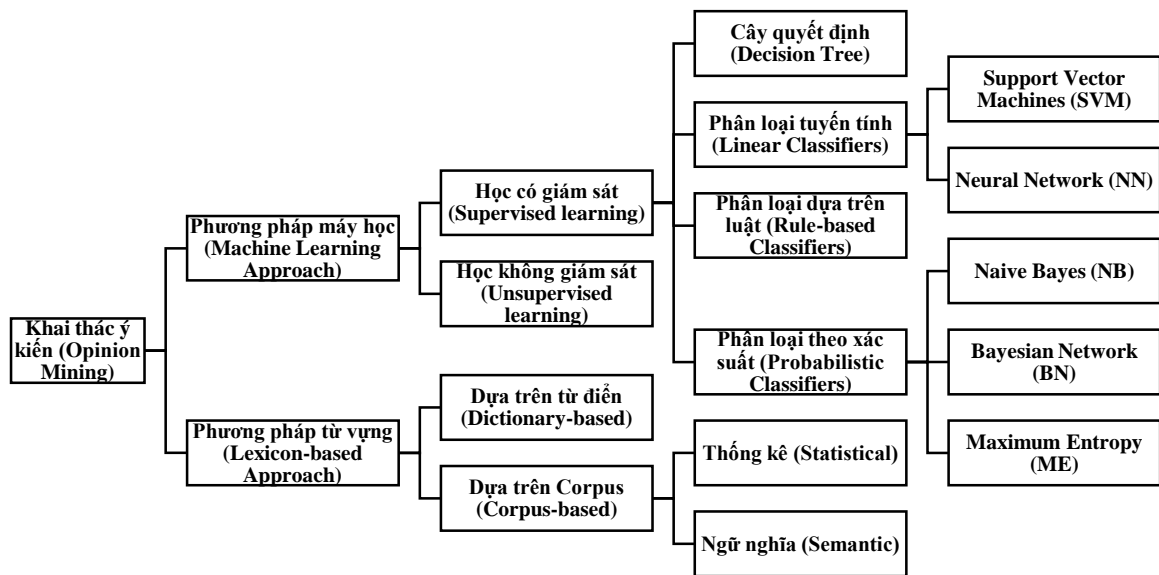
2.1 Các mô hình đánh giá chất lượng dịch vụ du lịch

Hiện nay, trong lĩnh vực du lịch, có rất nhiều mô hình đánh giá chất lượng dịch vụ được sử dụng rộng rãi như: Mô hình hiệu suất – tầm quan trọng (IPA) [21], Mô hình đánh giá chất lượng kỹ thuật, chức năng và hình ảnh của Gronroos [11], Mô hình khoảng cách chất lượng dịch vụ (giữa mức độ kỳ vọng - mức độ cảm nhận) SERVQUAL của Parasuraman và cộng sự vào năm 1988 [5, 6], Mô hình mức độ cảm nhận SERVPERF [22], là một biến thể của mô hình SERVQUAL, Mô hình RATER [41], là một sự phát triển của mô hình SERVQUAL. Tuy nhiên, việc đo lường mức độ cảm nhận của khách hàng để đánh giá và cải tiến chất lượng dịch vụ du lịch thường được thực hiện bằng cách thức truyền thống như: phỏng vấn trực tiếp, phỏng vấn qua điện thoại, thảo luận nhóm, khảo sát qua thư, khảo sát trực tuyến (ISO10004:2010). Theo [30], một trở ngại chung của những phương pháp này là rất nhiều công việc được tiến hành thủ công như: chuẩn bị các câu hỏi, tạo cơ sở dữ liệu cho người trả lời, gửi bảng câu hỏi, thu thập kết quả, phỏng vấn cá nhân, và chuẩn bị báo cáo,... Tất cả các thủ tục này khiến cho đợt khảo sát đánh giá chất lượng dịch vụ trở nên tốn kém. Sự cảm nhận của khách hàng được thể hiện dưới hình thức các chỉ số hài lòng trừu tượng làm cho khó hiểu, khó so sánh và giải thích kết quả. Các phương pháp phân tích dữ liệu thu thập được thông qua các thủ tục của ISO10004:2010 được khuyến cáo chỉ cho phép phát hiện các phụ thuộc tuyến tính. Việc phân tích dữ liệu bị giới hạn trong một khoảng thời gian nhất định và không đưa ra cái nhìn sâu sắc về xu hướng và sự vận động của sự hài lòng từ khách hàng. Điều này ảnh hưởng đến tốc độ ra quyết định quản lý. Bên cạnh đó, những phương pháp này không thể giám sát sự hài lòng của khách hàng một cách liên tục, và không có khả năng theo dõi xu hướng hài lòng của khách hàng trong dài hạn [30].

Ngày nay, những tiến bộ của lĩnh vực điện toán trên thế giới đã làm thay đổi cách thức truyền thông, đặc biệt là trong bối cảnh tác động của cuộc cách mạng Công nghiệp lần thứ 4, người sử dụng công nghệ dễ dàng truy cập thông tin và trao đổi ý kiến hài lòng về dịch vụ trên một quy mô lớn trong thời gian thực. Để khắc phục phương pháp đánh giá dịch vụ theo kiểu truyền thống, một hệ thống hỗ trợ ra quyết định là cần thiết để xử lý và phân loại dữ liệu, sử dụng các mô hình phù hợp để phân tích, và chuẩn bị các kết quả để cung cấp cho nhà quản lý ra quyết định [28].

2.2 Phương pháp khai thác ý kiến

Khai thác ý kiến là lĩnh vực nghiên cứu nhằm phân tích, đánh giá nhận định của con người về các đối tượng như: sản phẩm, dịch vụ, tổ chức, cá nhân, sự kiện, chủ đề và các thuộc tính của chúng [8, 9]. Một quy trình khai thác ý kiến thường gồm ba bước chính: (1) Thu thập ý kiến (Opinion Retrieval), (2) Phân loại ý kiến (Opinion Classification) và (3) Tổng hợp ý kiến (Opinion Summarization) [4, 7]. Trong đó, phân loại ý kiến được xem là bước quan trọng nhất nhằm mục đích phân lớp ý kiến theo các quan điểm: tích cực (positive), tiêu cực (negative) và trung lập (neutral). Phân loại ý kiến là một kỹ thuật khai thác dữ liệu dạng văn bản (Text Mining) trong lĩnh vực xử lý ngôn ngữ tự nhiên (Natural Language Processing). Có hai cách tiếp cận phổ biến trong phân loại ý kiến: dựa vào phương pháp máy học (Machine learning), dựa vào từ vựng (Lexicon based) [1, 34, 37, 45]. Ngoài ra, để gia tăng hiệu suất của việc phân loại ý kiến, các nghiên cứu đã dùng phương pháp lai kết hợp giữa hai phương pháp này.



