[Tạp chí Khoa học Dữ liệu SMU](https://scholar.smu.edu/datasciencereview)

[Tập 8](https://scholar.smu.edu/datasciencereview/vol8)

[Số 1 Mùa Xuân 2024](https://scholar.smu.edu/datasciencereview/vol8/iss1)

[Bài viết 4](https://scholar.smu.edu/datasciencereview/vol8/iss1/4)

Phân Tích Gợi Ý Trò Chơi Sử Dụng Hồ Sơ và Đánh Giá Trên Steam

Robert Blue

*Đại học Nam Methodist, robert.blue101@gmail.com*

Luis Garcia

*Đại học Nam Methodist, garcia.luis@mail.smu.edu*

Jacob Turner

*Đại học Nam Methodist, jturner@mail.smu.edu*

Theo dõi bài viết này và các tác phẩm bổ sung tại: https://scholar.smu.edu/datasciencereview Một phần của Tài nguyên Khoa học Dữ liệu

Trích dẫn Đề xuất

Blue, Robert; Garcia, Luis; và Turner, Jacob () "Phân Tích Gợi Ý Trò Chơi Sử Dụng Hồ Sơ và Đánh Giá Trên Steam," Tạp chí Khoa học Dữ liệu SMU: Tập 8: Số 1, Bài viết 4.

Có sẵn tại: https://scholar.smu.edu/datasciencereview/vol8/iss1/4

Bài viết này được cung cấp miễn phí và truy cập mở bởi SMU Scholar. Nó đã được chấp nhận để đưa vào Tạp chí Khoa học Dữ liệu SMU bởi một quản trị viên được ủy quyền của SMU Scholar. Để biết thêm thông tin, vui lòng truy cập http://digitalrepository.smu.edu.

**Phân Tích Gợi Ý Trò Chơi Sử Dụng Hồ Sơ và Đánh Giá Trên Steam**

Robert Blue, Luis Garcia, Jacob Andrew Turner, Ph.D. Thạc sĩ Khoa học Dữ liệu, Đại học Nam Methodist Dallas, TX 75275 USA

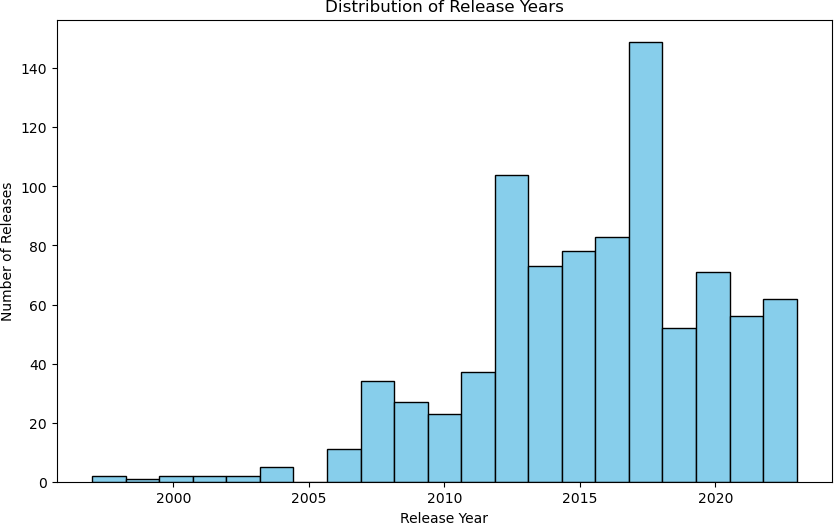
[rblue@smu.edu garcia.luis@smu.edu](mailto:rblue@smu.edu)

**Tóm tắt. Các studio trò chơi nhỏ hơn gặp bất lợi khi muốn sản phẩm của họ được người dùng chú ý. Nghiên cứu này nhằm cung cấp cái nhìn về cách các hệ thống gợi ý hoạt động để các studio nhỏ hơn có thể làm cho trò chơi của họ được chú ý trên Steam. Steam là một trong những dịch vụ phân phối trò chơi video lớn nhất và họ có một hệ thống gợi ý giúp quảng bá trò chơi đến người dùng. Nghiên cứu này đã sử dụng thông tin người dùng như số lượng trò chơi đã chơi, loại trò chơi và số giờ chơi để tạo ra các hệ thống gợi ý nhằm xác định các đặc điểm của trò chơi đang thúc đẩy các gợi ý.**

# Giới thiệu

Thị trường trò chơi video dự kiến sẽ có sự tăng trưởng bùng nổ trong 4 năm tới.

Theo Statistica, vào cuối năm 2023, thị trường dự kiến sẽ đạt 249,60 tỷ USD. Đến năm 2028, giá trị thị trường này dự kiến sẽ tăng lên 389,70 tỷ USD. (Statistica, 2023). Hình 1 cho thấy sự tăng trưởng bùng nổ trong 5 năm qua của ngành công nghiệp trò chơi với tỷ lệ tăng trưởng hàng năm trên 9% mỗi năm.



**Hình 1. Biểu đồ histogram về số lượng trò chơi được phát hành từ năm 1995 đến nay. Chu kỳ phát hành trò chơi đã bùng nổ kể từ những năm 2010 với sự nghiêng nặng về phần sau của dòng thời gian.**

Hình 2 làm nổi bật mười lăm studio hàng đầu đã phát hành trò chơi trên Steam kể từ năm 2019 với số lượng phát hành, tổng số chủ sở hữu và tổng số đánh giá. Chỉ có một studio được phân loại là "indie", Endnight Game Studio, nằm trong top mười và danh sách chỉ có bốn studio indie tổng cộng bao gồm Re-Logic, ConcernedApe và Klei Entertainment trong top mười lăm. Trong số đó, không có nhà phát hành indie nào bán được hơn 25.000.000 đơn vị trong khi nhà phát hành hàng đầu, Valve, đã bán được 143.000.000 đơn vị.



**Hình 2. 15 studio hàng đầu hiện có trên Steam.**

Steam đã nổi lên như một nền tảng thống trị cho game thủ, với hệ thống gợi ý đóng vai trò quan trọng trong việc hướng người dùng đến những sở thích trò chơi tiềm năng. Nghiên cứu này đi sâu vào cơ chế và những phức tạp của việc xây dựng một hệ thống gợi ý, được thiết kế tỉ mỉ để xem xét tất cả các khía cạnh của hồ sơ người dùng trên Steam. Số lượng trò chơi sở hữu hoặc tham gia trong các thể loại cụ thể của một người dùng là một chỉ số cơ bản, giúp hiểu rõ sở thích chơi game của họ. Thông tin này chứng tỏ rất quan trọng trong việc dự đoán những sở thích tiếp theo, điều này đóng vai trò then chốt trong việc hệ thống gợi ý chọn lọc các trò chơi nào. Thời gian chơi tích lũy của một cá nhân là một minh chứng không thể chối cãi cho sức hấp dẫn của một trò chơi và độ sâu của sự tham gia của người dùng. Việc chú ý đến chỉ số này đảm bảo rằng hệ thống gợi ý công nhận và tích hợp sự nhiệt tình mà người dùng thể hiện đối với các thể loại hoặc tiêu đề trò chơi cụ thể.

