

PHÂN TÍCH HÀNH VI NGƯỜI DÙNG ĐỐI VỚI QUẢNG CÁO MẠNG XÃ HỘI TRÊN BỘ DỮ LIỆU SOCIAL_NETWORK_ADS.CSV

Nguyễn Thanh Nguyên
Trường Đại học Tiền Giang

Tóm tắt: Trong thời đại công nghệ số, hành vi người dùng trên mạng xã hội ngày càng trở nên đa dạng và khó dự đoán, đòi hỏi các doanh nghiệp phải hiểu rõ thói quen và sở thích của khách hàng để xây dựng chiến lược tiếp cận hiệu quả. Từ nghiên cứu phân tích hành vi người dùng đối với quảng cáo trên mạng xã hội sẽ tập trung vào việc khai thác và phân tích dữ liệu từ bộ Social_Network_Ads.csv nhằm khám phá các yếu tố ảnh hưởng đến hành vi tương tác của người dùng với quảng cáo. Thông qua việc phân tích này giúp mô hình hóa hành vi người dùng, từ đó hỗ trợ doanh nghiệp tối ưu hóa chiến lược quảng cáo và nâng cao hiệu quả tiếp thị. Ngoài ra, nghiên cứu cũng góp phần phát triển kỹ năng xử lý dữ liệu, trực quan hóa thông tin và xây dựng mô hình dự đoán trong lĩnh vực khoa học dữ liệu và trí tuệ nhân tạo.

Từ khóa: hành vi người dùng, mạng xã hội, quảng cáo, doanh nghiệp.

ANALYSIS OF USER BEHAVIOR TOWARD SOCIAL MEDIA ADVERTISING BASED ON THE SOCIAL_NETWORK_ADS.CSV DATASET

Nguyen Thanh Nguyen
Tien Giang University

Abstract: In the digital era, user behavior on social media platforms has become increasingly diverse and difficult to predict, requiring businesses to understand customers' habits and preferences to develop effective marketing strategies. This study focuses on analyzing user behavior toward social media advertising by leveraging the Social_Network_Ads.csv dataset to uncover key factors influencing user interactions with ads. By applying data analysis and machine learning techniques, the study aims to model user behavior to support the optimization of advertising strategies and improve marketing effectiveness. Additionally, the research contributes to enhancing data processing, information visualization, and behavior prediction modeling skills within the domains of data science and artificial intelligence.

Keywords: user behavior, social media, advertising, business.

Nhận bài: 18/04/2025

Phản biện: 22/05/2025

Duyệt đăng: 26/05/2025

I. ĐẶT VẤN ĐỀ

Nội dung nghiên cứu trung vào việc sử dụng bộ dữ liệu thực tế Social_Network_Ads.csv, giúp mô phỏng hành vi người dùng trong môi trường số. Bộ dữ liệu có cấu trúc rõ ràng và chứa thông tin phù hợp để phân tích hành vi người dùng. Kết quả phân tích có thể ứng dụng trực tiếp trong việc cải thiện chiến lược quảng cáo nhằm thu hút và giữ chân khách hàng. Dựa trên dữ liệu này, các doanh nghiệp có mục tiêu nhắm đến khách hàng có tiềm năng, tiết kiệm chi phí quảng cáo và tăng doanh thu. Kết quả nghiên cứu sẽ cung cấp thêm bằng chứng khoa học về hành vi người dùng trên mạng xã hội, từ đó xây dựng mô hình dự đoán hiệu quả. Ngoài ra, nghiên cứu này còn đóng góp vào việc ứng dụng các kỹ thuật phân tích dữ liệu và trí tuệ nhân tạo trong marketing số. Mục tiêu đặt ra: Phân tích và đánh giá hành vi của người dùng đối với

quảng cáo trên mạng xã hội thông qua bộ dữ liệu Social_Network_Ads.csv nhằm khám phá các yếu tố ảnh hưởng đến quyết định mua hàng.

II. NỘI DUNG NGHIÊN CỨU

2.1. Cơ sở lý thuyết

2.1.1 Dataset

Bộ dữ liệu Social_Network_Ads.csv (website <https://www.kaggle.com>) là một tập dữ liệu thường được sử dụng trong các bài học về phân tích dữ liệu, học máy (machine learning) và thống kê. Bộ dữ liệu này thường chứa các thông tin liên quan đến người dùng trên mạng xã hội và hành vi của họ liên quan đến quảng cáo. Mục tiêu chính là dự đoán khả năng người dùng sẽ thực hiện một hành động cụ thể mua sản phẩm hay không mua sản phẩm) dựa trên các đặc điểm cá nhân của họ. Cấu trúc dữ liệu: Dữ liệu bao gồm 400 dòng và 5 cột với các thông tin như sau:

Bảng 1. Bộ dữ liệu Social_Network_Ads.csv

User ID	Gender	Age	EstimatedSalary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	0
...	...	...	...	...