Hình 1: Tổng hợp các phương pháp phân loại ý kiến trong nghiên cứu [45]

2.3 Các hệ thống hỗ trợ ra quyết định trong du lịch

Có nhiều nghiên cứu và nhiều cách tiếp cận khác nhau đối với việc hỗ trợ ra quyết định trong lĩnh vực du lịch [38]. Các nghiên cứu [14, 47] đã thảo luận về việc sử dụng hệ thống DSS trong việc hỗ trợ khách du lịch đưa ra quyết định lựa chọn điểm đến phù hợp với nhu cầu và ngân sách. Một số nghiên cứu khác tập trung vào việc sử dụng hệ thống DSS để hỗ trợ các bên liên quan trong ngành du lịch như hỗ trợ các nhà hoạch định du lịch [23], các nhà quản lý tổ chức quản lý điểm đến du lịch [32], các nhà dự báo nhu cầu du lịch [12], và cuối cùng là các nhà tiếp thị du lịch [18, 25], để đưa ra quyết định phù hợp. Một hướng tiếp cận khác là nghiên cứu các hệ thống hỗ trợ khách hàng trong việc ra quyết định, hay còn gọi hệ thống khuyến nghị (Recommender Systems). Hệ thống này được xem như là một đối thủ cạnh tranh với các đại lý du lịch vì nó cung cấp cho du khách những gợi ý để tạo thuận lợi cho quá trình ra quyết định của họ. Điển hình là các hệ thống Triplehop's TripMatcher, VacationCoach's Me-Print, DieToRecs của [17] và [3]. Tiếp cận theo hướng công nghệ, một số nghiên cứu đã sử dụng kỹ thuật nhà kho dữ liệu (Data Warehouse) như một công cụ để hỗ trợ quá trình ra quyết định trong doanh nghiệp thuộc lĩnh vực du lịch [2, 27, 38, 46]. Ngoài ra, một số nghiên cứu đã tích hợp công nghệ GIS (Geographic Information System) và Internet vào các hệ thống hỗ trợ ra quyết định trong du lịch nhằm tối ưu hóa quy trình lập kế hoạch và thu được lợi ích từ khả năng tiếp cận chính xác, trực quan hóa, xử lý dữ liệu và khả năng chia sẻ dựa trên vị trí địa lý [24, 36, 39].

2.4 Các hệ thống khai thác ý kiến trong du lịch

Lĩnh vực du lịch đã có những nghiên cứu khai thác ý kiến, có thể kể đến là nghiên cứu của [33] đề xuất một hệ thống tổng hợp ý kiến đánh giá cho một dịch vụ địa điểm (nhà hàng, khách sạn, điểm đến,...) bằng phương pháp lai giữa phương pháp máy học và phương pháp từ vựng. Cụ thể, nhóm tác giả tập trung vào các mô hình tóm tắt dựa trên khía cạnh, trong đó việc tóm lược được xây dựng bằng cách khai thác các khía cạnh liên quan đến dịch vụ, tổng hợp ý kiến cho mỗi khía cạnh và lựa chọn văn bản liên quan đến từng khía cạnh. Tiếp đến, [31] đã sử dụng các kỹ thuật máy học có giám sát để phân loại ý kiến của khách du lịch về các điểm đến du lịch ở Mỹ và Châu Âu. [41] đã sử dụng kỹ thuật Naïve Bayes (NB) và Self-Organizing Maps (SOM) trong việc phân loại và trực quan hóa ý kiến trên mạng xã hội Twitter của du khách về điểm đến du lịch tại Thái Lan. [49] cũng đã có nghiên cứu xây dựng hệ thống OpinionSeer bằng cách kết hợp phương pháp thống kê dựa trên ngôn ngữ và các kỹ thuật trực quan hóa dữ liệu để phân tích các phản hồi của du khách về khách sạn tại Hồng Kông. [44] đã xây dựng hệ thống BSAHOT, bằng cách thu thập dữ liệu từ các trang web và dựa trên phương pháp phân tích thống kê ký tự (n-grams) để phân loại ý kiến đánh giá của du khách về dịch vụ khách sạn. [15] đã trực quan hóa các kết quả phân tích ý kiến đánh giá trên Google Maps, cung cấp khả năng cho người dùng dễ dàng phát hiện các khách sạn và các khu vực tốt nhất để chọn ở. Tác giả đã sử dụng phương pháp máy học Naïve Bayes và bộ từ điển SentiWordNet chuẩn. [42] đã sử dụng kỹ thuật phân tích cảm xúc để khai thác 70.103 bài đánh giá được đăng trên các địa điểm trực

tuyến khác nhau từ năm 1999-2011 cho 86 khách sạn ở Washington. Khai thác ý kiến được thực hiện bằng phương pháp Naive Bayes kết hợp với mô hình kinh tế lượng đã giúp họ phân rã các đánh giá của người dùng thành 5 chiều để đo lường chất lượng dịch vụ khách sạn và kết quả phân tích cho thấy mức độ chính xác cao trong việc thu thập và đo lường chất lượng dịch vụ so với các nghiên cứu khai thác văn bản hiện trước đó. [16] đã phát triển khung kiến trúc chung cho công cụ khai thác ý kiến dựa trên khía cạnh, sau đó tạo mẫu thử nghiệm và phân tích ý kiến từ TripAdvisor trong bối cảnh ngành du lịch ở Los Lagos (Chi Lê). Nhóm tác giả sử dụng phương pháp khai thác ý kiến theo khía cạnh của [8] dựa trên việc thống kê theo các quy tắc ngôn ngữ tự nhiên và áp dụng cho lĩnh vực du lịch. Kết quả cho thấy phương pháp này có độ chính xác cao và hiệu quả hơn hẳn. Tiếp đến, [10] đã đề xuất một hệ thống tự động thu thập và tổng hợp ý kiến đánh giá khách sạn của du khách trên trang web du lịch TripAdvisor. Tác giả đã tận dụng bộ từ điển SentiWordNet kết hợp với phương pháp máy học không giám sát (unsupervised learning) trong việc phân loại ý kiến đánh giá. [26] đã xây dựng hệ thống phân loại ý kiến của du khách trên TripAdvisor, bằng phương pháp máy học kết hợp với các thuật toán logic mờ. Kết quả nghiên cứu cho thấy phương pháp khai thác ý kiến của nhóm tác giả có độ chính xác cao hơn các phương pháp truyền thống. [48] cũng đã có nghiên cứu khai thác ý kiến bằng phương pháp LDA (latent dirichlet allocation). Bộ dữ liệu được thu thập trên TripAdvisor bao gồm 266.544 đánh giá trực tuyến cho 25.670 khách sạn ở 16 quốc gia. Kết quả nghiên cứu đã xác định 19 khía cạnh dịch vụ khách sạn quan trọng. [20] đã đề xuất hệ thống tự động khai thác ý kiến của du khách về các khách sạn trên trang booking.com. Nhóm tác giả đã sử dụng phương pháp lai kết hợp giữa LDA và Naive Bayes. Kết quả nghiên cứu được đánh giá là rất đáng khích lệ và chỉ ra rằng hệ thống này khá nhanh, có thể mở rộng và hầu hết chính xác trong việc phân tích các đánh giá của khách hàng. [13] cũng đã có nghiên cứu tiến hành phân tích ý kiến đánh giá của người Nhật về các điểm tham quan tại Trung Quốc bằng cách sử dụng phương pháp thống kê và tập luật, kết hợp phương pháp trích chọn đặc trưng TF-IDF. Kết quả nghiên cứu cung cấp các hướng dẫn sử dụng thực tế trong việc phân tích ý kiến đánh giá du lịch của Nhật Bản và tạo điều kiện cải thiện dịch vụ điểm đến du lịch.

Trong nước cũng bắt đầu có những nghiên cứu sử dụng phương pháp khai thác ý kiến. Điển hình là nghiên cứu của [29], đã tiến hành thực nghiệm việc phân loại ý kiến bằng các kỹ thuật máy học: Naive Bayes, Support Vector Machines (SVM) và Maximum Entropy (ME), minh họa bằng bộ dữ liệu về các ý kiến bình luận khách sạn tại Việt Nam. Kế đến, [40] đã có nghiên cứu so sánh các thuật toán phân loại ý kiến bằng tiếng Việt, cũng thực nghiệm dựa trên bộ dữ liệu là các ý kiến về khách sạn tại Việt Nam. Kết quả của cả hai nghiên cứu này cho thấy phương pháp SVM là tốt nhất.