Ngoài khía cạnh gameplay, người dùng thể hiện cảm xúc của họ về các trò chơi thông qua các đánh giá. Một phân tích tỉ mỉ về các thành phần ngôn ngữ và cảm xúc tổng thể trong các đánh giá này có thể làm sáng tỏ những khía cạnh phức tạp của sự hài lòng hoặc không hài lòng của người dùng. Phân tích này cũng có thể được áp dụng cho thời gian chơi và sử dụng phân tích cảm xúc trên chính đánh giá để xác định tính hợp lệ của đánh giá đó.

Một đánh giá chỉ có một vài giờ chơi sẽ không hữu ích bằng một đánh giá có hàng trăm giờ chơi. Tương tự, một đánh giá chỉ phê bình hoặc khen ngợi một phần nhỏ của trò chơi sẽ không có giá trị bằng một đánh giá đi sâu vào các cơ chế hoặc khía cạnh cụ thể mà là tích cực hoặc tiêu cực.

Mặc dù sức mạnh không thể phủ nhận của hệ thống gợi ý hiện có của Steam, nhưng nó có những thiên lệch rõ ràng. Thuật toán hiện tại có xu hướng thiên về các studio lớn và đã nổi tiếng cũng như các tiêu đề được đánh giá cao. Do đó, điều này thường làm lu mờ những đóng góp từ các studio nhỏ hơn hoặc các nhà phát triển độc lập, từ đó làm giảm sự hiện diện và khả năng hiển thị của họ trên nền tảng. Trong khi chiến lược này chứng tỏ hiệu quả trong việc giảm thiểu sự xuất hiện của "shovelware", hay các trò chơi được biết đến nhiều hơn vì số lượng hơn là giá trị - nó cũng dấy lên những lo ngại. Có nguy cơ tiềm ẩn là vô tình hạn chế sự tiếp xúc cho các nội dung chất lượng cao xuất phát từ các thực thể không có sự hỗ trợ lớn hơn hoặc ngân sách tiếp thị mà thường thấy ở các studio và tiêu đề chính thống hơn.

The importance of recommendations systems is not solely a feature of the gaming market. Across diverse sectors, from digital literature to movie streaming platforms, personalized recommendations exert a profound impact on product uptake. A well- structured recommendation infrastructure not only augments product prominence but also refines the overarching user interaction. Personalized consumer experiences have the potential to uplift sales figures by an estimated 20%. An overwhelming majority, approximately 80% of consumers, demonstrate a preference for e-commerce platforms that offer customized and personalized recommendation experiences (Sahin 2023). In the gaming sector, enthusiasts are inclined to write comprehensive and extended reviews for titles they harbor deep-rooted affinities for and into which they have invested considerable resources, whether it is time, money, or both. In fact, comprehensive reviews are often an outcome of elevated game acquisition costs (Lin, Bezemer, Zou, Hassan 2019).

March 2020 marked a significant juncture with Valve introducing a re-envisioned Interactive Recommendation engine. Using a foundation built on a robust machine learning infrastructure, the system gleans insights from the playtime trajectories of a vast user base on Steam. Rather than using a traditional tagging system, the emphasis is instead focused on player behavior and established patterns, forming the foundation for game recommendations. This new system still shares the same pitfalls as its predecessor, particularly a bias towards games with substantial marketing or widespread recognition (Robertson 2019). Notably, this overhaul appeared, to a significant extent, as a redressal to apprehensions voiced by smaller studios, who felt marginalized during major steam promotional periods like the Steam Summer Sale (Grayson 2019).

Given the current situation and identified gaps, this research paper hopes to answer a simple question: How do large studios get their games recommended on Steam over smaller studios?

# Literature Review

The Literature Review focuses on three areas: Analysis of Game Reviews, NLP and sentiment analysis methods used on reviews, and matrix factorization.

### Analysis of Game Reviews

Game reviews hold valuable information that can help decipher why a game was liked or disliked by a reviewer (Lin et al. 2019). Prior work analyzing words in reviews yielded associations between negative reviews after many playing hours and bad patches as well as negative reviews with few playing hours and severe bugs or bad game design (Lin et al. 2019). An analysis of player behavior concluded that gamers have little patience when it comes to faulty servers (Chambers et al. 2005), further supporting the previous association between short play hours and bad game design. One contributing factor for effectively addressing bad game designs are the type of update strategies that the developers use, of which updates that happen less than weeks apart from one another tend to have more back-to-back updates (Lin, Bezemer, Hassan 2017). Another finding was that negative reviews are usually posted with less than half of the playing time than those of positive reviews (Lin et al. 2019). Lastly, it would seem like price affects the user’s willingness to rate the game differently depending on the tags (Toy et al. 2023)

An analysis of the reviews of games, considering the genre, yielded the results that game reviews vary in length and playtime when reviewed depending on the genre of the game (Guzsvinecz et al. 2023). Another observation was that positive reviews were more prevalent during Early-Release windows so during these times someone can expect the reviews to have positive language (Lin et al. 2018). Building recommendation systems while considering genres allow for recommendations to be better aligned with user’s interests (Andersson, J 2022). One can still create simple recommendation systems without knowing the genre and only using data regarding what games the target user is playing (R.R 2021).

### NLP

While previous game review analysis provides trends to look out for when scrapping Steam reviews, deeper dives into the reviews can be done with NLP and Sentiment Analysis for better recommendations just like Gameopedia has been doing with their sentiment analysis tool (Karthikeyan, K. 2021). Another way that NLP methods have been used are to identify nouns in reviews and pair them with adjectives to identify patterns in reviews (Zhu et al. 2015). The analysis of game reviews also used sentiment analysis with classification of words being done in accordance with the NRC Emotion Lexicon which is a list of English words and their association with sentiment and found that the intensity of like and dislike in the reviews varied depending on the genre (Guzsvinecz et al. 2023). Reviews were portioned into equal lengths and sentiment per portion was taken to analyze how sentiment differed from starting a review to ending it (Guzsvinecz et al. 2023). Incorporating sentiment of reviews when making recommendations improves the recommendation accuracy (Roy et al. 2021).

One thing to consider when analyzing sentiment in reviews is that some reviews may be fake reviews, luckily there exist ways to identify fake reviews and fake reviewers which can be used to deduce legitimate reviews (Liu, B 2012). Sentiment polarity categorization, scoring a word from extremely negative to incredibly positive, is an issue that comes with sentiment analysis but a study on Amazon reviews proposed a

three-phase process that helped researchers yield successful results for sentence and review-level sentiment analysis (Fang et al. 2015).

