

CÁC VẤN ĐỀ DU LỊCH VIỆT NAM QUA PHÂN TÍCH

THẢO LUẬN MẠNG XÃ HỘI REDDIT

TOURISM ISSUES IN VIETNAM THROUGH ANALYSIS

OF REDDIT DISCUSSIONS

TÔN NGUYỄN TRỌNG HIỀN^(*)

THÔNG TIN	TÓM TẮT
Ngày nhận bài: 16-11-2024 Ngày biên tập xong: 02-01-2025 Ngày duyệt đăng: 22-3-2025 Mã số: TCKH49-06-2025 ISSN: 2525 – 2429	Nghiên cứu sử dụng phương pháp khai phá văn bản ngôn ngữ tự nhiên thu thập dữ liệu từ mạng xã hội Reddit nhằm khám phá cảm nhận và thảo luận của du khách về du lịch tại Việt Nam. Kết quả cho thấy mức độ quan tâm của du khách đối với các địa phương tại Việt Nam có sự khác biệt đáng kể, với sự chú ý chủ yếu tập trung vào một số khu vực nhất định. Phân tích cảm xúc cho thấy phần lớn các bình luận đều mang sắc thái tích cực, mặc dù vẫn tồn tại những mối quan tâm, trong đó vấn đề thị thực nhập cảnh được nhắc đến thường xuyên. Nghiên cứu khuyến nghị các nhà quản lý du lịch cần thúc đẩy quảng bá các điểm đến ít được biết đến, nhằm phân bổ dòng khách đồng đều và giảm áp lực tại các điểm du lịch quá tải. Việc phát triển các điểm đến chưa được khai thác nhiều và tạo sự khác biệt rõ rệt trong trải nghiệm du lịch sẽ không chỉ giảm thiểu sự cạnh tranh giữa các điểm đến mà còn thúc đẩy sự phát triển bền vững du lịch tại Việt Nam. Đồng thời, chiến lược này còn giúp củng cố vị thế cạnh tranh của du lịch Việt Nam trong khu vực ASEAN, tạo ra cơ hội phát triển cho các khu vực chưa được khai thác và làm phong phú thêm các lựa chọn cho du khách quốc tế. Ngoài ra, khuyến nghị phát triển các kênh tư vấn trực tuyến chính thống, tư vấn kịp thời để nâng cao trải nghiệm du khách.
Từ khóa: du lịch; khai phá văn bản; Reddit; Việt Nam. Key words: tourism; text-minining; Reddit; Vietnam.	ABSTRACT: This study employs natural language processing methods to analyze data collected from the social media platform Reddit, aiming to explore tourists' perceptions and discussions regarding tourism in Vietnam. The results show that the level of interest among tourists in different regions of Vietnam varies significantly, with attention mainly focused on certain areas. Sentiment analysis reveals that the majority of comments are positive, although there are still notable concerns, particularly regarding visa

^(*) ThS. Trường Đại học Văn Lang, hien.tnt@vlu.edu.vn

issues, which are frequently mentioned. The study recommends that tourism managers should promote lesser-known destinations to distribute tourist traffic more evenly and reduce the pressure on overcrowded tourist spots. Developing underexplored destinations and creating distinct differences in the travel experience will not only minimize competition between destinations but also foster sustainable tourism development in Vietnam. Additionally, this strategy will help strengthen the competitive position of Vietnam's tourism within the ASEAN region, creating development opportunities for underserved areas and enriching the choices available to international tourists. Additionally, it is essential to develop official online consultation channels and provide timely advice to enhance the tourist experience.

1. ĐẶT VẤN ĐỀ

Du lịch là một trong những ngành kinh tế mũi nhọn của nhiều quốc gia trên thế giới. Ngành công nghiệp không khói này không chỉ thúc đẩy tăng trưởng kinh tế mà còn đóng vai trò quan trọng trong việc bảo tồn và phát huy các giá trị di sản văn hóa của mỗi quốc gia. Trong bối cảnh toàn cầu hóa và sự phát triển của các phương tiện giao thông, việc di chuyển giữa các quốc gia ngày càng trở nên thuận tiện. Sự bùng nổ du lịch sau đại dịch COVID-19 đã góp phần tăng trưởng mạnh mẽ cho ngành du lịch toàn cầu, trong đó có Việt Nam. Góc nhìn những năm đầu 1990, du lịch Việt Nam sớm được nhận ra đây là bức tranh mới và rất sáng. Du lịch hiện nay đã trở thành một trong những ngành kinh tế có tốc độ tăng trưởng nhanh và đóng góp đáng kể vào tổng sản phẩm quốc nội (GDP). Theo ước tính của tập đoàn đầu tư và quản lý tài sản VinaCapital, tổng đóng góp của ngành du lịch cho nền kinh tế Việt Nam bao gồm cả trực tiếp và gián tiếp, sẽ chiếm hơn 15% GDP năm 2024 [15]. Ngoài đóng góp trực tiếp này, du lịch còn có tác động gián tiếp đến nhiều ngành khác, từ dịch vụ lưu trú, nhà hàng,