2.5 Nhận dạng các khoảng trống nghiên cứu

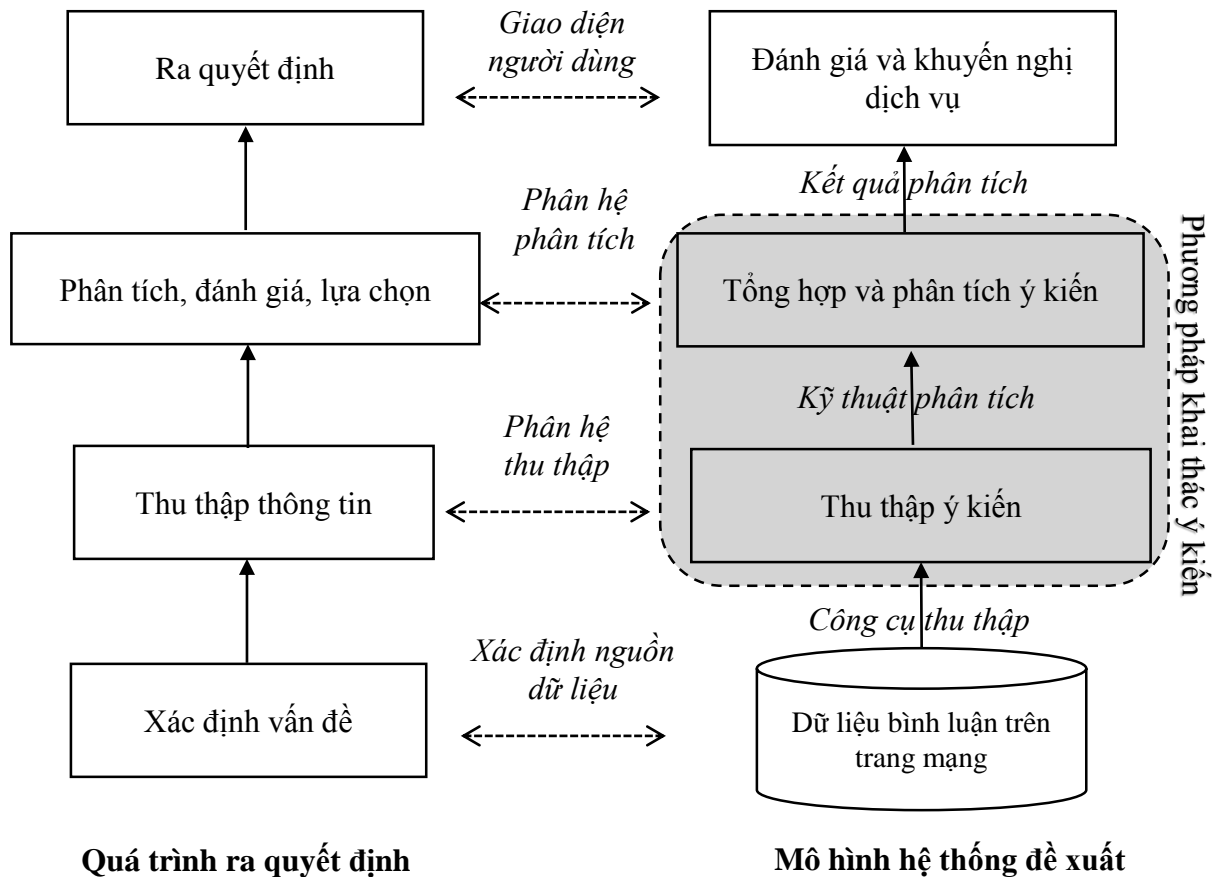
Từ kết quả lược khảo các công trình nghiên cứu, một số khoảng trống nghiên cứu được nhận dạng làm cơ sở để đề xuất mô hình nghiên cứu tổng quát:

- Hướng nghiên cứu về hệ thống hỗ trợ ra quyết định trong lĩnh vực du lịch đã được cộng đồng các nhà nghiên cứu quan tâm rất nhiều, đa dạng về hình thức tiếp cận, cả về công nghệ lẫn ứng dụng để hỗ trợ giải quyết nhiều vấn đề liên quan đến du lịch. Tuy nhiên, gần như chưa có nghiên cứu nào trong và ngoài nước đề cập đến ứng dụng trong hỗ trợ đánh giá dịch vụ du lịch. Bên cạnh đó, cũng có rất ít nghiên cứu hệ thống hỗ trợ ra quyết định tiếp cận dựa trên phương pháp khai thác ý kiến.
- Phần lớn các nghiên cứu khai thác ý kiến trong du lịch thường tiếp cận trên quan điểm kỹ thuật tập trung xây dựng hoặc áp dụng các thuật toán xử lý ngôn ngữ tự nhiên vào việc phân loại và tổng hợp ý kiến du khách, rất ít nghiên cứu tiếp cận theo hướng hành vi để hỗ trợ các đối tượng người dùng trong việc ra quyết định. Do đó, khai thác ý kiến ứng dụng trong phân tích thái độ, sở thích, hành vi, quá trình ra quyết định chọn dịch vụ và đánh giá chất lượng dịch vụ du lịch dựa trên dữ liệu lớn là một hướng nghiên cứu cần được mở rộng hơn cho cộng đồng các nhà khoa học tại Việt Nam.
- Trong lĩnh vực du lịch, có rất ít bằng chứng thực nghiệm cho thấy có nghiên cứu thực hiện khai thác ý kiến đối với ngôn ngữ tiếng Việt. Phải chăng khai thác ý kiến trên ngôn ngữ tiếng Việt là một thách thức đối với cộng đồng các nhà nghiên cứu Việt Nam?.
- Bối cảnh nghiên cứu tại Việt Nam gần như **chưa có nghiên cứu toàn diện về hệ thống hỗ trợ đánh giá và khuyến nghị dịch vụ du lịch dựa trên thế mạnh của phương pháp khai thác ý**

kiến khách hàng trực tuyến. Đây cũng chính là khoảng trống lớn nhất mà mục tiêu của nghiên cứu này hướng đến nghiên cứu xây dựng hệ thống này.