### Collaborative Filtering and Matrix Factorization

Lọc cộng tác (CF) là một kỹ thuật cơ bản được sử dụng bởi các hệ thống gợi ý quy mô lớn. CF phân tích dữ liệu tương tác, chẳng hạn như đánh giá của người dùng và độ phổ biến của mặt hàng thông qua phản hồi tích cực, để kết nối các mẫu giữa người dùng. Phân tích này giúp xây dựng các mô hình dự đoán gợi ý các mặt hàng mà một người dùng cụ thể đã thích trong quá khứ. Quá trình này bao gồm hai phương pháp chính: các phương pháp lân cận, tập trung vào các tương tác trực tiếp giữa người dùng và mặt hàng, và các kỹ thuật phân tích ma trận, tương tự như phân tích thành phần chính, giảm độ phức tạp của dữ liệu bằng cách suy diễn các yếu tố tiềm ẩn mô tả các tương tác giữa người dùng và mặt hàng (Batra et al. 2023). Tuy nhiên, CF dễ bị ảnh hưởng bởi một số thiên kiến; các mặt hàng có mức độ tương tác cao hơn thường nhận được sự chú ý không tương xứng, có thể làm lu mờ các tùy chọn ít phổ biến hơn. Thiên kiến về độ phổ biến này có thể khởi đầu một vòng phản hồi, nơi các mặt hàng phổ biến được đẩy lên đầu hàng đợi gợi ý. Thêm vào đó, khi hệ thống mở rộng để đáp ứng nhiều người dùng và mặt hàng hơn, độ phức tạp của các mối quan hệ này tăng lên, thách thức khả năng mở rộng của mô hình. Để có cái nhìn sâu hơn về lọc cộng tác, bao gồm phân tích ma trận, người đọc được hướng dẫn đến các tài nguyên bổ sung như hướng dẫn của Real Python về lọc cộng tác (Ajisaria 2021) và tài liệu học thuật liên quan. Nền tảng này có thể nâng cao hiểu biết về các đầu vào và đầu ra dữ liệu thiết yếu cho CF, cung cấp cái nhìn rõ ràng hơn về cơ chế hoạt động của nó.

Để vượt qua những hạn chế này, các công ty thường sử dụng một kỹ thuật gọi là Phân tích Ma trận. Gã khổng lồ streaming Netflix sử dụng thuật toán này cũng như các công ty lớn khác như Amazon và Spotify. Phân tích ma trận hiệu quả trong các hệ thống gợi ý vì nhiều lý do. Một trong những lý do chính để chọn phân tích ma trận là khả năng xác định các yếu tố tiềm ẩn. Một yếu tố tiềm ẩn là một biến chỉ có thể được quan sát thông qua các phương pháp toán học chứ không phải chỉ đơn thuần quan sát dữ liệu. Điều này rất quan trọng vì thuật toán có thể phát hiện các mối tương quan có thể bị bỏ qua, điều này sẽ xây dựng một tương tác người dùng-mặt hàng mạnh mẽ hơn. Phân tích ma trận cũng giải quyết vấn đề với dữ liệu thưa thớt bằng cách "lấp đầy" các mối tương quan bị thiếu bằng những yếu tố tiềm ẩn này. Điều này cung cấp một hệ thống gợi ý vững chắc dựa trên sở thích của người dùng. Phân tích ma trận cũng có thể mở rộng bằng cách sử dụng các hệ thống như ALS (bình phương tối thiểu luân phiên) có thể chạy song song trên nhiều môi trường khác nhau, điều này có thể tối ưu hóa đầu ra.

# Phương pháp

Chúng tôi sẽ sử dụng các công trình trước đây đã được thực hiện trong hệ thống gợi ý, đặc biệt là công trình của Tiến sĩ Julian McAuley từ Đại học California, San Diego (Wan, McAuley 2018). Chúng tôi cũng sẽ sử dụng API Steam để lấy thông tin người dùng và trò chơi.

Dữ liệu cho nghiên cứu này được lấy từ hai API khác nhau vì dữ liệu cho người dùng và trò chơi có thể truy cập thông qua một API duy nhất. API Steamworks Web chứa các yêu cầu được sử dụng để lấy danh sách trò chơi cùng với các khóa đặc biệt của chúng, lấy thông tin về số lượng trò chơi mà một người chơi sở hữu và tần suất họ chơi những trò chơi đó, cũng như danh sách các trò chơi đã được chơi gần đây. API gợi ý của Steam chứa một yêu cầu để lấy dữ liệu cho các đánh giá.

Các thẻ cho các trò chơi được trích xuất thủ công bằng cách sử dụng requests.get và BeautifulSoup để thu thập dữ liệu từ trang web tài liệu của Steam. Việc làm sạch dữ liệu thêm được thực hiện trên kết quả từ trang để chúng ta có thể trích xuất các thẻ cho thể loại và thể loại phụ. Phương thức json.loads sẽ được sử dụng khi gọi cả hai API.

Trước khi bất kỳ API nào có thể được sử dụng, Steam yêu cầu phải tạo một tài khoản Steam và tạo một khóa API Web Steam. Tài khoản Steam có các ID steam duy nhất liên kết các đánh giá với người dùng và, nếu người dùng có thể nhìn thấy, cho phép API Steamworks trích xuất thông tin về xu hướng chơi game của người dùng.

Hành động API đầu tiên cần thực hiện là trích xuất tên của các trò chơi cũng như ID trò chơi được lưu dưới dạng appid. Những thông tin này được trích xuất thông qua giao diện ISteamApps của Steamworks API bằng phương thức GetAppList, không cần bao gồm tham số bổ sung nào. Kết quả của truy vấn sẽ được lưu trên một bảng Nx2 để tên của các trò chơi có thể được liên kết sau này thông qua một khóa thứ cấp với bất kỳ bảng nào trong tương lai. Từ đây, trang web của các trò chơi sẽ cần được thu thập thông tin để lấy các thẻ, vì không có API nào cung cấp thông tin này. Các thẻ từ danh mục thể loại và tiểu thể loại sẽ được so sánh và chỉ những thẻ đó sẽ được giữ lại.

Điều thứ hai cần làm với các API là trích xuất các đánh giá từ các trò chơi. Các trò chơi được truy vấn bằng phương thức getreviews bằng cách cung cấp appid và tham số ngôn ngữ để chỉ nhận được các đánh giá bằng tiếng Anh. Lời gọi trả về số lượng đánh giá, steamid của người dùng, thời gian chơi tại thời điểm đánh giá, thời gian tạo đánh giá, liệu đánh giá có được thực hiện trong giai đoạn Early Access hay không, nội dung đánh giá và nhiều trường khác sẽ được xem xét khi đánh giá tính hợp lệ của các đánh giá. Tất cả các trò chơi được trích xuất từ các cuộc gọi API trước đó sẽ được tra cứu và chỉ những trò chơi có hơn 100 đánh giá mới được giữ lại để phân tích. Từ những trò chơi này, chỉ những người dùng có num\_games\_owned lớn hơn 0 sẽ được xem xét vì một giá trị 0 trong trường này cho thấy rằng mặc dù chúng tôi có thể thấy đánh giá của họ, nhưng chúng tôi không có quyền truy cập vào thư viện của họ, và đây không phải là những người dùng mà chúng tôi muốn đưa vào Ma trận Phân tích.