đến các ngành công nghiệp sản xuất hàng lưu niệm và dịch vụ vận tải. Trong tháng 9-2024, Việt Nam đã đón khoảng 1,3 triệu lượt khách quốc tế. Tổng cộng 9 tháng đầu năm 2024, lượng khách quốc tế đạt hơn 12,7 triệu lượt, tăng 43,0% so với cùng kỳ năm ngoái, và vượt qua con số 12,6 triệu lượt của cả năm 2023 [13]. Việt Nam được biết đến với nhiều danh lam thắng cảnh nổi tiếng như Vịnh Hạ Long, Hội An, Đà Lạt, Phú Quốc... Những địa điểm này đã thu hút hàng triệu lượt khách du lịch quốc tế, góp phần quan trọng trong việc tạo ra công ăn việc làm và thúc đẩy sự phát triển kinh tế tại các địa phương. Bên cạnh đó, Việt Nam đang tập trung triển khai các chiến lược phát triển du lịch bền vững, với mong muốn khai thác tối đa tiềm năng du lịch của đất nước đồng thời bảo tồn cảnh quan thiên nhiên và văn hóa địa phương [2]. Những điều kiện thuận lợi cùng với những nỗ lực kể trên đang giúp tăng cường khả năng cạnh tranh của du lịch Việt Nam trên thị trường quốc tế, đồng thời góp phần nâng cao hình ảnh quốc gia trong bối cảnh hội nhập toàn cầu. Trong bối cảnh đó, việc nắm bắt và phân tích quan điểm, đánh giá của du khách trở

thành một yếu tố then chốt để cải thiện chất lượng dịch vụ và điều chỉnh hiệu quả các chiến lược phát triển du lịch tại Việt Nam.

Trong bối cảnh cạnh tranh quốc tế ngày càng khốc liệt, việc hiểu rõ nhu cầu và đánh giá của du khách quốc tế trở thành yếu tố sống còn để nâng cao vị thế của du lịch Việt Nam. Hiện nay, với sự phát triển của công nghệ thông tin, du khách có thể dễ dàng chia sẻ hiểu biết của chính họ thông qua các nền tảng mạng xã hội và các trang web đánh giá. Đáng chú ý, đại dịch COVID-19 đã làm thay đổi đáng kể hành vi của du khách, thúc đẩy xu hướng sử dụng các nền tảng trực tuyến để tìm kiếm thông tin và chia sẻ kinh nghiệm. Sự phát triển của các nền tảng trực tuyến này đã tạo ra một không gian nơi du khách có thể trao đổi ý kiến, đánh giá các điểm đến, và nhận thông tin từ những người đã trải nghiệm thực tế, đặc biệt trong bối cảnh hạn chế đi lại và các biện pháp giãn cách xã hội. Trong đó, Reddit là một trong những nền tảng lớn và phổ biến, nơi du khách nước ngoài thường dùng có thể tham gia vào các diễn đàn để thảo luận về nhiều chủ đề khác nhau, bao gồm du lịch. Reddit là một mạng xã hội và diễn đàn trực tuyến, nơi người dùng có thể đăng các nội dung, đặt câu hỏi, chia sẻ ý kiến hoặc tham gia thảo luận về các chủ đề cụ thể. Nhờ tính chất mở và sự tham gia đa dạng của người dùng đến từ khắp nơi trên thế giới, Reddit đã trở thành một nguồn dữ liệu phong phú cho các nhà nghiên cứu muốn phân tích ý kiến và xu hướng thảo luận của cộng đồng về các vấn đề khác nhau, bao gồm du lịch. Do đó, không gì ngạc nhiên khi cho rằng nền tảng mạng xã hội này đã thu hút sự chú ý của nhiều nhà nghiên cứu nhờ vào sự phổ biến toàn cầu của nó.