2.6 Mô hình nghiên cứu đề xuất

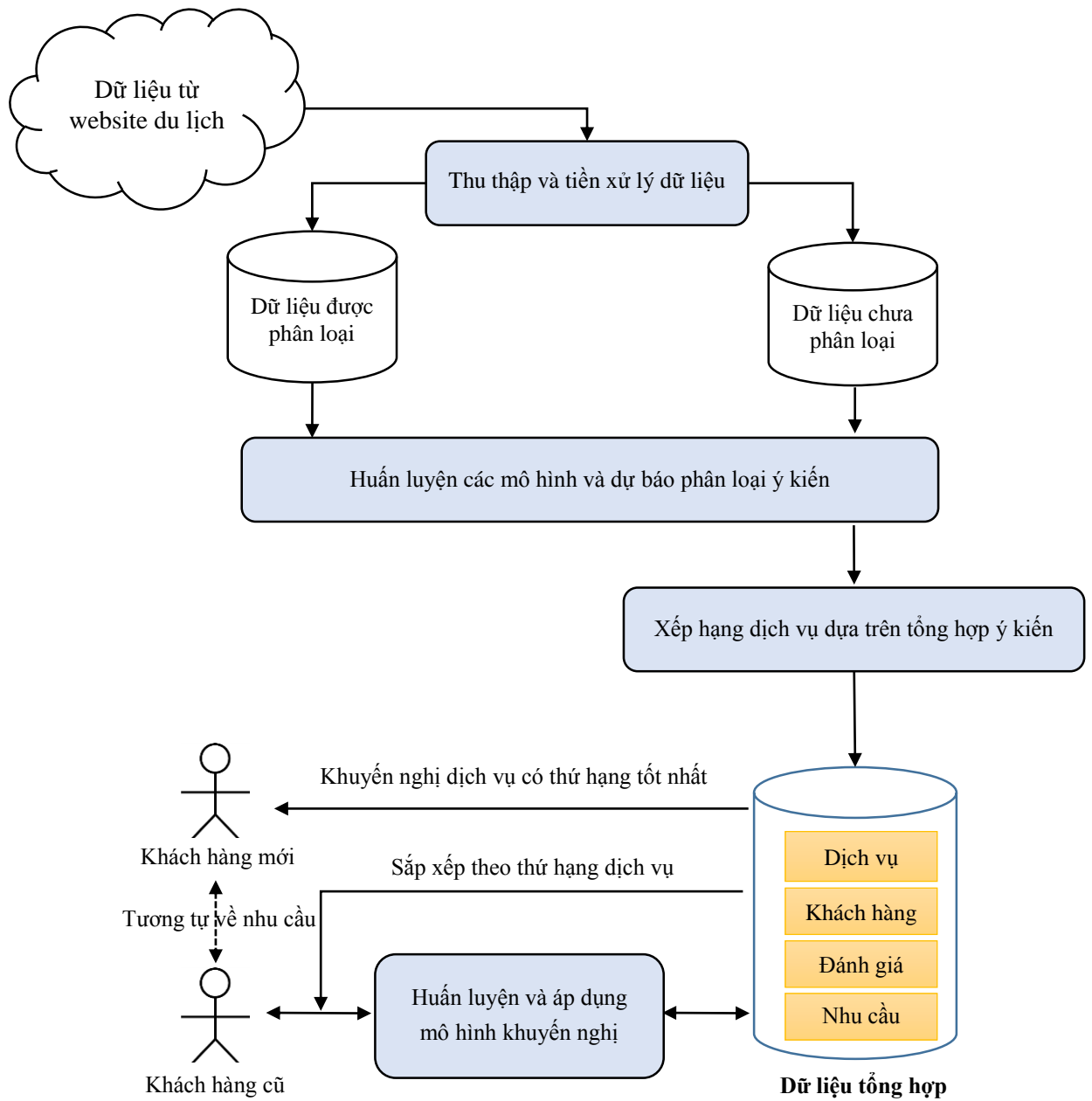
Xuất phát từ cơ sở lý thuyết, các công trình nghiên cứu liên quan và các khoảng trống nghiên cứu đã được nhận dạng, mô hình nghiên cứu tổng quát được đề xuất như hình sau:



Hình 2: Mô hình nghiên cứu tổng quát

3 PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu này được tiến hành theo phương pháp khai thác dữ liệu theo chuẩn công nghiệp CRISP-DM (Cross Industry Standard Process for Data Mining). Quy trình gồm các bước: (1) Thu thập dữ liệu và Tiền xử lý dữ liệu, (2) Huấn luyện các mô hình và dự báo phân loại ý kiến, (3) Xếp hạng dịch vụ dựa trên tổng hợp ý kiến, và (4) Huấn luyện và áp dụng mô hình khuyến nghị dịch vụ. Môi trường thực nghiệm nghiên cứu được cài đặt bằng ngôn ngữ lập trình Python với sự hỗ trợ của công cụ tách từ Python Vietnamese Toolkit (dành cho ngôn ngữ tiếng Việt) và các thư viện có sẵn.



Hình 3: Mô hình hệ thống hỗ trợ đánh giá và khuyến nghị dịch vụ du lịch

3.1 Thu thập và tiền xử lý dữ liệu

Nghiên cứu này đã tiến hành thu thập dữ liệu bằng chương trình tự động, dữ liệu lấy từ trang web Agoda.com. Đây là phương pháp thu thập nội dung tự động từ các trang HTML của bất kỳ tài nguyên Internet bằng các chương trình hoặc mã lệnh đặc biệt. Với đối tượng và phạm vi nghiên cứu hướng đến là ngôn ngữ tiếng Việt, do đó dữ liệu chỉ sử dụng những bình luận của khách hàng bằng tiếng Việt. Tiếp đến, nghiên cứu đã tiến hành tiền xử lý dữ liệu bằng cách loại bỏ những dữ liệu khuyết, những bình luận không chứa đựng thông tin cần thiết để tiến hành bước xử lý tiếp theo.

3.2 Huấn luyện các mô hình và dự báo phân loại ý kiến

Đây là giai đoạn quan trọng nhất của một nghiên cứu khai thác ý kiến, nhằm mục đích xác định một bình luận của khách hàng là “tích cực” hay “tiêu cực”. Nghiên cứu này ứng dụng một số thuật toán phân loại thuộc nhóm máy học giám sát (Supervised Machine Learning) được cho là tốt nhất dựa trên kết quả tổng hợp từ các nghiên cứu trước có liên quan đến đề tài để tìm ra mô hình phù hợp nhất đối với tập dữ liệu là các bình luận đã được phân loại, từ đó tiến hành dự báo cho các dữ liệu bình luận chưa được phân loại hoặc các dữ liệu bình luận mới phát sinh mà không cần phải huấn luyện lại. Nghiên cứu này tiến hành huấn luyện

bằng 6 thuật toán, bao gồm: Naïve Bayes (NB), Support Vector Machines (SVM), Logistic regression (LR), Neural Network (NN), DecisionTree (DT), RandomForest (RF).

Nghiên cứu này dùng cách đánh giá phổ biến là dựa trên các chỉ số tính toán trong ma trận sai lầm (Confusion Matrix). Thông thường, hiệu quả của mô hình phân loại ý kiến được đánh giá dựa trên 4 chỉ số: Accuracy, Precision, Recall, F1.

Trong đó:

- $Accuracy = \frac{TN+TP}{TN+TP+FP+FN}$
- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$

3.3 Xếp hạng dịch vụ dựa trên tổng hợp ý kiến

Sau khi thực hiện phân loại ý kiến bằng phương pháp máy học được xem là tốt nhất trong bước huấn luyện và đánh giá mô hình, kết quả có một tập dữ liệu các bình luận thể hiện quan điểm về khách sạn. Việc tổng hợp và xếp hạng dịch vụ du lịch theo quan điểm của du khách có thể dựa vào công thức đơn giản như sau:

$$H_i = \frac{TP_i - TN_i}{TR_i} \quad (1)$$

Trong đó:

- H_i : là điểm xếp hạng của dịch vụ thứ i , với $i = 1, 2, \dots, m$ (trong đó m là tổng số dịch vụ).
- TP_i : là số bình luận được đánh giá là “tích cực” của dịch vụ thứ i
- TN_i : là số bình luận được đánh giá là “tiêu cực” của dịch vụ thứ i
- TR_i : là tổng số bình luận nhận được của dịch vụ thứ i

Ngoài ra, ta có thể chuẩn hóa theo thang đo khác tùy thuộc vào mục đích so sánh, đánh giá các dịch vụ:

$$R'_i = a + \frac{(H_i - \text{Min}\{H_1, H_2, \dots, H_n\}) \times (b - a)}{\text{Max}\{H_1, H_2, \dots, H_n\} - \text{Min}\{H_1, H_2, \dots, H_n\}} \quad (2)$$

Trong đó:

- R'_i : là điểm xếp hạng của dịch vụ thứ i đã được chuẩn hóa
- a và b là cận dưới và cận trên của thang đo (ví dụ nếu đưa về thang đo 1 đến 5 điểm, lúc này $a = 1$, $b = 5$)

3.4 Huấn luyện và áp dụng mô hình khuyến nghị dịch vụ

Mục đích của giải pháp này là khuyến nghị cho một khách hàng cụ thể (kể cả những người đã có hoặc chưa có trong hệ thống) quyết định chọn dịch vụ nào phù hợp nhất.