Bây giờ, khi các đánh giá trò chơi, thẻ trò chơi và người dùng có thư viện công khai đã được trích xuất, thông tin thư viện của người dùng có thể được truy vấn. Sử dụng phương thức GetOwnedGames của giao diện IPlayerService, các trò chơi của một người dùng có thể được lấy bằng cách cung cấp steamid duy nhất của người dùng, được lấy từ các đánh giá. Điều này trả về tổng thời gian chơi cho một trò chơi, thời gian chơi trong hai tuần qua, cũng như lần cuối cùng một trò chơi được chơi, và nó bao gồm steam\_appid. Thông tin về lần cuối cùng chơi được trả về ở định dạng Epoch và phải được chuyển đổi thành datetime để có thể hiểu được, và thời gian chơi được chuyển đổi từ phút sang giờ để rõ ràng hơn.

Truy vấn cho các đánh giá bỏ qua các đánh giá không liên quan và trả về một điểm số cho đánh giá dựa trên tính hữu ích, điều này có thể được sử dụng để xác định mức độ suy nghĩ của các đánh giá đó. Cường độ của các đánh giá sẽ được phân tích thông qua phân tích cảm xúc để tăng thêm trọng số cho chúng. Các phương pháp thống kê sẽ được áp dụng cho các đánh giá để đảm bảo rằng chúng có ý nghĩa thống kê sau khi xác định phân phối của các đánh giá.

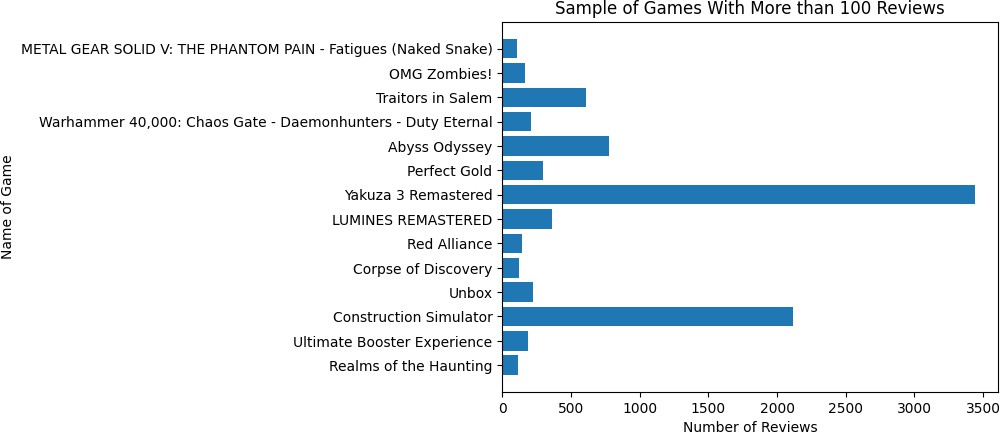
theo như để đáp ứng các giả định của nhiều mô hình có thể được sử dụng (Dror et. al 2020).

Trong quá trình phát triển hệ thống gợi ý của chúng tôi, cả hai phương pháp dựa trên người dùng và dựa trên mục đã được sử dụng với dữ liệu truy cập qua API. Cụ thể, đối với hệ thống dựa trên mục, "Yếu tố Gắn kết" được xác định bởi tổng thời gian người dùng dành cho việc chơi game, trong khi "Yếu tố Phổ biến" dựa trên số lượng đánh giá tích cực mà mỗi trò chơi nhận được. Dữ liệu này được tổ chức thành một ma trận dày có cấu trúc, trong đó các hàng đại diện cho từng trò chơi và các cột bao gồm các chỉ số chính này, cùng với tổng số lượt tải xuống giúp phân loại và phân cụm. Định dạng ma trận này tạo điều kiện cho việc áp dụng các phép đo khoảng cách Euclidean để xác định và nhóm các trò chơi có thuộc tính tương tự dựa trên mức độ gắn kết và độ phổ biến của chúng. Về cơ bản, ma trận này không chỉ đơn thuần là một tập hợp dữ liệu thô mà là một thực thể đã được chuyển đổi phục vụ như một ma trận tương đồng, cho phép phân cụm hiệu quả các trò chơi có đặc điểm tương đồng. Cách tiếp cận này nhấn mạnh sự tương tác tinh vi giữa các loại dữ liệu khác nhau trong hệ thống của chúng tôi và làm nổi bật tính nghiêm ngặt trong phương pháp phân đoạn và phân tích các thuộc tính của trò chơi.

Tương tự, hệ thống dựa trên người dùng đã tận dụng các yếu tố tiềm ẩn về Sự tham gia và Độ phổ biến, được rút ra từ hành vi của từng người dùng. Thay vì sử dụng tổng thời gian chơi và đánh giá, hệ thống này đã phân tích hơn 800 thư viện Steam của người dùng và chỉ tập trung vào thời gian chơi của riêng họ cùng với tổng số đánh giá tích cực mà mỗi trò chơi nhận được. Phân tích chi tiết này đã giúp chúng tôi phát hiện ra các yếu tố tiềm ẩn đóng vai trò là các mẫu và sở thích chính trong sở thích chơi game của từng cá nhân, từ đó cho phép chúng tôi tạo ra một mô hình dự đoán. Cụ thể, chúng tôi đã sử dụng một mô hình áp dụng phân tích giá trị riêng (SVD), một kỹ thuật phân tích ma trận phân rã dữ liệu thành các vector và giá trị riêng. Phương pháp này xác định và định lượng hiệu quả các cấu trúc tiềm ẩn trong tương tác giữa người dùng và trò chơi, và xây dựng một dự đoán dựa trên các yếu tố này. Chúng tôi sau đó sử dụng thời gian chơi dự đoán mà mô hình xuất ra để tạo ra một hệ thống gợi ý mạnh mẽ. Cách tiếp cận dự đoán này đã sử dụng sức mạnh của phân tích ma trận để cung cấp các gợi ý trò chơi cá nhân hóa, nâng cao sự tham gia của người dùng bằng cách điều chỉnh các gợi ý phù hợp với sở thích và thói quen chơi game của từng cá nhân.

# Kết quả

Mặc dù chúng tôi đã giới hạn phạm vi các trò chơi được truy vấn chỉ cho những trò chơi có hơn 100 đánh giá, mức độ tương tác của những trò chơi này rất khác nhau. Hình 3 dưới đây nhấn mạnh thực tế rằng mặc dù chúng tôi đã có một lựa chọn có chủ đích để chọn những trò chơi nhận được phản hồi từ người dùng, nhưng độ lớn của nó không đồng nhất.



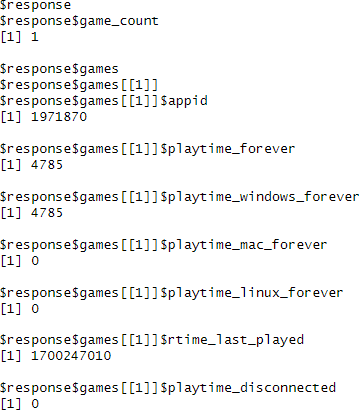
**Hình 3. Mặc dù các appid có hơn 100 đánh giá, nhưng số lượng đánh giá thực tế lại khác nhau.**

Các trò chơi trên Steam được gán thẻ với thể loại cấp cao, thể loại, tiểu thể loại và nhiều loại thẻ đa dạng khác. Hình 4 dưới đây cho thấy 10 thẻ thể loại cấp cao, nhiều trong số đó là các thẻ có mặt trong ma trận mục được tạo ra để đưa ra các gợi ý dựa trên mục.