Trong đó:

- + User ID: Mã định danh của người dùng (không ảnh hưởng đến phân tích).
- + Gender (Giới tính): Được biểu diễn dưới dạng phân loại (Nam/Nữ).
- + Age (Tuổi): Biến số liên tục, phản ánh độ tuổi người dùng.
- + EstimatedSalary (Thu nhập ước tính): Giá trị liên tục thể hiện mức thu nhập hàng năm.
- + Purchased (Đã mua): Biến nhị phân (0: không mua hàng hoặc 1: mua hàng), chỉ ra người dùng đã mua sản phẩm hay chưa sau khi xem quảng cáo.

2.1.2 Các nghiên cứu liên quan

Nghiên cứu khảo sát 200 người mua sắm trực tuyến tại NCR (Gurugram, Delhi, Noida) xác định bốn yếu tố chính ảnh hưởng đến hành vi mua hàng qua mạng xã hội: Nguồn cập nhật, Người có ảnh hưởng, Sự tiện lợi và Chia sẻ trải nghiệm. Trong đó, “Người có ảnh hưởng” là yếu tố quan trọng nhất. Kết quả giúp nhà bán lẻ và nhà phát triển tận dụng hiệu quả mạng xã hội (Chand Prakash và các cộng sự, 2021); Với hơn 4 tỷ người dùng, mạng xã hội ảnh hưởng lớn đến hành vi tiêu dùng, đặc biệt trong tìm kiếm thông tin và quyết định mua hàng. Người tiêu dùng dễ tiếp nhận quảng cáo nhắm mục tiêu và chịu ảnh hưởng từ người nổi tiếng. Doanh nghiệp cần điều chỉnh chiến lược tiếp thị để tăng hiệu quả tương tác (Pappu Sindhuja và các cộng sự, 2023); Nghiên cứu tiếp thị tập trung vào vai trò của mạng xã hội trong thúc đẩy ý định mua hàng và duy trì lòng trung thành của người tiêu dùng. Mạng xã hội ảnh hưởng đến hành vi, tâm lý và thái độ tiêu dùng, giúp doanh nghiệp tối ưu chiến lược tiếp thị, nâng cao thương hiệu và lợi nhuận trong cạnh tranh ngày càng gay gắt (Guoqing Zhang, 2023); Influencer marketing đã thay đổi hành vi người tiêu dùng và trở thành công cụ hiệu quả để tiếp cận khách hàng mục tiêu. Với độ tin cậy cao và khả năng tác động mạnh, influencer giúp thúc đẩy quyết định mua hàng. So với quảng cáo truyền thống, hình thức này tạo sự kết nối và tương tác tốt hơn. Nghiên cứu tập trung phân tích các yếu tố ảnh hưởng, cơ hội và thách thức trong việc khai thác influencer marketing hiệu quả. (Dr. Mustafizul Haque và các cộng sự, 2024)

2.3 Phân tích mạng

Sử dụng Gephi để phân tích mạng trên bộ dữ

liệu Social_Network_Ads.csv với các đặc điểm như sau:

- Cấu trúc mạng: 564 nút (node) là các User, Gender (Giới tính), Age (Tuổi), EstimatedSalary (thu nhập ước tính), Purchased (Đã mua hay chưa mua) và 1600 cạnh (Edges) là các đường nối giữa các điểm dữ liệu biểu diễn mối quan hệ hoặc sự tương tác của người dùng có chung hành vi hoặc chung đặt trưng từ bộ dữ liệu được tạo lập từ bộ dữ liệu Social_Network_Ads.csv có tỷ lệ Male (nam giới) được đánh giá là 49% và Female (nữ giới) là 51%. Trong đó phân thành 02 cụm lớn đại diện cho giới tính: Male (Nam), Female (Nữ). Các nút lớn đại diện cho các nhóm có mức độ ảnh hưởng cao hơn hoặc có nhiều kết nối.