- Trường hợp khách hàng mới hoàn toàn, không cung cấp bất kỳ thông tin về nhu cầu hoặc sở thích du lịch, hệ thống sẽ dựa trên kết quả xếp hạng dịch vụ đã được xác định trong công thức (2) để khuyến nghị cho khách hàng.
- Trường hợp khách hàng mới, có cung cấp thông tin về nhu cầu hoặc sở thích du lịch, hệ thống sẽ tìm một khách hàng cũ (có trong dữ liệu) tương tự nhất với khách hàng hiện hành, và dựa kết quả của mô hình khuyến nghị sẽ đề xuất một danh sách các dịch vụ được dự đoán là khách hàng này yêu thích (được gán nhãn là “tích cực”). Sau đó, hệ thống sẽ kết hợp thêm kết quả xếp hạng dịch vụ đã được xác định trong công thức (2) để khuyến nghị cho khách hàng mới.

Dựa dữ liệu thực tế thu thập được và kết quả tính toán ở các giai đoạn trên, nghiên cứu này áp dụng mô hình khuyến nghị dựa trên lọc cộng tác (collaborative filtering). Nghiên cứu này tiến hành huấn luyện và dự đoán bằng các thuật toán để tìm ra giải pháp tốt nhất, các thuật toán bao gồm: Normal Distribution, K-Nearest Neighbors, Co-Clustering và Matrix Factorization. Phương pháp huấn luyện K-Fold (K=5) được sử dụng trong bước này.

Thuật toán dự đoán tốt nhất sẽ có sai số RMSE (Root Mean Squared Error) và thời gian xử lý thấp nhất. Thuật toán này sẽ được dùng để xây dựng chức năng khuyến nghị cho khách hàng lựa chọn dịch vụ. Sai số RMSE được tính bởi công thức sau:

$$RMSE = \frac{1}{|R|} \sum_{r_{ui} \in R} (r_{ui} - r'_{ui})^2 \quad (3)$$

Trong đó:

- R là tập dữ liệu dùng để kiểm thử (dữ liệu test)

- r_{ui}^2 là giá trị dự đoán khách hàng u đánh giá cho dịch vụ i .
- r_{ui} là giá trị đánh giá của khách hàng u đối với dịch vụ i .

4 KẾT QUẢ NGHIÊN CỨU

4.1 Kết quả thu thập và tiền xử lý dữ liệu

Kết quả thu thập dữ liệu được 39.976 bình luận từ năm 2012 đến năm 2019, bao gồm cả tiếng Anh và tiếng Việt. Sau khi lọc bỏ các bình luận bằng tiếng Anh, chỉ giữ lại các bình luận bằng Tiếng Việt, dữ liệu còn lại để tiến hành thực nghiệm là 14.559 bình luận của 521 khách sạn ở 40 tỉnh thành trên cả nước, được phân bố như trong bảng sau:

Bảng 1: Kết quả thu thập và tiền xử lý dữ liệu

STT	Tỉnh/Thành phố	Số khách sạn	Số lượt bình luận	Trung bình
1	Buôn Ma Thuột	13	338	26,0
2	Bình Dương	2	15	7,5
3	Bạc Liêu	2	24	12,0
4	Bắc Giang	2	26	13,0
5	Bến Tre	3	70	23,3
6	Châu Đốc	8	186	23,3
7	Cà Mau	2	39	19,5
8	Cần Thơ	24	819	34,1
9	Hà Nội	43	685	15,9
10	Hà Tĩnh	5	51	10,2
11	Hòa Bình	3	26	8,7
12	Hạ Long	21	548	26,1
13	Hải Phòng	16	152	9,5
14	Hồ Chí Minh	38	1094	28,8
15	Kon Tum	2	20	10,0
16	Long Xuyên	3	22	7,3
17	Lạng Sơn	3	28	9,3
18	Mỹ Tho (Tiền Giang)	3	26	8,7
19	Nam Định	10	66	6,6
20	Nha Trang	38	1326	34,9
21	Ninh Bình	19	263	13,8
22	Phan Rang - Tháp Chàm (Ninh Thuận)	4	92	23,0
23	Phan Thiết	41	1659	40,5
24	Phú Thọ	2	15	7,5
25	Pleiku (Gia Lai)	6	133	22,2
26	Quy Nhơn (Bình Định)	17	362	21,3
27	Quảng Ngãi	7	71	10,1
28	Rạch Giá (Kiên Giang)	4	67	16,8
29	Sầm Sơn (Thanh Hóa)	11	255	23,2
30	Tam Kỳ (Quảng Nam)	2	37	18,5
31	Thái Nguyên	2	12	6,0
32	Tuy Hòa (Phú Yên)	12	311	25,9
33	Tây Ninh	3	21	7,0
34	Vinh	9	231	25,7
35	Vĩnh Phúc	3	44	14,7
36	Vũng Tàu	37	1606	43,4
37	Đà Lạt	41	1665	40,6
38	Đà Nẵng	38	1756	46,2

39	Đông Hà (Quảng Trị)	3	36	12,0
40	Đồng Hới (Quảng Bình)	19	362	19,1
Tổng cộng		521	14.559	27,9

4.2 Kết quả huấn luyện mô hình và dự báo phân loại ý kiến

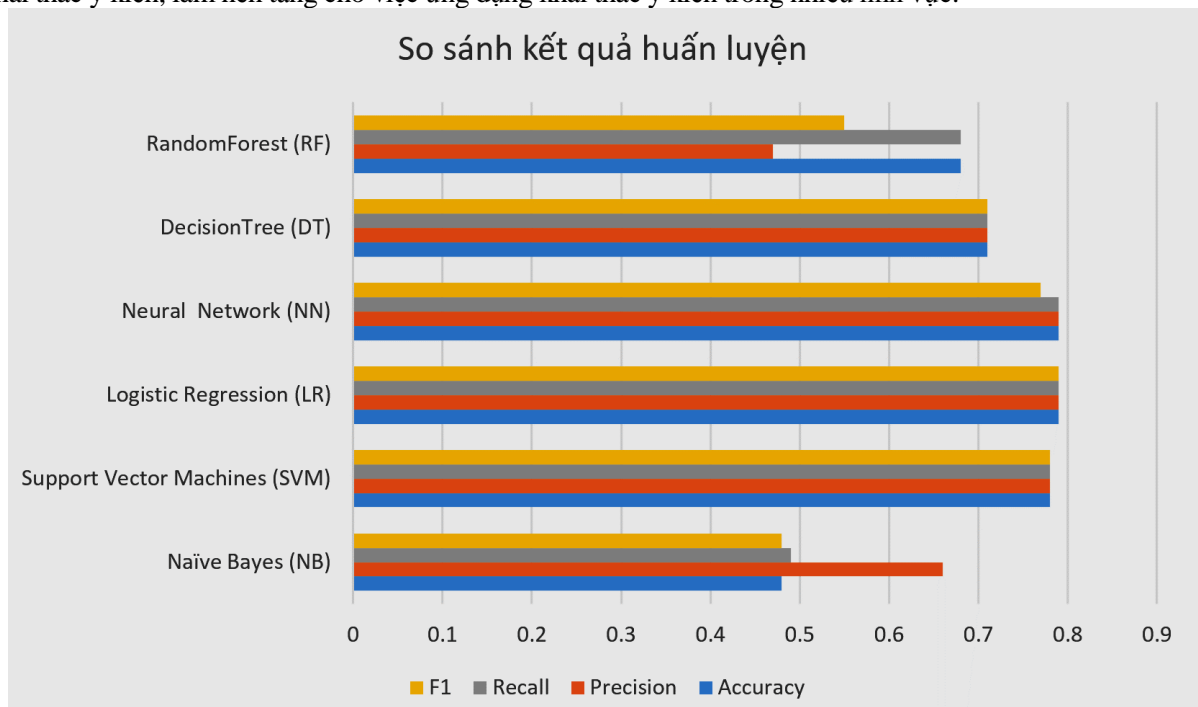
Để thực hiện việc huấn luyện bằng các mô hình máy học cần phải chuẩn bị tập dữ liệu đã được gán nhãn (hay đã được phân loại) đủ lớn để đưa vào làm tập dữ liệu huấn luyện. Thông thường đối với các nghiên cứu ứng dụng phương pháp máy học, tập dữ liệu này sẽ được xây dựng bằng thủ công. Tuy nhiên trong nghiên cứu này, sau khi xem xét ngẫu nhiên nội dung của tập dữ liệu bình luận đã thu thập được và dựa vào kết quả điểm đánh giá (trường rating trong tập dữ liệu), nghiên cứu này nhận thấy các bình luận có điểm đánh giá nhỏ hơn 7,0 mang ý nghĩa tiêu cực (negative) và ngược lại các bình luận có điểm đánh giá lớn hơn 7,0 mang ý nghĩa tích cực (positive). Do đó tập dữ liệu huấn luyện được xác định có 14.559 bình luận, trong đó có 4.506 bình luận là tiêu cực (được gán nhãn 0) và 10.053 bình luận là tích cực (được gán nhãn 1).