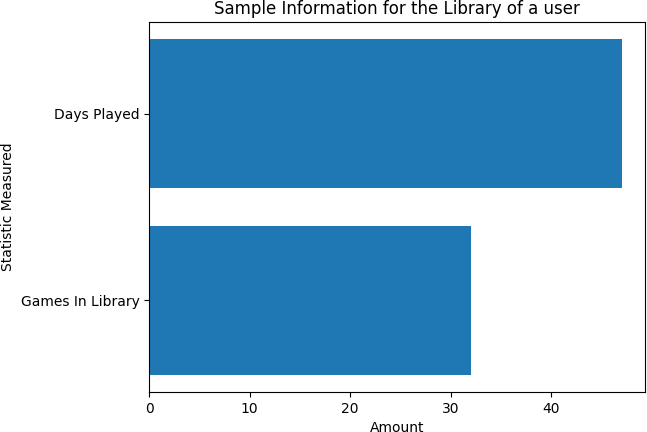
**Hình 4. 10 trong số 423 thẻ có thể có của một trò chơi.**

API Steam cho phép người dùng điều tra thư viện của một người dùng Steam và xem họ đang chơi trò chơi nào và bao nhiêu, nhưng chỉ khi họ chia sẻ thông tin đó công khai. Hình 5 cho thấy thời gian chơi của một trò chơi có thể được lấy từ thư viện của người dùng bằng cách cung cấp trò chơi mà bạn quan tâm. Quy trình từ trò chơi đến thời gian chơi này đã được sử dụng để lấy tổng thời gian chơi mà một trò chơi có trên mẫu người dùng đã được truy vấn thư viện Steam của họ thông qua API Steam như một phần của nghiên cứu này.

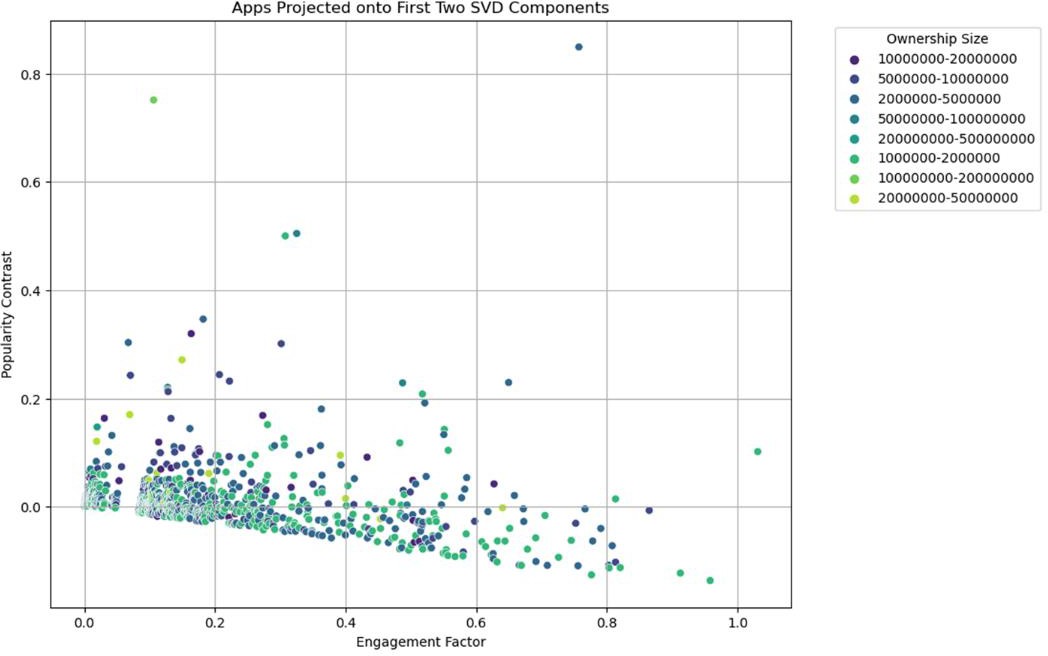


**Hình 5. Hình này cho thấy đầu ra trong R của kết quả mà chúng tôi nhận được khi truy vấn một appid cụ thể, chúng tôi có thể lấy tất cả thông tin người dùng ở trên liên quan đến một trò chơi mà họ sở hữu.**

Nhóm đối tượng quan tâm cho nghiên cứu này là những người dùng có một số lượng trò chơi và thời gian chơi hợp lý. Hình 6 cho thấy phân phối của các trò chơi và thời gian chơi mà họ có thể có. Mặc dù Hình 5 ở trên cho thấy số lượng trò chơi mà một người dùng có được cung cấp bởi một yêu cầu của SteamAPI, trường này đã bị bỏ qua vì những trò chơi có 0 thời gian chơi vẫn được tính vào số lượng trò chơi của họ.



**Hình 6. Hình này cho thấy chúng ta có thể tìm thấy những người dùng có hồ sơ công khai và số lượng trò chơi đã chơi đáng kể cũng như thời gian chơi những trò chơi đó khá nhiều. Đây là những người dùng mà chúng ta muốn nhắm đến.**



**Hình 7. Hình này cho thấy hai yếu tố tiềm ẩn, yếu tố tương tác và nội dung phổ biến, mà**

được rút ra sau khi thực hiện phân tích ma trận trên ma trận mục chứa thông tin trò chơi. Mỗi chấm là một trò chơi, và các chấm được mã màu dựa trên cụm Kích thước Sở hữu mà nó thuộc về để giúp hình dung cách các trò chơi tương tự với cùng số lượt tải xuống. Bằng cách sử dụng các giá trị của các yếu tố tiềm ẩn, khoảng cách có thể được rút ra giữa bất kỳ trò chơi nào.

Từ các trò chơi được hiển thị trong Hình 7, 'Counter-Strike' được sử dụng làm cơ sở và 5 trò chơi gần nhất với nó đã được xác định. Bảng 1 dưới đây cho thấy các trò chơi có khoảng cách Euclidian thấp nhất đến 'Counter-Strike' và khoảng cách đó là gì. Với việc thời gian chơi và số lượng đánh giá tích cực là các yếu tố tiềm ẩn, những gợi ý này được thúc đẩy bởi sự tương tác của người dùng thay vì sự tương đồng giữa các mục, đặc biệt là vì phần tiếp theo của nó, 'Counter-Strike 2', không phải là một gợi ý.

