

NGHIÊN CỨU PHÂN NHÓM KHÁCH HÀNG TIÊU DÙNG SẢN PHẨM THỜI TRANG NỘI ĐỊA CỦA CÔNG TY TRÁCH NHIỆM HỮU HẠN COMTEC

RESEARCH ON THE CLUSTERING OF CUSTOMERS CONSUMING DOMESTIC FASHION PRODUCTS OF COMPANY LIMITED COMTEC

Nguyễn Thị Lệ^{1,*}, Lê Thị Hoài Thu²

DOI: <http://doi.org/10.57001/huiv5804.2025.020>

TÓM TẮT

Hiểu rõ nhu cầu, đặc điểm tiêu dùng nhằm đáp ứng tốt nhu cầu về sản phẩm thời trang của khách hàng mục tiêu có vai trò rất quan trọng đối với các công ty may Việt Nam để tiêu thụ sản phẩm ở thị trường nội địa. Bài báo này trình bày kết quả nghiên cứu phân nhóm khách hàng tiêu dùng sản phẩm thời trang nội địa của một công ty may (Công ty Trách nhiệm hữu hạn Comtec). Dữ liệu bán hàng tại một cửa hàng chính của công ty trong một năm được thu thập và ứng dụng giải thuật K-means clustering để phân nhóm khách hàng. Các chỉ số thống kê thể hiện đặc điểm tiêu dùng sản phẩm thời trang của ba nhóm khách hàng được xác định. Mô hình tối ưu thể hiện mối quan hệ giữa số lần mua hàng, số lượng sản phẩm và tuổi với mức chi cho sản phẩm thời trang của khách hàng cũng được xác định bằng Bayesian Model Average.

Từ khóa: Phân nhóm khách hàng, phân tích thành phần chính, K-means clustering, sản phẩm thời trang.

ABSTRACT

Understanding consumer needs and characteristics to well meet the fashion product needs of target customers plays a very important role for Vietnamese garment companies that consume products in the domestic market. This article presents the results of research on clustering consumers of domestic fashion products of a garment company (Company Limited Comtec). Sales data at a company's main store for one year was collected and the K-means clustering algorithm was applied to cluster customers. Three customer clusters have been identified along with statistical indicators showing the fashion product consumption characteristics of each cluster. The optimal model showing the relationship between the number of purchases, number of products and age with the customer's spending on fashion products is also determined using the Bayesian Model Average.

Keywords: Customer clustering, Principal component analysis, K-means clustering, fashion products.

¹Trường Đại học Công nghiệp Hà Nội

²Trường Cao đẳng Công nghiệp Thực phẩm

*Email: le.nguyenthi@hauivn.edu.vn

Ngày nhận bài: 23/5/2024

Ngày nhận bài sửa sau phản biện: 23/8/2024

Ngày chấp nhận đăng: 26/01/2025

1. GIỚI THIỆU

Hiểu rõ nhu cầu, đặc điểm tiêu dùng nhằm đáp ứng tốt nhu cầu về sản phẩm thời trang của khách hàng mục tiêu có vai trò rất quan trọng đối với các công ty may. Để thực hiện được điều đó cần phân chia khách hàng thành các nhóm dựa trên những đặc trưng có liên quan tới khách hàng và sự tiêu dùng sản phẩm. Một số nghiên cứu đã được thực hiện nhằm xác định các nhu cầu, đặc điểm tiêu dùng sản phẩm thời trang đã được thực hiện. Tác giả Đào Thị Hà Anh đã nghiên cứu những yếu tố ảnh hưởng đến sự trung thành với thương hiệu sản phẩm may mặc của người tiêu dùng Việt Nam [1]. Tác giả Đặng Thị Kim Thoa đã xác định các yếu tố ảnh hưởng đến sự sẵn sàng mua hàng may mặc nội địa của người tiêu dùng Việt Nam [2]. Tác giả Nguyễn Thị Mai Hoa đã nghiên cứu đặc điểm tiêu dùng sản phẩm thời trang công sở tại Thành phố Hà Nội [3]. Ảnh hưởng của một số yếu tố liên quan với khách hàng tới mức chi cho sản phẩm thời trang của nữ giới trong độ tuổi từ 30 - 55 tuổi tại thành phố Hà Nội cũng đã được xác định và có thể phân nhóm khách hàng dựa trên

mức chi cho sản phẩm thời trang [4, 5]. Các nghiên cứu trên đã đề cập đến mức độ ảnh hưởng của tiếp thị sản phẩm thời trang đến nhu cầu người tiêu dùng, đặc điểm tiêu dùng của khách hàng, sự tác động của các yếu tố đến sự sẵn sàng mua hàng hóa may mặc nội địa của người tiêu dùng và tập trung tìm hiểu về thị trường mục tiêu, những yếu tố ảnh hưởng, đặc điểm tiêu dùng sản phẩm thời trang nội địa; phân tích môi trường bên trong và bên ngoài hoặc phân nhóm khách hàng theo một vài tiêu chí cụ thể [6, 7] mà chưa đề cập đến việc phân nhóm khách hàng tiêu dùng sản phẩm thời trang của doanh nghiệp với nhiều thuộc tính thể hiện các khía cạnh khác nhau dựa trên cơ sở khoa học một cách hiệu quả.

Mục đích của nghiên cứu này nhằm phân nhóm khách hàng tiêu dùng sản phẩm thời trang nội địa tại một công ty may dựa vào dữ liệu bán hàng tại cửa hàng bán lẻ tiêu biểu của công ty, để góp phần xây dựng cơ sở khoa học cho hoạch định chiến lược sản phẩm và giải pháp nâng cao hiệu quả trong sản xuất và kinh doanh.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu thực hiện phân nhóm khách hàng tại một cửa hàng bán lẻ điển hình của Công ty TNHH Comtec với 1012 khách hàng đã mua sản phẩm thời trang trong thời gian từ 01/01/2022 đến 31/12/2022.