Trong nghiên cứu này, phương pháp khai phá văn bản được tiến hành để phân tích các ý kiến, đánh giá và thảo luận của du khách về du lịch Việt Nam. Khai phá văn bản là quá trình tự động trích xuất các thông tin có giá trị từ các

tập dữ liệu văn bản lớn. Cụ thể, phương pháp này bao gồm các bước như thu thập dữ liệu, làm sạch dữ liệu, phân loại văn bản và phân tích để tìm ra các xu hướng chính hoặc những vấn đề nổi bật. Thông qua khai phá văn bản, chúng tôi có thể xác định được những tình hình nào tại Việt Nam đang thu hút sự quan tâm của du khách quốc tế và những vấn đề nào đang được thảo luận nhiều nhất. Mục tiêu của nghiên cứu này là cung cấp một cái nhìn toàn diện về quan điểm của du khách quốc tế về du lịch Việt Nam, từ đó giúp các nhà quản lý du lịch có cơ sở để phát triển các chiến lược phù hợp. Bên cạnh việc xác định những địa phương thu hút nhiều du khách nhất, chúng tôi cũng tập trung phân tích các vấn đề nổi bật mà du khách đang tập trung thảo luận, như chất lượng dịch vụ, giá cả, hạ tầng du lịch, cách thức marketing và các trải nghiệm văn hóa... Điều này không chỉ giúp nhận diện các cơ hội phát triển mới mà còn chỉ ra những thách thức cần phải đối mặt để nâng cao trải nghiệm của du khách. Thông qua nghiên cứu, chúng tôi hy vọng sẽ cung cấp khuyến nghị nhằm nâng cao trải nghiệm du khách và khả năng cạnh tranh của Việt Nam trong ngành du lịch toàn cầu.

2. NỘI DUNG

2.1. Lược sử nghiên cứu

Khảo sát qua các nghiên cứu trước đây, phần lớn cung cấp cái nhìn về những yếu tố quyết định sự hài lòng của du khách thực hiện tại nhiều địa phương thông qua sử dụng kết quả khảo sát, ví dụ nghiên cứu nhu cầu du lịch trải nghiệm của người dân Thành phố Cần Thơ [1]; Các nhân tố ảnh hưởng đến sự phát triển du lịch làng nghề ở thành phố Phú Quốc [3]; Các nhân tố tác động đến quyết định lựa chọn điểm đến du lịch nông thôn dựa vào cộng đồng vùng Bắc Trung Bộ [4]. Mặc dù các nghiên cứu này đã cung cấp những hiểu biết có giá trị về các địa phương cụ thể, nhưng phạm vi của chúng bị giới hạn ở các vùng địa phương. Việc mở rộng phạm vi nghiên cứu ra toàn quốc là cần

thiết để có được cái nhìn toàn diện về các thách thức và cơ hội trong du lịch Việt Nam, bởi vì những vấn đề du khách gặp phải không chỉ giới hạn ở một địa phương mà còn xuất hiện ở nhiều điểm đến khác nhau trên khắp cả nước. Đối với nghiên cứu về phát triển du lịch Việt Nam nói chung, bên cạnh khảo sát du khách được thực hiện, các nghiên cứu còn tập trung vào quan điểm của nhà cung cấp dịch vụ hoặc các chuyên gia trong ngành và cũng chủ yếu sử dụng các phương pháp như phỏng vấn, khảo sát để đánh giá tác động của du lịch đối với các khía cạnh kinh tế, xã hội, và môi trường.

Các nghiên cứu trước đây như nghiên cứu Tseng và cộng sự [11], đề xuất bốn khía cạnh chính gồm văn hóa - xã hội, kinh tế, môi trường và sự hợp tác, với 24 tiêu chí được sử dụng để phân tích bằng phương pháp ra quyết định đa mục tiêu, kết quả cho thấy sự hợp tác bao gồm chia sẻ thông tin, minh bạch, sự điều chỉnh là thuộc tính có ảnh hưởng mạnh nhất, đóng vai trò quan trọng nhất trong việc thúc đẩy các thuộc tính khác trong du lịch bền vững. Nghiên cứu khác xem xét dưới góc nhìn của nhà tổ chức dịch vụ du lịch, nghiên cứu sử dụng 23 cuộc phỏng vấn bán cấu trúc với các quản lý cấp cao và lãnh đạo doanh nghiệp từ năm nhà điều hành tour hàng đầu tại Việt Nam. Kết quả cho thấy, trong bối cảnh gặp nhiều khó khăn hiện tại, các doanh nghiệp du lịch Việt Nam đã phát triển nhiều chiến lược ứng phó khác nhau, đó là vấn đề đáng quan tâm [12]. Có thể thấy, chưa có nghiên cứu nào tập trung vào việc phân tích ý kiến của du khách - những người trực tiếp trải nghiệm dịch vụ du lịch thông qua khai phá dữ liệu văn bản từ các nền tảng trực tuyến như mạng xã hội hoặc các trang web đánh giá. Khảo sát truyền thống được thay dần bằng phương pháp khai phá dữ liệu, đặc biệt là trong những vấn đề mới chưa khám phá, bởi việc phân tích được lượng lớn dữ liệu phân bố ngẫu nhiên, người trả lời không bị áp lực trước câu hỏi, hay người phỏng vấn. Ngoài ra, một