**Bảng 1. Bảng này cho thấy 5 trò chơi có khoảng cách Euclidean thấp nhất từ ‘Counter-**

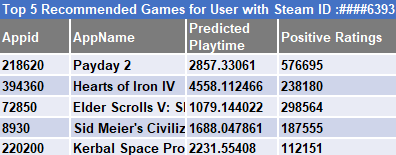
Độ dài Euclid được tính toán dựa trên các giá trị Hệ số Giao tiếp và Nội dung Phổ biến từ Hình F làm đầu vào.

|  |  |  |
| --- | --- | --- |
| mã ứng dụng | tên | Khoảng cách |
| 400 | Cổng | .002737 |
| 320 | Half-Life 2: Deathmatch | .018651 |
| 620 | Portal 2 | .033383 |
| 108600 | Dự án Zomboid | .057210 |
| 272060 | Serena | .115725 |

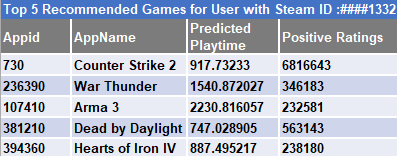
Trong khi các khuyến nghị trước đó, các khuyến nghị dựa trên mặt hàng, được đưa ra bằng cách lấy dữ liệu của một trò chơi và so sánh nó với các trò chơi khác dựa trên mức độ tương đồng về tổng thời gian chơi và đánh giá tích cực của trò chơi, các khuyến nghị dựa trên người dùng tận dụng dữ liệu người dùng để đưa ra dự đoán. Một ma trận người-hàng được tạo ra dựa trên thống kê thời gian chơi của người dùng. Mỗi hàng trong ma trận là một người dùng cụ thể và mỗi cột trong ma trận là một trò chơi cụ thể. Các mục trong ma trận là tổng thời gian chơi của người dùng đó cho một trò chơi. Kết quả của việc thực hiện phân tích ma trận trên điều này là hai ma trận mới, một ma trận chứa các yếu tố tiềm ẩn cho người dùng và một ma trận chứa các yếu tố tiềm ẩn cho các trò chơi. Những ma trận này được sử dụng để tạo ra thời gian chơi dự đoán cho các trò chơi mà người dùng chưa chơi. Sử dụng thời gian chơi dự đoán cụ thể cho người dùng mà SVD đã giúp tạo ra và các đánh giá tích cực cho một trò chơi, người dùng đã nhận được các khuyến nghị phù hợp. Bảng 2 và Bảng 3 cho thấy Heart of Iron IV là một khuyến nghị nhưng thời gian chơi dự đoán thực tế cho nó khác nhau vì thống kê thời gian chơi của người dùng khác nhau. Rõ ràng rằng sự phổ biến của một trò chơi đang thúc đẩy các khuyến nghị vì hầu hết các trò chơi được khuyến nghị đều từng được đánh giá cao hoặc có một cộng đồng người chơi tích cực.

cơ sở, nhưng các trò chơi ở đây có mức độ phổ biến khác nhau so với những trò chơi từ mô hình trước.

**Bảng 2. Bảng này cho thấy các trò chơi được khuyến nghị, được chỉ định bởi appid, mà người dùng ####6393 nhận được. Những khuyến nghị này bị ảnh hưởng mạnh mẽ bởi mức độ chơi của một trò chơi và mức độ tích cực của các đánh giá cho trò chơi đó. Một quan sát đáng chú ý là trò chơi ở hàng 2 có ít đánh giá tích cực hơn và thời gian chơi dự đoán nhiều hơn so với các trò chơi ở hàng 1 và hàng 3.**



**Bảng 3. Bảng này cho thấy các trò chơi được khuyến nghị, được chỉ định bởi appid, mà người dùng ####1332 nhận được. So sánh với Bảng 2, trò chơi có appid 394360 được thấy trong cả hai bảng nhưng thời gian chơi dự đoán và xếp hạng của khuyến nghị khác nhau giữa chúng.**



# Thảo luận

Phân tích khám phá hiệu quả của một hệ thống gợi ý so với các cơ chế hiện có của Steam, tập trung vào việc đại diện cho các studio indie. Nó đặt ra câu hỏi liệu các công ty được giới thiệu trên Steam có nên sử dụng thông tin công khai để nhắm đến khách hàng tiềm năng hay không, cân nhắc giữa rủi ro và lợi ích của các chiến lược như vậy. Nghiên cứu cũng nêu bật những phát hiện độc đáo và những thách thức gặp phải, đặc biệt là những khó khăn tính toán trong việc xử lý danh mục trò chơi rộng lớn và các đánh giá của người dùng trên Steam. Với gần 200.000 trò chơi và một số trò chơi có hơn 100.000 đánh giá, nhiệm vụ này thật sự khó khăn và tốn thời gian.

Một quyết định có chủ ý đã được đưa ra để loại trừ những người dùng có thư viện trò chơi hoặc thời gian chơi hạn chế, thừa nhận rằng lựa chọn này đã bỏ qua một phân khúc của cộng đồng Steam. Nghiên cứu đã gặp phải những hạn chế trong việc truy vấn dữ liệu, đặc biệt là với giới hạn của API Steam chỉ hiển thị 100 đánh giá tại một thời điểm, cần phải thực hiện nhiều yêu cầu. Phương pháp nghiên cứu đã sử dụng một kích thước mẫu tương đối nhỏ—860 trò chơi và thông tin thư viện của 1.000 người dùng—được coi là đủ để hiểu và phát triển động cơ gợi ý mặc dù chỉ đại diện một phần nhỏ cho hệ sinh thái Steam rộng lớn.

Tập trung chủ yếu vào người dùng có thời gian chơi game dài và bộ sưu tập game lớn, cụ thể là lọc các trò chơi có 100 đánh giá trở lên để loại trừ các trò chơi ngách hoặc chỉ có nội dung tải xuống. Cách tiếp cận này đảm bảo rằng nghiên cứu tập trung vào các trò chơi trong thể loại indie thay vì một tập hợp trò chơi rộng hơn, ít được xác định hơn. Các hạn chế mở rộng đến việc chỉ xem xét các trò chơi có nhiều đánh giá, vô tình loại trừ các bản phát hành mới và yêu cầu nỗ lực thủ công để thu thập thẻ trò chơi do hạn chế của API. Thêm vào đó, chỉ có các đánh giá bằng tiếng Anh được phân tích, dẫn đến việc loại trừ dữ liệu đáng kể.

## Đạo đức

Từ góc độ đạo đức, nghiên cứu này nhằm hỗ trợ các studio indie bị các thực thể lớn hơn che khuất và duy trì quyền riêng tư của người dùng thông qua việc bỏ qua một phần Steam ID. Biện pháp này tôn trọng quyền riêng tư của người dùng trong khi thừa nhận tính công khai của dữ liệu như vậy. Có nhiều mối quan tâm đạo đức liên quan đến việc có thể xác định người dùng có thể bị thu hút bởi một sản phẩm.