Nghiên cứu này tiến hành huấn luyện bằng 6 thuật toán, bao gồm: Naïve Bayes (NB), Support Vector Machines (SVM), Logistic regression (LR), Neural Network (NN), DecisionTree (DT), RandomForest (RF).

Bảng 2: Kết quả huấn luyện thuật toán phân loại ý kiến

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

Kết quả huấn luyện cho thấy các mô hình LR, SVM và NN có độ chính xác khá cao (lần lượt là 0,79; 0,78 và 0,79). Nghĩa là các mô hình này tương đối phù hợp với tập dữ liệu huấn luyện. Tuy nhiên, nếu xét thêm yếu tố thời gian huấn luyện thì chỉ nên chọn mô hình LR và SVM là tốt nhất. Do đó, các ứng dụng tiếp theo có thể dùng 2 mô hình này như một công cụ để phân loại ý kiến cho các dữ liệu bình luận chưa được phân loại hoặc các dữ liệu bình luận mới phát sinh mà không cần phải huấn luyện lại. Kết quả nghiên cứu này đã 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 nền tảng cho việc ứng dụng khai thác ý kiến trong nhiều lĩnh vực.



Hình 4: Kết quả huấn luyện mô hình phân loại ý kiến

4.3 Kết quả xếp hạng dịch vụ

Sau khi thực hiện phân loại ý kiến bằng phương pháp máy học được xem là tốt nhất trong bước huấn luyện và đánh giá mô hình, kết quả có một tập dữ liệu các bình luận thể hiện quan điểm về dịch vụ khách sạn. Việc xếp hạng khách sạn theo quan điểm của du khách được thực hiện như trong công thức (1) và (2). Nghiên cứu này sử dụng thang điểm từ 1 đến 5 và thực hiện làm tròn điểm số. Kết quả cho thấy khách sạn được đánh giá 4 điểm chiếm số lượng cao nhất (209 khách sạn) và khách sạn được đánh giá 1 điểm chiếm số lượng thấp nhất (12 khách sạn).



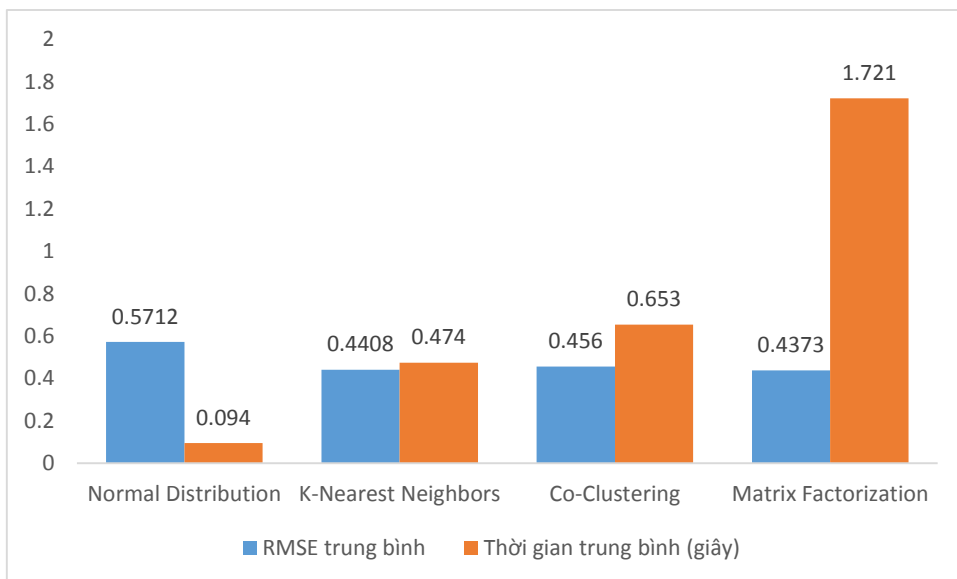
Hình 5: Kết quả xếp hạng khách sạn

4.4 Kết quả huấn luyện và áp dụng mô hình khuyến nghị dịch vụ

Nghiên cứu này tiến hành huấn luyện và dự đoán bằng các thuật toán để tìm ra giải pháp khuyến nghị tốt nhất, các thuật toán được xác định trong phần thiết kế giải pháp, bao gồm: Normal Distribution, K-Nearest Neighbors, Co-Clustering và Matrix Factorization

Bảng 3: Kết quả huấn luyện các mô hình khuyến nghị

STT	Tên thuật toán	RMSE trung bình	Thời gian trung bình (giây)
1	Normal Distribution	0,5712	0,094
2	K-Nearest Neighbors	0,4408	0,474
3	Co-Clustering	0,4560	0,653
4	Matrix Factorization	0,4373	1,721



Hình 6: Kết quả huấn luyện các mô hình khuyến nghị

Kết quả trên cho thấy phương pháp Matrix Factorization có sai số thấp nhất là 0,4373; kế đến là phương pháp K-Nearest Neighbors có sai số 0,4408; thứ ba là Co-Clustering với sai số 0,456 và cuối cùng là Normal Distribution có sai số cao nhất là 0,5712. Do chênh lệch về thời gian xử lý giữa các phương pháp không đáng kể nên nghiên cứu này quyết định chọn phương pháp Matrix Factorization làm giải pháp để xây dựng chức năng khuyến nghị dịch vụ.

4.5 Thảo luận kết quả nghiên cứu

Từ các kết quả nghiên cứu, tác giả đưa ra một số khuyến nghị để có thể triển khai hệ thống trong thực tiễn của ngành du lịch.