nghiên cứu về sự chấp nhận thông tin truyền miệng điện tử (eWOM) và niềm tin thương hiệu trong lĩnh vực du lịch Việt Nam cho rằng truyền miệng trên truyền thông có vai trò là cầu nối giữa sự tin nhiệm của nguồn tin, sự chấp nhận thông tin và việc xây dựng hình ảnh cũng như niềm tin thương hiệu trong tâm trí khách hàng [5]. Nhiều nghiên cứu gần đây trong lĩnh vực du lịch và dịch vụ khách sạn đã ghi nhận tác động của cả truyền miệng điện tử tác động tích cực lẫn tiêu cực đối với các sản phẩm du lịch. Điều này cho thấy rằng nghiên cứu về các cuộc thảo luận trên mạng xã hội là khả thi và phù hợp [9], vì các nền tảng này là nơi du khách thể hiện trực tiếp cảm nhận, đánh giá và thảo luận về trải nghiệm của họ. Do đó, việc phân tích eWOM giúp hiểu rõ hơn về thái độ của khách du lịch và cung cấp thông tin giá trị cho các nhà quản lý du lịch. Phân tích dữ liệu theo thời gian thực qua các nền tảng mạng xã hội cung cấp cái nhìn trực tiếp và chính xác về cách người tiêu dùng đánh giá và thảo luận về du lịch. Bằng cách theo dõi và phân tích các cuộc trò chuyện này, thương hiệu, doanh nghiệp, tổ chức hay chính quyền có thể quản lý và điều chỉnh hình ảnh thương hiệu của mình dựa trên phản hồi từ người tiêu dùng.

Khảo sát các nghiên cứu mạng xã hội du lịch. Mirzaalian & Halpenny [6] thực hiện lược khảo tài liệu và kết luận đến sự phát triển nhanh chóng của phân tích mạng xã hội trong lĩnh vực du lịch và khách sạn kể từ năm 2014. Cùng nhận định Nusair và cộng sự [10] cho rằng, phân tích mạng xã hội trong vấn đề du lịch là xu hướng cần thiết, nhóm tác giả bổ sung việc cần hiểu sâu về sự tương tác trên mạng xã hội, bởi đây là nơi các du khách thảo luận, chia sẻ trải nghiệm và thể hiện nhu cầu cũng như mong muốn. Qua việc phân tích tương tác này, các thương hiệu có thể xây dựng các chiến lược hiệu quả nhằm tạo ra cộng đồng gắn kết và thu hút khách hàng trong các nhóm đối tượng khác nhau. Cụ thể, Bookings.com

hay Trip.com, Tripadvisor... là những mạng xã hội chứa các đánh giá, liên quan trực tiếp đến trải nghiệm lưu trú hoặc dịch vụ, đây là nguồn dữ liệu đã được nhiều nghiên cứu trước đây sử dụng nhằm mục đích phân tích hành vi khách hàng. Trong khi đó, Reddit là mạng xã hội đang được sử dụng trong các nghiên cứu gần đây [7], [8]. Reddit là một diễn đàn mở, nơi người dùng có thể thảo luận một cách tự do về nhiều khía cạnh khác nhau của du lịch, không chỉ giới hạn ở nơi lưu trú mà còn bao gồm văn hóa, điểm đến, dịch vụ, những trải nghiệm cá nhân. Điều này mang lại cái nhìn đa chiều và toàn diện hơn về những vấn đề và xu hướng du lịch. Do đó, Reddit được lựa chọn trong nghiên cứu này vì là một nền tảng quốc tế với hàng triệu người dùng đa dạng, tạo điều kiện phản ánh xu hướng thảo luận từ nhiều góc nhìn khác nhau. So với các nền tảng như TripAdvisor hay Bookings... Reddit có tính chất mở và khả năng khuyến khích thảo luận sâu rộng, khiến nó trở thành nguồn dữ liệu độc đáo và phong phú.