Như đã được chỉ ra với hệ thống gợi ý dựa trên mặt hàng, được tạo ra từ việc phân tích ma trận thông tin trò chơi với 19 đặc điểm được phân tách thành các yếu tố tiềm ẩn, các mặt hàng có thể được sử dụng để gợi ý các mặt hàng khác, vì vậy bằng cách sử dụng các đầu vào khác nhau, người ta có thể dễ dàng sử dụng người dùng để gợi ý cho những người dùng khác. Thông tin có sẵn qua API của Steam cho phép các công ty có cơ hội tìm kiếm những người dùng chưa tiếp xúc với sản phẩm của họ và tìm thấy trang Steam của họ. Các công ty có thể dễ dàng tìm thấy những người dùng mà họ đã mang lại lợi nhuận nhiều nhất, thông qua việc xem xét số tiền họ đã chi cho các giao dịch vi mô hoặc số lượng trò chơi của họ mà họ đã mua. Khi những người dùng này đã được xác định, họ có thể được truy vấn thư viện của mình để tạo ra một ma trận người dùng-mặt hàng mà sau đó sẽ được phân tích. Từ đây, tất cả những gì cần làm là các công ty lấy mẫu thư viện của người dùng ngẫu nhiên theo cách tương tự để tìm ra những người dùng chưa mua sản phẩm của họ nhưng lại giống với những người dùng đang tạo ra nhiều doanh thu cho sản phẩm của họ. Đến thời điểm hiện tại, dường như không có chỉ báo nào cho thấy Steam cho phép các công ty trực tiếp quảng bá sản phẩm của họ đến những người dùng cụ thể, nhưng nếu tính năng đó bao giờ được thêm vào, thì nó sẽ dễ dàng bị khai thác thông qua phương pháp đã được giải thích ở trên.

## Cơ hội nghiên cứu trong tương lai

Mặc dù trò chơi điện tử là một cách tốt để giết thời gian, việc sử dụng thời gian mà ai đó sẽ dành để chơi một trò chơi điện tử để gợi ý một trò chơi cho họ có thể gây hại. Giả sử thời gian chơi dự đoán cho một người dùng là 18 giờ mỗi ngày cho một trò chơi mà họ không sở hữu. Nếu một gợi ý về trò chơi đó được đưa ra cho người dùng và họ tình cờ

tuân theo hành vi dự đoán, thì điều đó là không đạo đức. Việc cố ý đề xuất một hoạt động giải trí sẽ chiếm hầu hết thời gian trong ngày của ai đó với mục đích duy nhất là doanh thu của công ty là hoàn toàn tham lam. Người dùng nên được đề xuất các mục với một số giới hạn về thời gian chơi dự đoán để không có tác động tiêu cực đến năng suất của họ hoặc khả năng thực hiện các thói quen hàng ngày tiêu chuẩn khác như ăn uống, ngủ nghỉ và tắm rửa.

Các hướng nghiên cứu trong tương lai bao gồm mở rộng tập dữ liệu để bao gồm các trò chơi có ít hoặc không có đánh giá, các tựa game chưa phát hành, và tận dụng thông tin từ các đánh giá một cách toàn diện hơn, chẳng hạn như thời gian và nội dung cập nhật.

# Kết luận

Trong phân tích của chúng tôi, chúng tôi nhận thấy rằng yếu tố Độ phổ biến được trọng số nặng hơn so với yếu tố Sự tham gia trong cả hệ thống gợi ý dựa trên mục và dựa trên người dùng. Xu hướng này nhấn mạnh một thách thức đáng chú ý trong các hệ thống gợi ý phân tích ma trận quy mô lớn, nơi mà các trò chơi được sản xuất bởi các studio có ngân sách tiếp thị cao hơn và có độ hiển thị cao hơn thường được ưu ái. Sự thiên lệch này đặt các studio nhỏ và độc lập vào thế bất lợi vì các trò chơi của họ có thể bị lu mờ bởi các studio lớn hơn và có thể không được giới thiệu trên trang cửa hàng. Độ hiển thị thấp sẽ dẫn đến doanh số bán hàng thấp hơn và sẽ làm giảm nhu cầu hơn nữa.

Tuy nhiên, phân tích này cũng đưa ra một cơ hội chiến lược tiềm năng cho các studio nhỏ hơn. Bằng cách hiểu cơ chế và cấu trúc của hệ thống gợi ý của Steam, nơi mà độ phổ biến ảnh hưởng lớn đến doanh số bán game và các gợi ý, các nhà phát triển và đội ngũ marketing có thể nhắm đến các chiến lược và nỗ lực tiếp cận rộng rãi hơn để cải thiện khả năng hiển thị của họ. Các chiến thuật như tiếp cận trên mạng xã hội, tương tác với các streamer nổi bật và các influencer trong ngành game, hoặc tăng cường sự hiện diện trên mạng xã hội có thể nâng cao mức độ tương tác của họ và giúp các studio tập trung nỗ lực vào những lĩnh vực đúng đắn. Bằng cách sử dụng những chiến thuật này, họ có thể làm việc xung quanh các thuật toán gợi ý và giúp cân bằng sân chơi.