- **Về phạm vi triển khai hệ thống:** Với đặc tính của hệ thống này dựa trên nguồn dữ liệu thứ cấp có sẵn, nên đề xuất cho các doanh nghiệp có thể: (1) xây dựng và triển khai hệ thống theo dạng tích hợp vào hệ thống quản trị quan hệ khách hàng (CRM) để đánh giá chất lượng dịch vụ của chính doanh nghiệp đó; (2) xây dựng và triển khai như một hệ thống độc lập, có thể khai thác các nguồn dữ liệu trên mạng để đánh giá chất lượng dịch vụ của nhiều doanh nghiệp khác nhau trong ngành du lịch.
- **Về công nghệ lưu trữ và xử lý dữ liệu:** Khả năng hỗ trợ cho các nhà quản trị ra các quyết định trong đánh giá và cải tiến chất lượng dịch vụ có được chính xác và nhanh chóng phụ thuộc nhiều vào độ lớn của dữ liệu và năng lực xử lý của hệ thống máy tính. Do đó, khi triển khai thực tế, doanh nghiệp cần tính đến ứng dụng các công nghệ lưu trữ dữ liệu lớn (Big data), bên cạnh đó là xem xét khả năng trang bị hệ thống tính toán hiệu suất cao (High Performance Computing). Tuy nhiên, vấn đề này cũng cần phải đánh giá tính khả thi của dự án, giữa chi phí đầu tư hạ tầng công nghệ thông tin và lợi ích kỳ vọng mà hệ thống này mang lại.
- **Về quy trình vận hành hệ thống:** Đây là một dạng hệ thống hỗ trợ ra quyết định, việc sử dụng các kết quả phân tích từ hệ thống thường mang tính định kỳ, tùy thuộc vào nhu cầu đánh giá chất lượng dịch vụ và tần suất ra quyết định cải tiến chất lượng dịch vụ của nhà quản trị. Do đó, các chức năng thu thập dữ liệu, tiền xử lý dữ liệu, huấn luyện lại các mô hình phân loại ý kiến và khuyến nghị dịch vụ cũng nên được thực hiện một cách định kỳ. Bên cạnh đó cũng tùy thuộc vào tốc độ tăng trưởng và biến động của nguồn dữ liệu bình luận của khách hàng trên các trang mạng và các nền tảng xã hội.
- **Về người sử dụng hệ thống:** Ngoài đối tượng sử dụng chính của hệ thống là nhà quản trị doanh nghiệp và khách hàng, các chức năng thu thập dữ liệu, tiền xử lý dữ liệu, huấn luyện, đánh giá và lựa chọn các mô hình phân loại ý kiến và khuyến nghị dịch vụ nên được thực hiện bởi các chuyên gia tri thức, đặc biệt là các chuyên gia về khoa học dữ liệu (Data scientists).

5 KẾT LUẬN

Nghiên cứu này đã đề xuất một mô hình kiến trúc hệ thống cùng với các giải pháp hỗ trợ đánh giá và khuyến nghị dịch vụ du lịch dựa trên phương pháp khai thác ý kiến khách hàng trực tuyến. Kết quả thực nghiệm các giải pháp đã cung cấp một nền tảng giúp cho các công ty kinh doanh dịch vụ du lịch, các nhà quản trị có cơ sở để so sánh, đánh giá các dịch vụ du lịch. Đồng thời có thể khuyến nghị, khuyến nghị cho khách hàng lựa chọn được dịch vụ du lịch phù hợp. Tuy nhiên, nghiên cứu này vẫn còn nhiều hạn chế, có thể cải tiến tốt hơn hoặc trong những nghiên cứu tiếp theo.

Thứ nhất, hạn chế về đối tượng và phạm vi nghiên cứu: Nghiên cứu này chỉ thu thập dữ liệu là các bình luận của khách hàng về khách sạn trên trang web Agoda. Nghiên cứu có thể mở rộng thu thập các nhận xét về bất kỳ các sản phẩm hoặc dịch vụ khác trên các website thương mại điện tử hoặc các trang mạng xã hội.

Thứ hai, hạn chế về dữ liệu nghiên cứu: Về thu thập dữ liệu, do phạm vi của đề tài chỉ giới hạn thu thập dữ liệu từ trang web du lịch Agoda. Nghiên cứu có thể mở rộng thu thập thêm các nhận xét về dịch vụ khách sạn trên các trang như Booking, Tripadvisor, hoặc mạng xã hội. Tuy nhiên, nghiên cứu sẽ gặp một thách thức lớn trong việc hợp nhất và đồng bộ các nguồn dữ liệu khác nhau.

Thứ ba, hạn chế về phương pháp nghiên cứu: Nghiên cứu này chỉ phân loại ý kiến khách hàng theo thang đo hai mức: tích cực (positive) và tiêu cực (negative). Hướng nghiên cứu kế tiếp có thể sử dụng thang đo nhiều mức hơn (ví dụ theo thang đo Likert 5 mức). Bên cạnh đó, về kỹ thuật phân loại ý kiến, nghiên cứu này chỉ sử dụng phương pháp máy học có giám sát, nếu kết hợp lại với phương pháp từ vựng dựa trên ngữ nghĩa có thể sẽ cho kết quả tốt hơn. Tuy nhiên, hiện tại chưa có nhiều công cụ hỗ trợ xử lý tiếng Việt tốt

như đối với ngôn ngữ tiếng Anh. Điều này cũng mở ra nhiều cơ hội cho cộng đồng các nhà nghiên cứu trong và ngoài nước.

Cuối cùng, hạn chế trong cách đánh giá kết quả nghiên cứu: Kết quả nghiên cứu mới chỉ dừng lại ở việc đánh giá các giải pháp xử lý dữ liệu, nghiên cứu này cần thêm thời gian để hoàn chỉnh hệ thống, sau đó triển khai và lấy ý kiến người dùng để đánh giá và kiểm nghiệm hệ thống vận hành thực tế.

TÀI LIỆU THAM KHẢO

- [1]. A. Dhokrat, S. Khillare, and C. N. Mahender, Review on techniques and tools used for opinion mining, *International Journal of Computer Applications Technology and Research*, 2015, vol. 6, no. 4, pp. 419-424.
- [2]. A. Hendawi, and H. El-Shishny, Data Warehouse Prototype for the Tourism Industry: A Case Study from Egypt, in *International Conference on Informatics and Systems*, 2008.
- [3]. A. Hinze, and S. Junmanee, Advanced recommendation models for mobile tourist information, in *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, 2006, pp. 643-660.
- [4]. A. Kumar, S. Abirami, Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method, *Information Sciences*, 2018, vol. 460, pp. 23-41.
- [5]. A. Parasuraman, V. A. Zeithaml, and L. L. Berry, A conceptual model of service quality and its implications for future research, *The Journal of Marketing*, 1985, pp. 41-50.
- [6]. A. Parasuraman, V. A. Zeithaml, and L. L. Berry, SERVQUAL: A multi-item scale for measuring customer perceptions of service quality, *Journal of Retailing*, 1988, vol. 1, no. 64, pp. 12-40.
- [7]. A. S. A. Ali, Opinion Mining Techniques, *International Journal of Innovative Science, Engineering and Technology-IJISSET*, 2015, vol. 6, no. 2, pp. 752-755.
- [8]. B. Liu, Sentiment analysis and opinion mining, *Synthesis Lectures on Human Language Technologies*, 2012, vol. 1, no. 5, pp. 1-167.
- [9]. B. Pang, and L. Lee, Opinion mining and sentiment analysis, *Foundations and trends in information retrieval*, 2008, vol. 2, no. 2, pp. 1-135.
- [10]. C. Bucur, Using Opinion Mining Techniques in Tourism, *Procedia Economics and Finance*, 2015, vol. 23, pp. 1666-1673.
- [11]. C. Gronroos, A service quality model and its marketing implications, *European Journal of Marketing*, 1984, vol. 4, no. 18, pp. 36-44.
- [12]. C. Petropoulos, A. Patelis, K. Metaxiotis, K. Nikolopoulos, and V. Assimakopoulos, SFTIS: A decision support system for tourism demand analysis and forecasting, *Journal of Computer Information Systems*, 2003, vol. 1, no. 44, pp. 21-32.
- [13]. C. Yu, X. Zhu, B. Feng, L. Cai, and L. An, Sentiment Analysis of Japanese Tourism Online Reviews, *Journal of Data and Information Science*, 2019, vol. 1, no. 4, pp. 89-113.