2.2. Phương pháp nghiên cứu

2.2.1. Phương pháp

Trong nghiên cứu này, chúng tôi sử dụng phương pháp khai phá văn bản nhằm khám phá các chủ đề chính và cảm xúc ẩn chứa trong dữ liệu văn bản thu thập được. Để đạt được mục tiêu này, chúng tôi áp dụng phần mềm Communalitics phiên bản thương mại, một công cụ hỗ trợ mạnh mẽ cho việc tiền xử lý dữ liệu, xây dựng mô hình phân loại và phân cụm. Thông qua quá trình này, chúng tôi có thể trích xuất những thông tin có ý nghĩa từ dữ liệu, đồng thời trực quan hóa kết quả bằng các mạng lưới hoặc biểu đồ, giúp làm rõ hơn mối quan hệ giữa các chủ đề và cảm xúc trong dữ liệu. Công cụ Communalitics được sử dụng với hai mô hình trí tuệ nhân tạo chính là Detoxify và Google's Perspective API. Detoxify hỗ trợ phân tích văn bản trong đó bao gồm các ngôn ngữ như Tiếng Anh, Pháp, Tây Ban Nha, Ý, Bồ Đào Nha, Thổ Nhĩ Kỳ, và Nga. Trong khi đó,

Perspective API hỗ trợ thêm nhiều ngôn ngữ khác như tiếng Ả Rập, Trung Quốc, Séc, Đức, Hindi, Nhật, Hàn, và Thụy Điển, với tổng cộng 15 ngôn ngữ được phân tích.

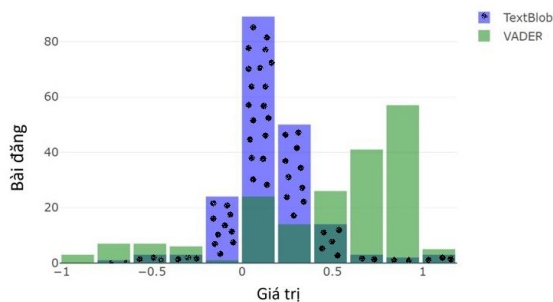
2.2.2. Dữ liệu nghiên cứu

Để truy xuất dữ liệu từ Reddit cần sử dụng Reddit API hợp lệ. Sau khi đăng ký tài khoản và thiết lập kỹ thuật, dữ liệu từ Reddit, bao gồm các bài viết, bình luận được truy xuất đầy đủ hợp lệ. Việc trích xuất từ khóa 'Vietnam Tourism' (Du lịch Việt Nam) với được thực hiện vào lúc 19:30 17-9-2024 GMT +7. Dữ liệu trước khi phân tích Communalitics sẽ tự động xác định ngôn ngữ của mỗi bài viết và loại bỏ các bài viết không được hỗ trợ, cũng như loại bỏ các từ không cần thiết (stopwords) 15 ngôn ngữ, ví dụ động từ 'am, is, are' trong tiếng Anh. Ngoài ra, phần mềm cũng loại bỏ các yếu tố như tên người dùng, hashtag, và URL để giảm thiểu kết quả sai lệch.

2.3. Kết quả nghiên cứu và thảo luận

Kết quả trích xuất thu được 196 bài viết với 1.061 người tham gia theo dõi và bình luận với chủ đề tìm kiếm. Thông qua sàng lọc mức độ phổ biến với điều kiện từ khóa xuất hiện ít nhất 100 lần, nhận thấy văn bản thu được hầu như là tiếng Anh. Kết quả phân tích cảm xúc cho thấy, sự phân phối của các giá trị độ cực trong một tập dữ liệu với tổng cộng 191 bài viết từ hai thư viện phân tích cảm xúc VADER và TextBlob. VADER phân tích cảm xúc dựa trên từ điển, phù hợp với văn bản ngắn như bình luận mạng xã hội, trong khi TextBlob kết hợp học máy và từ điển, có ưu thế với văn bản dài hơn. Theo kết quả phân tích dữ liệu (hình 1), VADER phân loại 12,57% bài viết có cảm xúc tiêu cực, 10,99% có cảm xúc trung tính, và 76,44% có cảm xúc tích cực. Tương tự, TextBlob chỉ ra 12,57% là tiêu cực, 26,18% là trung tính, và 61,26% là tích cực. Sự khác biệt giữa hai thư viện thể hiện qua phương pháp xử lý, nhưng đồng thời, sự đồng thuận trung bình (Cohen's kappa statistic = 0,414) vẫn cung cấp

mức độ đáng tin cậy cho các kết quả. Điều này cho thấy, mặc dù có một số khác biệt, việc sử dụng cả hai công cụ giúp tăng tính toàn diện và minh bạch trong phân tích cảm xúc.