# Lời cảm ơn.

Jacob Andrew Turner, Tiến sĩ – Cố vấn

# Tài liệu tham khảo

1. Rizani, M. N., Khalid, M. N. A., & Iida, H. (2023). Ứng dụng khái niệm Meta-Gaming vào nền tảng xuất bản: Phân tích nền tảng trò chơi Steam. Thông tin, 14(2), 110. https://doi.org/10.3390/info14020110
2. Sahin, B. (Tháng 2 năm 2023). Cá nhân hóa Thương mại điện tử: Hướng dẫn hoàn chỉnh của bạn. Bloomreach. https://www.bloomreach.com/en/blog/2017/ecommerce-[cá nhân hóa.](https://www.bloomreach.com/en/blog/2017/ecommerce-personalization)
3. Giới thiệuCáiHơi nướcTương tácHệ thống gợi ý.Tháng Ba2020). https://steamcommunity.com/games/593110/announcements/detail/1716373422378 712841
4. Robertson, A. (Tháng 7 năm 2019). Hệ thống gợi ý tương tác mới của Steam được xây dựng để tìm kiếm.ẩnđá quý.CáiVerge. https://www.theverge.com/2019/7/11/20690231/valve-steam-labs-interactive-[công cụ máy học gợi ý trò chơi.](https://www.theverge.com/2019/7/11/20690231/valve-steam-labs-interactive-recommender-game-recommendation-machine-learning-tool)
5. Grayson, N. (Tháng 7 năm 2019). Đợt giảm giá mùa hè Steam năm nay thật hỗn loạn, các nhà phát triển game nói với Kotaku. https://kotaku.com/this-years-steam-summer-sale-was-a-mess-game-developer-1836215859
6. Lin, Dayi & Bezemer, Cor-Paul & Hassan, Ahmed E.. (2018). Nghiên cứu thực nghiệm về các trò chơi truy cập sớm trên nền tảng Steam. Kỹ thuật phần mềm thực nghiệm. 23. https://doi.org/10.1007/s10664-017-9531-3
7. Lin, D., Bezemer, CP. & Hassan, A.E. Nghiên cứu các cập nhật khẩn cấp của các trò chơi phổ biến trên nền tảng Steam. Kỹ thuật phần mềm thực nghiệm 22, 2095–2126 (2017). https://doi.org/10.1007/s10664-016-9480-2
8. Seif El-Nasr, M., Drachen, A., & Canossa, A. (Biên soạn) (2013). Phân tích trò chơi: Tối đa hóa giá trị của dữ liệu người chơi. Springer. https://doi.org/10.1007/978-1-4471-4769-5
9. Ajisaria, Abinava. Xây dựng một động cơ gợi ý với lọc hợp tác. https://realpython.com/build-recommendation-engine-collaborative-[lọc/.](https://realpython.com/build-recommendation-engine-collaborative-filtering/)2021
10. Lin, Dayi & Bezemer, Cor-Paul & Zou, Ying & Hassan, Ahmed E.. (2019). Nghiên cứu thực nghiệm về đánh giá trò chơi trên nền tảng Steam. Kỹ thuật phần mềm thực nghiệm. 24. https://doi.org/10.1007/s10664-018-9627-4
11. Karthikeyan, K. (2021, ngày 2 tháng 11). Gameopedia. Cải thiện Đề xuất Trò chơi Video: Giải quyết Thách thức và Cơ hội trong Thương mại Điện tử. https://www.gameopedia.com/problems-with-game-recommendations/
12. Grayson, Nathan. (2019, ngày 12 tháng 9). Kotaku. Đề xuất của Steam sẽ không hiển thị các trò chơi phổ biến thường xuyên nữa. https://kotaku.com/steams-recommendations-will-now-show-popular-games-less-1838073592
13. SirTapTap. (2016, 16 tháng 9). SirTapTap. Chính sách đánh giá mới của Steam gây ra nhiều vấn đề hơn là giải quyết. https://sirtaptap.com/articles/steams-new-review-policy-causes-more-problems-than-it-solves/#google\_vignette
14. R.R. (2021, ngày 9 tháng 5). Sử dụng dữ liệu của Steam để tìm và gợi ý các trò chơi tương tự. Hệ thống gợi ý trò chơi điện tử. https://medium.com/web-mining-is688-spring-2021/video-game-recommendation-system-b9bcb306bf16
15. Singhal, A., Sinha, P., & Pant, S. (2017, tháng 12) Sử dụng Học sâu trong Hiện đạiKhuyến nghịHệ thống:ATóm tắtcủaGần đâyHoạt động. https://www.researchgate.net/publication/321846860\_Use\_of\_Deep\_Learning\_in\_Modern\_Recommendation\_System\_A\_Summary\_of\_Recent\_Works
16. Guzsvinecz, Tibor & Szűcs, Judit. (2023). Phân tích độ dài và cảm xúc của các đánh giá về các thể loại trò chơi video hàng đầu trên nền tảng Steam. Máy tính trong con ngườiHành vi.149.

[https://www.researchgate.net/publication/373882591\_Độ\_dài\_và\_phân\_tích\_cảm\_xúc\_của\_các\_đánh\_giá\_về\_các\_thể\_loại\_game\_video\_cấp\_cao\_trên\_nền\_tảng\_steam](https://www.researchgate.net/publication/373882591_Length_and_sentiment_analysis_of_reviews_about_top-level_video_game_genres_on_the_steam_platform)

1. Chambers, Chris & Feng, Wu-chang & Sahu, Sambit & Saha, Debanjan. (2005). Đặc trưng dựa trên đo lường của một bộ sưu tập các trò chơi trực tuyến. https://www.researchgate.net/publication/228956555\_Measurement-based\_Characterization\_of\_a\_Collection\_of\_On-line\_Games
2. Sahin, B. (2023, ngày 2 tháng 2). Cá nhân hóa Thương mại điện tử: Hướng dẫn hoàn chỉnh của bạn.Thời gian thựcCá nhân hóa. https://www.bloomreach.com/en/blog/2017/ecommerce-personalization
3. Dror, R., Peled-Cohen, L., Shlomov, S. & Reichart, R. (2020). Kiểm tra ý nghĩa thống kê cho Xử lý ngôn ngữ tự nhiên. Springer. https://doi.org/10.1007/978-3-031-02174-9
4. Cohen, S. (2019). Phân tích Bayes trong Xử lý Ngôn ngữ Tự nhiên, Phiên bản Thứ hai. Springer. https://doi.org/10.1007/978-3-031-02170-1
5. Liu, B. (2012). Phân tích cảm xúc và khai thác ý kiến. Springer. https://doi.org/10.1007/978-3-031-02145-9
6. Jackson, D. (2017, ngày 7 tháng 7). Giải thưởng Netflix: Cách một cuộc thi trị giá 1 triệu đô la đã thay đổiXem phim liên tụcMãi mãi.Thrillist. https://www.thrillist.com/entertainment/nation/the-netflix-prize
7. Roy, D., Ding, C. (2021). Đề xuất phim dựa trên nhiều nguồn với đánh giá và thông tin bổ sung. Tạp chí Phân tích Mạng Xã hội và Khai thác Dữ liệu 11, Bài viết 76.

<https://doi.org/10.1007/s13278-021-00785-5>

1. Zhu, M., Fang, X. (2015). Phân tích từ vựng về danh từ và tính từ từ các đánh giá trò chơi trực tuyến. Trong: Kurosu, M. (biên soạn) Tương tác giữa con người và máy tính: Công nghệ tương tác. HCI 2015. Tài liệu bài giảng trong Khoa học Máy tính(), tập 9170. Springer, Cham. https://doi.org/10.1007/978-3-319-20916-6\_62
2. Fang, X., Zhan, J. (2015) Phân tích cảm xúc sử dụng dữ liệu đánh giá sản phẩm. Tạp chí Dữ liệu Lớn 2, Bài viết 5. https://doi.org/10.1186/s40537-015-0015-2
3. Andersson, J. (2022). Phương pháp thống kê trong hệ thống gợi ý. https://doi.org/10.1007/978-1-4842-7765-2\_7
4. A. K. Balazs Hidasi, “Gợi ý dựa trên phiên với Mạng nơ-ron hồi tiếp.

Mạng lưới,ICLR,pp.1–10,2016. https://www.researchgate.net/publication/284579100\_Gợi\_ý\_dựa\_trên\_phiên\_bằng\_mạng\_nơ\_ron\_tái\_cuộc

1. Davis, N., Khám Phá Dữ Liệu Steam (2019), GitHub, https://github.com/nik-davis/steam-data-science-project
2. *Video game - toàn cầu: Dự báo thị trường Statista. Statista. (Tháng 12 năm 2023). https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide*
3. M. Wan, J. McAuley, “Đề xuất tuần tự tự chú ý” (2018),

*ICDM, https://cseweb.ucsd.edu/~jmcauley/pdfs/icdm18.pdf*

1. Batra, S., Sharma, V., Sun, Y., Wang, X., & Wang, Y. (2023). Hệ thống gợi ý Steam. arXiv preprint arXiv:2305.04890
2. W. Kang, J. McAuley, “Đề xuất tuần tự tự chú ý” (2018)

*ICDM, https://cseweb.ucsd.edu/~jmcauley/pdfs/icdm18.pdf*

1. A. Pathak, K. Gupta, J. McAuley, “Tạo ra và cá nhân hóa các gợi ý góitrênHơi nước(2017),SIGIR, https://cseweb.ucsd.edu/~jmcauley/pdfs/sigir17.pdf