- [14]. Đ. Bunja, B. K. Miočić, and P. Nekić, Possibilities for implementation of the decision support system in the Croatian tourism industry, in *30th International Convention MIPRO 2008-CONFERENCE ON BUSINESS INTELLIGENCE SYSTEMS*, 2007.
- [15]. E. Bjørkelund, T. H. Burnett, and K. Nørnvåg, A study of opinion mining and visualization of hotel reviews, in *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*, 2012, pp. 229-238.
- [16]. E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, A novel deterministic approach for aspect-based opinion mining in tourism products reviews, *Expert Systems with Applications*, 2014, vol. 17, no. 4, pp. 7764-7775.
- [17]. F. Ricci, *Travel recommender systems*, IEEE Intelligent Systems, 2002, vol. 6, no. 17, pp. 55-57.
- [18]. G. Stalidis, D. Karapistolis, and A. Vafeiadis, Marketing decision support using Artificial Intelligence and Knowledge Modeling: application to tourist destination management, *International Conference on Strategic Innovative Marketing*, 2014.
- [19]. I. N. Wahab, Role of information technology in tourism industry: Impact and growth, *International Journal of Innovative Research in Computer and Communication Engineering*, 2017, vol. 2, no. 5, pp. 260-263.
- [20]. I. Perikos, K. Kovas, F. Grivokostopoulou, and L. Hatzilygeroudis, A System for Aspect-based Opinion Mining of Hotel Reviews, in *WEBIST*, 2017, pp. 388-394.
- [21]. J. A. Martilla, and J. C. James, Importance-performance analysis, *Journal of marketing*, 1997, vol. 41, no. 1, pp. 77-79.
- [22]. J. J. Cronin, and S. A. Taylor, Measuring service quality: a reexamination and extension, *Journal of marketing*, 1992, vol. 56, no. 3, pp. 55-68.
- [23]. J. P. Bousset, D. Skuras, J. Těšitel, J. B. Marsat, A. Petrou, E. Fiallo-Pantziou, and M. Bartoš, A decision support system for integrated tourism development: Rethinking tourism policies and management strategies', *Tourism Geographies*, 2007, vol. 4, no. 9, pp. 387-404.
- [24]. J. Wang, and C. Li, Design and implementation of tourist WebGIS based on J2EE, In *22nd International Cartographic Conference*, 2005.
- [25]. K. W. Wöber, Information supply in tourism management by marketing decision support systems, *Tourism Management*, 2003, vol. 3, no. 24, pp. 241-255.
- [26]. M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews, *Advances in Fuzzy Systems*, 2016.
- [27]. M. Danubianu, T. Socaciu, and A. Barila, Some aspects of data warehousing in tourism industry, *The USV Annals of Economics and Public Administration*, 2009, vol. 9, no. 1, pp. 290-296.

- [28]. N. Becser, Z. Zoltay-Paprika, A Decision Support Model for Improving Service Quality, SQI-DSS: A New Approach, *Budapest University of Economics and Public Administration, Department of Business Economics*, 2003.
- [29]. N. T. Duyen, N. X. Bach, and T. M. Phuong, An empirical study on sentiment analysis for Vietnamese, in *Advanced Technologies for Communications (ATC), 2014 International Conference*, 2014, pp. 309-314.
- [30]. N. Yussupova, G. Kovács, M. Boyko, and D. Bogdanova, Models and Methods for Quality Management Based on Artificial Intelligence Applications, *Acta Polytechnica Hungarica*, 2016.
- [31]. Q. Ye, Z. Zhang, and R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, *Expert Systems with Applications*, 2009, vol. 3, no. 36, pp. 6527-6535.
- [32]. R. Baggio and L. Caporarello, Decision support systems in a tourism destination: literature survey and model building, in *U: Proceedings itAIS-2nd Conference of the Italian chapter of AIS*, 2005.
- [33]. S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, and J. Reynar, Building a sentiment summarizer for local service reviews, in *WWW workshop on NLP in the information explosion era*, 2008, pp. 339-348.
- [34]. S. K. Yadav, Sentiment analysis and classification: A survey, *International Journal of Advance Research in Computer Science and Management Studies*, 2015.
- [35]. S. M. Mudambi, and D. Schuff, What makes a helpful review? A study of customer reviews on Amazon.com, *MIS quarterly*, 2010, vol. 1, no. 34, pp. 185-200.
- [36]. S. P. Singh, J. Sharma, and P. Singh, A web-based tourist decision support system for Agra City, *International Journal of Instrumentation, Control & Automation (IJICA)*, 2011, vol. 1, no. 1, pp. 51-54.
- [37]. S. Sun, C. Luo, and J. Chen, A review of natural language processing techniques for opinion mining systems, *Information Fusion*, 2017, vol. 36, pp. 10-25.
- [38]. T. Abdul-Aziz, I. Moawad, and W. M. Abu-Alam, Decision Support System Utilizing Data Warehouse Technique for the Tourism Sector in Egypt, in *The 7th International Conference on Information Technology*, 2015.
- [39]. T. Masron, N. Ismail, and A. Marzuki, The conceptual design and application of web-based tourism decision support systems, *Theoretical and Empirical Researches in Urban Management*, 2016, vol. 2, no. 11, pp. 64-75.
- [40]. T. S. Bang, C. Haruechaiyasak, and V. Sornlertlamvanich, Vietnamese sentiment analysis based on term feature selection approach, in *Proceedings of The Tenth International Conference on Knowledge, Information and Creativity Support Systems*, 2015, pp. 12-14.

- [41]. V. A. Zeithaml, A. Parasuraman, and L. L. Berry, *Delivering quality service: Balancing customer perceptions and expectations*, 1990.
- [42]. W. B. Claster, M. Cooper, and P. Sallis, Thailand-Tourism and conflict: Modeling sentiment from Twitter tweets using naïve Bayes and unsupervised artificial neural nets, in *Computational Intelligence, Modelling and Simulation (CIMSIM), 2010 Second International Conference*, 2010, pp. 89-94.
- [43]. W. Duan, Q. Cao, Y. Yu, and S. Levy, Mining online user-generated content: using sentiment analysis technique to study hotel service quality, in *System Sciences (HICSS), 2013 46th Hawaii International Conference*, 2013, pp. 3119-3128.
- [44]. W. Kasper, M. Vela, Sentiment analysis for hotel reviews, in *Computational linguistics-applications conference*, 2011, pp. 45-52.
- [45]. W. Medhat, A. Hassan, and H. Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 2014, vol. 4, no. 4, pp. 1093-1113.
- [46]. X. Qiao, L. Zhang, N. Li, and W. Zhu, Constructing a data warehouse based decision support platform for China tourism industry, in *Information and Communication Technologies in Tourism 2014*, 2014, pp. 883-893.
- [47]. Y. Asafe, A. Bolaji, A. Enaholo, and O. Olubukola, Web-based expert decision support system for tourism destination management in Nigeria, *International Journal of Advanced Research in Artificial Intelligence*, 2013, vol. 4, no. 2, pp. 59-63.
- [48]. Y. Guo, S. J. Barnes, and Q. Jia, Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation, *Tourism Management*, 2017, vol. 59, pp. 467-483.
- [49]. Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, OpinionSeer: interactive visualization of hotel customer feedback, *IEEE transactions on visualization and computer graphics*, 2010, vol. 6, no. 16, pp. 1109-1118.

Ngày nhận bài: 04/07/2020

Ngày chấp nhận đăng: 17/06/2020