Hình 1. Biểu đồ phân bố giá trị cảm xúc

Nguồn: Tác giả tổng hợp và phân tích



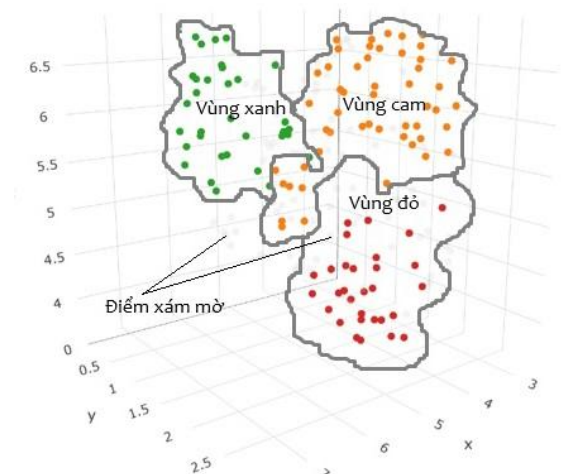
Hình 2. Phân tích wordcloud những từ vựng và biểu tượng cảm xúc phổ biến

Nguồn: Tác giả tổng hợp và phân tích

Tiếp tục với phân tích ‘word cloud’ (hình 2), từ khóa ‘Hanoi’ xuất hiện nổi bật nhất, cho thấy sự quan tâm lớn đến thủ đô Hà Nội. Những địa danh nổi tiếng như ‘Hoi An’ (Hội An), ‘Halong’ (Hạ Long), ‘Saigon’ (Thành phố Hồ Chí Minh), ‘Sapa’ (Sa pa), ‘Ha Giang’ (Hà Giang), ‘Da Nang’ (Đà Nẵng) cũng được đề cập, chứng tỏ đây là các điểm đến thu hút sự chú ý của du khách, chỉ ra sự quan tâm của du khách tới các điểm đến nổi bật tại Việt Nam, đặc biệt là khu vực miền Trung và miền Bắc. Các từ khóa khác: ‘trip’ (chuyến đi), ‘days’ (số ngày), ‘visa’ (thị thực), ‘around’ (xung quanh), ‘experience’ (trải nghiệm) phản ánh các mối quan tâm phổ biến của du khách liên quan đến hành trình du lịch, số ngày tham quan, thủ tục

xin thị thực, trải nghiệm tổng quan; Cho thấy đây là một trong những rào cản lớn với du khách. Những chủ đề về ‘planning’ (lập kế hoạch), ‘recommendations’ (đề xuất), ‘places’ (địa điểm), ‘itinerary’ (hành trình) cho thấy, du khách thường thảo luận về việc lập kế hoạch du lịch và tìm kiếm các gợi ý từ cộng đồng. Sự xuất hiện một số địa danh ngoài Việt Nam, ví dụ Thái Lan, có thể du khách đang so sánh và xin lời khuyên trải nghiệm giữa hai điểm đến. Đặc biệt, việc sử dụng biểu tượng cảm xúc trong ‘word cloud’ phản ánh cảm xúc đa dạng của du khách, cụ thể, đọc và phân tích những biểu tượng cảm xúc để tìm ra những vấn đề cần quan tâm, phát hiện rằng, cảm xúc tiêu cực liên quan vấn đề thị thực nhập cảnh, số ngày nhập cảnh và số lần được nhập cảnh là những vấn đề du khách đang quan tâm hay lo ngại.

Kết quả phân tích về du lịch Việt Nam sử dụng thuật toán HDBSCAN với các tham số epsilon là 0,05 và mẫu chung tối thiểu là 10, đã tạo ra biểu đồ phân tán 3D (hình 3). Hình ảnh hiển thị một biểu đồ phân tán 3D với các điểm có màu khác nhau đã chỉ ra được các nhóm chủ đề tiềm ẩn trong dữ liệu liên quan đến du lịch Việt Nam.



Hình 3. Biểu đồ tương đồng ngữ nghĩa

Nguồn: Tác giả tổng hợp và phân tích

Màu xám đại diện cho các điểm dữ liệu không liên quan đến bất kỳ chủ đề, khía cạnh cụ thể nào.