- Mật độ kết nối: Các cụm có mật độ kết nối cao thể hiện rằng các dữ liệu trong cụm có mối liên hệ chặt chẽ. Điều này có thể phản ánh tính đồng nhất trong hành vi mua sắm của từng giới. Sử dụng Layout dạng Force Atlas (Repulsion strength: 10000.0) và các thông số đặc trưng của Appearance (Node hiển thị Ranking: Degree có Min size: 10 và Max size: 50; Edges: Weight với bộ màu tương ứng).

Thuật toán Force Atlas là một công cụ mạnh mẽ để trực quan hóa đồ thị, giúp các nhà phân tích hiểu rõ hơn về cấu trúc mạng và các mối quan hệ giữa các node. Với khả năng điều chỉnh cao và ứng dụng đa dạng, nó là một trong những thuật toán sắp xếp đồ thị được sử dụng rộng rãi nhất hiện nay.

$$F_{repulsion}(u, v) = -k \cdot \frac{1}{\|x_u - x_v\|^2}$$

Trong đó:

+ k: Hằng số tỷ lệ.

+ x_u, x_v : Vị trí của node u và v.

Node u và v có cạnh nối bị hút về nhau:

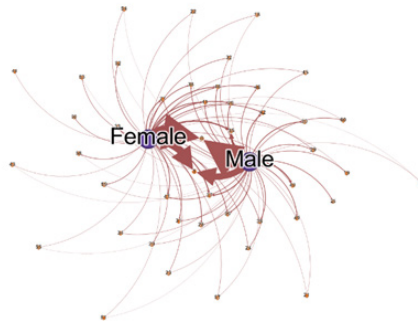
$$F_{attraction}(u, v) = -k \cdot \log(\|x_u - x_v\|)$$

Cập nhật vị trí các node sau mỗi vòng lặp vị trí các node được điều chỉnh lại dựa trên:

$$\Delta x_u = \sum F_{repulsion} + \sum F_{attraction}$$

2.3.1 Phân tích Cộng đồng theo Gender (giới tính), Age (tuổi) và Purchased (mua hàng hay không mua hàng)

Cấu trúc dữ liệu phân tích theo giới tính, độ tuổi và hành vi mua hàng hoặc không mua hàng: có 47 node và 800 edge.



Hình 2.3.1. Phân tích Cộng đồng theo Gender, Age (tuổi) và Purchased (mua hàng hay không mua hàng)

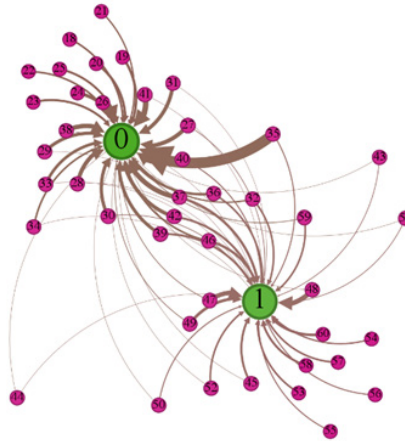
Qua hình 2.3.1 cho thấy tỷ lệ mua hàng của nữ giới cao hơn so với nam giới, tuy nhiên tỷ lệ chênh lệch không cao (mua hàng: nữ giới là 37.75% và nam giới là 33.67%; không mua hàng: nữ giới: 62.25% và nam giới là 66.33%).

➤ Vì vậy phân tích khía cạnh giới tính nam và giới tính nữ có tỷ lệ mua hàng hay không mua

hàng gần như tương đồng.

2.3.2 Phân tích Cộng đồng theo Age (tuổi) và Purchased (mua hàng hay không mua hàng)

Cấu trúc dữ liệu phân tích theo độ tuổi và hành vi mua hàng hoặc không mua hàng: có 45 node và 400 edge.

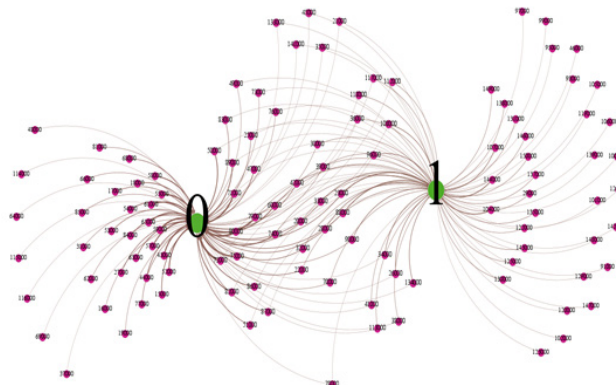


Hình 2.3.2. Phân tích Cộng đồng theo Age (tuổi) và Purchased (mua hàng hay không mua hàng)

Qua hình 2.3.2 cho thấy hành vi mua hàng hay không mua hàng được thể hiện rõ rệt hơn ở từng độ tuổi. Mật độ các node (độ tuổi) đi vào node số 0 nhiều hơn node số 1, nghĩa là tỷ lệ không mua hàng (64.25%) cao hơn rất nhiều so với mua hàng (35.75%) nhưng tùy thuộc vào lứa tuổi có tỷ lệ mua hàng khác nhau.