thể nào trong dữ liệu về du lịch Việt Nam. Có thể thấy, các điểm nhiễu này tương đối nhiều. Tiếp tục khám phá ý nghĩa các hướng thảo luận chính thông qua đọc cảm quan và dưới sự hỗ trợ thuật toán ‘mistral 7b instruct’- mô hình ngôn ngữ trí tuệ nhân tạo được huấn luyện để xử lý các tác vụ hướng dẫn và tương tác với con người một cách chính xác và tự nhiên, phát hiện rằng, vùng cụm màu xanh đỏ xác nhận các thảo luận về chính sách thị thực, số ngày cho phép nhập cảnh, số lần nhập cảnh, thời gian xét duyệt thị thực, phương tiện cho phép nhận cảnh. Vùng cụm màu cam liên quan đến thảo luận tư vấn lịch trình các điểm đến du lịch Bắc Bộ. Trong khi ‘word cloud’ giúp nhận ra bức tranh tổng thể về những từ phổ biến, trong khi phân tích ngữ nghĩa giúp chúng ta hiểu rõ hơn về ý nghĩa sâu xa và bối cảnh của những từ đó. Có thể có nhiều vấn đề hoặc trải nghiệm được nhắc đến liên quan đến Bắc Bộ, như thời tiết, ẩm thực, hoặc các vấn đề trong việc di chuyển giữa các địa điểm này, làm cho khu vực Bắc Bộ trở thành chủ đề chính trong các cuộc thảo luận. Điều này chỉ ra rằng, du khách có xu hướng chia sẻ nhiều hơn về trải nghiệm và cảm nhận của họ tại các địa điểm này, có thể vì lý do văn hóa, lịch sử, hoặc trải nghiệm cá nhân phong phú hơn. Bổ sung cho khía cạnh này, vùng cụm màu đỏ chỉ ra các vấn đề thắc mắc cần lời khuyên liên quan đến trải nghiệm văn hóa, món ăn, thời điểm du lịch, phương tiện đi lại, du lịch về đêm. Kết quả phân tích không chỉ làm sáng tỏ xu hướng cảm xúc và các mối quan tâm chính của du khách, mà còn cung cấp cơ sở dữ liệu quan trọng để cải thiện năng lực cạnh tranh của du lịch Việt Nam.

3. KẾT LUẬN

Kết quả phân tích bình luận trên Reddit cho thấy du khách quốc tế đặc biệt quan tâm đến các điểm du lịch tại miền Bắc Việt Nam, đặc biệt là thủ đô Hà Nội. Phần lớn các bình luận mang tính tích cực, phản ánh sự hài lòng với trải nghiệm du lịch tại đây. Kết quả này phù hợp với số liệu từ Tổng cục Thống kê quý I/2024, cho thấy Hà Nội cùng các địa

phương như Quảng Ninh, Khánh Hòa, và Thành phố Hồ Chí Minh đạt tăng trưởng ấn tượng về doanh thu và lượng khách [14]. Tuy nhiên, nghiên cứu cũng chỉ ra vấn đề về chính sách thị thực, đặc biệt là thời hạn lưu trú và số lần nhập cảnh, là mối bận tâm lớn của một số du khách. Mặc dù các bình luận tiêu cực chỉ chiếm tỷ lệ nhỏ, nhưng vấn đề này ảnh hưởng đáng kể đến trải nghiệm và sự thuận tiện của du khách. Các chính sách như miễn thị thực hoặc đơn giản hóa thủ tục có thể là giải pháp hiệu quả, học hỏi từ kinh nghiệm thành công của các nước trong việc thu hút khách du lịch. Chúng ta cũng thấy được du khách so sánh các điểm đến đang xảy ra. Để đánh giá việc Việt Nam có nên áp dụng chính sách tương tự hay không, chúng ta cần xem xét kỹ hơn những lợi ích và thách thức mà nó mang lại. Ngoài ra, cần thúc đẩy quảng bá các điểm đến ít được biết đến để phân bổ dòng khách đồng đều, giảm áp lực tại các điểm du lịch phổ biến, và hướng đến phát triển du lịch bền vững. Việc phân bổ này không chỉ giúp bảo vệ các tài nguyên du lịch mà còn tạo ra cơ hội phát triển kinh tế cho các khu vực chưa được khai thác, củng cố vị thế cạnh tranh của Việt Nam trong khu vực ASEAN. Bên cạnh đó, để giải quyết kịp thời những thắc mắc thường gặp của du khách liên quan đến các vấn đề như trải nghiệm văn hóa, ẩm thực địa phương, thời điểm du lịch lý tưởng, phương tiện di chuyển phù hợp và các hoạt động du lịch về đêm, các cơ quan quản lý du lịch và các bên liên quan cần chú trọng đến việc phát triển các nền tảng thông tin chính thống dễ tiếp cận và đáng tin cậy. Điều này có thể bao gồm việc cập nhật và cung cấp thông tin chi tiết qua các kênh trực tuyến như website du lịch, ứng dụng di động và các trang mạng xã hội. Trong tương lai, khuyến nghị các nghiên cứu tiếp theo tập trung vào việc đo lường tác động của những thay đổi chính sách, thị thực và phân tích chuyên sâu về các khu vực chưa được khai thác hiệu quả. Những cải tiến này sẽ là nền tảng để phát triển chiến lược du lịch toàn diện, nâng cao trải nghiệm du khách và tăng cường tính cạnh tranh của Việt Nam trên thị trường du lịch quốc tế.

TÀI LIỆU THAM KHẢO

- [1] Nguyễn Thị Tú Trinh, Nguyễn Hồng Đào, Khuru Ngọc Huyền (2018), *Nghiên cứu nhu cầu du lịch trải nghiệm của người dân thành phố Cần Thơ*, *Tạp chí Khoa học Đại học Cần Thơ*, 54(7).
- [2] Hoàng Thùy Linh, Ngô Thị Kim Liên (2020), *Quản lý, bảo tồn di sản văn hóa và sự kết nối, phát triển du lịch Việt Nam*, *Tạp chí Khoa học Đại học Mở Thành Phố Hồ Chí Minh*, 15(6).
- [3] Trương Trí Thông (2021), *Các nhân tố ảnh hưởng đến sự phát triển làng nghề phục vụ du lịch ở thành phố Phú Quốc tỉnh Kiên Giang*, *Tạp chí Khoa học Trường Đại học Sư phạm Thành phố Hồ Chí Minh*, 18(7).
- [4] Phạm Quốc Trí, Trịnh Thị Thu (2022), *Các nhân tố tác động đến quyết định lựa chọn điểm đến du lịch nông thôn dựa vào cộng đồng vùng Bắc Trung Bộ*, *Tạp chí Khoa học xã hội miền Trung*, 1(75).
- [5] Antoni Serra-Cantalalops, José Ramon-Cardona, Fabiana Salvi (2018), *The impact of positive emotional experiences on eWOM generation and loyalty*, *Spanish Journal of Marketing - ESIC*, 22(2).
- [6] Farshid Mirzaalian, Elizabeth Halpenny (2019), *Social media analytics in hospitality and tourism*, *Journal of Hospitality and Tourism Technology*, 10(4).
- [7] Ismail Shaheer, Neil Carr (2022), *Social representations of tourists' deviant behaviours: An analysis of Reddit comments*, *International Journal of Tourism Research*, 24(5).
- [8] Jenny Nilsson Vestola (2023), *Perceptions of climate change and tourism: A deep dive into the forums of Reddit*, *Tourism and Hospitality Research*.
- [9] Jurui Zhang (2019), *What's yours is mine: exploring customer voice on Airbnb using text-mining approaches*, *Journal of Consumer Marketing*, 36(5).
- [10] Khaldoon Nusair, Irfan Butt, S.R. Nikhashemi (2019), *A bibliometric analysis of social media in hospitality and tourism research*, *International Journal of Contemporary Hospitality Management*, 31(7).
- [11] Ming-Lang Tseng, Kuo-Jui Wu, Chia-Hao Lee, Ming K. Lim, Tat-Dat Bui, Chih-Cheng Chen (2018), *Assessing sustainable tourism in Vietnam: A hierarchical structure approach*, *Journal of Cleaner Production*, 195.
- [12] Thịnh Gia Hoang, Anh Thi Van Nguyen, Huan Huu Nguyen, Vu Minh Ngo (2023), *Current obstacles, contemporary practices, and potential solutions for recovery in Vietnam tourism after the COVID-19 pandemic: Tour operators' perspectives*, *Tourism and Hospitality Research*, 24(3).
- [13] Cục du lịch Quốc gia Việt Nam (2024), *Cơ sở dữ liệu trực tuyến*, <https://thongke.tourism.vn/index.php/news/items/221>, ngày truy cập: 31-8-2024.
- [14] Tổng cục Thống kê (2024), *Cơ sở dữ liệu trực tuyến*, <https://www.gso.gov.vn/du-lieu-va-so-lieu-thong-ke/2024/04/doanh-thu-dich-vu-luu-tru-an-uong-va-lu-hanh-tang-cao-trong-quy-i-2024>, ngày truy cập: 31-8-2024.
- [15] VinaCapital (2024), *Cơ sở dữ liệu trực tuyến*, <https://vinacapital.com/wp-content/uploads/2024/06/VinaCapital-Insights-Du-lich-tang-truong-thuc-day-nen-kinh-te-Viet-Nam.pdf>, ngày truy cập: 31-8-2024.