2.3.3 Phân tích Cộng đồng theo EstimatedSalary (thu nhập ước tính) và Purchased (mua hàng hay không mua hàng)

Cấu trúc dữ liệu phân tích theo thu nhập ước tính và hành vi mua hàng hoặc không mua hàng: có 119 node và 400 edge.



Hình 2.3.3. Phân tích Cộng đồng theo thu nhập ước tính và mua hàng hay không mua hàng

Hình 2.3.3 thể hiện rõ sự phân vùng hành vi mua và không mua hàng dựa trên thu nhập ước tính. Hai nút trung tâm (0: không mua, 1: mua) lớn hơn, các nút nhỏ xung quanh gần nhãn thu nhập. Các cạnh nối biểu thị mối liên kết giữa thu nhập và trạng thái mua hàng. Hình minh họa ba vùng rõ rệt: bên trái nút 0 là nhóm không mua hàng (100%) chủ yếu thu nhập thấp; vùng giữa nút 0 và 1 là nhóm hỗn hợp, đa phần thu nhập trung bình nhưng số người mua ít hơn không mua; bên phải nút 1 là nhóm mua hàng (100%) với thu nhập cao. Sau quá trình phân tích nêu trên thì chúng ta cần tiếp cận với nhóm người có thu nhập cao thì khả năng mua hàng cao hơn rất nhiều so với nhóm người có thu nhập thấp.

III. KẾT LUẬN

Phân tích hành vi người dùng đối với quảng

cáo trên mạng xã hội chỉ ra rằng độ tuổi và thu nhập ảnh hưởng rõ rệt đến hành vi mua hàng, trong khi giới tính không đáng kể (chênh lệch chỉ 4.08%). Nhóm tuổi từ 36 trở lên, đặc biệt là từ 46 tuổi, có tỷ lệ mua hàng cao (>80%). Thu nhập giúp phân biệt nhóm khách hàng tiềm năng, từ đó định hướng chiến lược quảng cáo hiệu quả hơn, tập trung vào nhóm thu nhập cao để tối ưu chuyển đổi. Đồng thời, hiểu rõ nhóm thu nhập thấp và trung bình giúp thiết kế chính sách phù hợp.

Hướng phát triển đề tài là xây dựng hệ thống hỗ trợ quyết định cho doanh nghiệp trong tối ưu ngân sách quảng cáo, phân bổ nguồn lực và theo dõi hiệu quả chiến dịch dựa trên hành vi người dùng. Nghiên cứu cũng mở rộng ứng dụng sang các lĩnh vực giáo dục, y tế, bất động sản và mô hình thương mại điện tử.

TÀI LIỆU THAM KHẢO

- [1]. Vũ Huy Thông (2010), Giáo trình Hành vi người tiêu dùng, Nxb Đại học Kinh tế quốc dân, Hà Nội.
- [2]. Chand Prakash, Dr. Sunil Kumar, Amit Dangi, and Kanchan Yadav (2021), Exploring the role of social media in shaping consumer buying behavior: a factor analysis approach, *Academy of Marketing Studies Journal*, Vol 25.
- [3]. Pappu Sindhuja, Akankhya Panda, S.V.S.P.P. Jaya Sankar Krishna (2023), Influence of Social Media on Consumer Buying Behavior, *International Journal of Scientific Research in Engineering and Management (IJSREM)*, Vol 7.
- [4]. Guoqing Zhang (2023), The Influence of Social Media Marketing on Consumers' Behavior, *Proceedings of the 2023 International Conference on Management Research and Economic Development*, DOI: 10.54254.
- [5]. Dr. Mustafizul Haque, Dr. S. Md. Shakir Ali, Dr. Nasa Dhanraj, Dr. Renu Bala, Mr. Vikram Bajaj, (2024), Impact of Influencer Marketing on Consumer Behavior : An Analytical Study, *Tuijin Jishu/Journal of Propulsion Technology*, Vol 44.