Dữ liệu bán hàng của năm 2022 lưu trữ trên máy tính của công ty được trích xuất và biên tập lại dưới dạng tệp Excel chứa các thông tin về khách mua hàng bao gồm: số điện thoại, tuổi, giới tính, mã sản phẩm, tên sản phẩm, giá bán của từng sản phẩm, ngày mua hàng, số sản phẩm mà khách hàng mua, số tiền khách hàng thanh toán,... để thuận tiện cho việc xử lý và phân tích dữ liệu trên phần mềm R.

Hệ số tương quan r (*correlation coefficient*) đo lường mối quan hệ giữa tuổi (*age*), số lần mua hàng (*number*), số sản phẩm (*quantity*) khách hàng đã mua và mức chi trả (*pay*) cho các sản phẩm thời trang của từng khách hàng được xác định bằng hàm *pair.panel* (hoặc hàm *cor*) của *Caret package* trong phần mềm R [8]. Kỹ thuật BMA được áp dụng để xác định mô hình đa biến tối ưu thể hiện mối quan hệ giữa tuổi (*age*), số lần mua hàng (*number*), số sản phẩm (*quantity*) khách hàng đã mua và mức chi trả (*pay*) cho các sản phẩm thời trang của từng khách hàng [8].

Khách hàng được phân loại bằng cách ứng dụng giải thuật k-means clustering [9]. Phương pháp Gap Statistic và Elbow được sử dụng để xác định số nhóm tối ưu [10]. Quá trình được thực hiện thông qua các bước: chuẩn hóa dữ liệu, hiển thị ma trận khoảng cách, xác định k tối ưu, xác định tâm các nhóm, khoảng cách của các đối tượng

trong nhóm tới tâm và đánh giá sự phân nhóm [11, 12]. Sau khi phân nhóm, các đặc trưng thống kê của từng nhóm cũng được các định.

3. KẾT QUẢ VÀ BÀN LUẬN

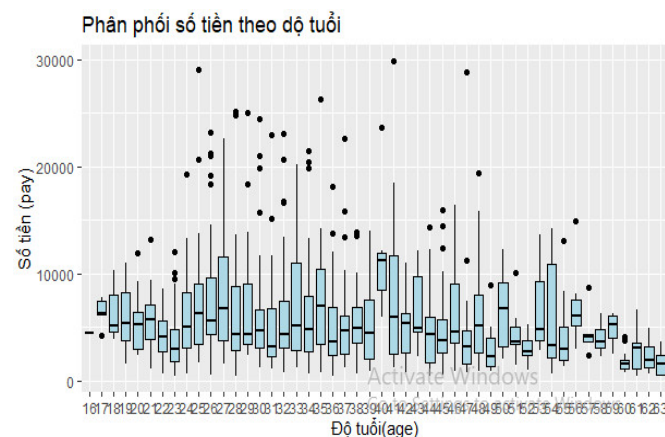
Dữ liệu bán hàng năm 2022 tại cửa hàng bán lẻ điển hình của công ty may được phân tích mô tả trên phần mềm R và thu được kết quả như bảng 1.

Bảng 1. Đặc trưng thống kê của các biến số

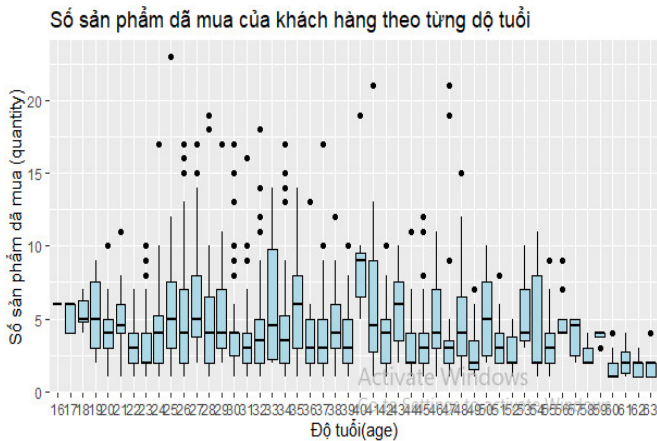
Tên biến	Nhỏ nhất	Lớn nhất	Trung bình	Trung vị	Độ lệch chuẩn	Khoảng tin cậy 95%
<i>age</i> (tuổi)	16,00	63,00	35,77	33,00	11,43	(24,34; 47,20)
<i>number</i> (số lần mua/năm)	1,00	11,00	2,35	2,00	1,74	(0,61; 4,10)
<i>quantity</i> (số lượng sản phẩm đã mua/năm)	1,00	23,00	4,39	4,00	3,39	(1,00; 7,79)
<i>pay</i> (mức chi: nghìn đồng/năm)	399	29843	5695	4358	4670	(1020,80; 10369,29)

Số lượng khách hàng mua sản phẩm tại cửa hàng là 1012 người, trong đó 307 là nam, 705 là nữ (hơn gấp đôi số khách hàng nam). Độ tuổi trung bình của khách hàng là 35,77 tuổi, dao động từ 16 đến 63 tuổi. Mức chi trung bình của khách hàng là 5.695.000 đồng, dao động từ 399.000 đồng đến 29.843.000 đồng. Khách hàng nữ 41 tuổi chi trả số tiền mua quần áo thời trang cao nhất năm 2022 tại cửa hàng với số tiền là 29.843.000 đồng với 21 sản phẩm. Những khách hàng ngoài 50 tuổi chi trả cho các sản phẩm ít hơn so với khách hàng trẻ tuổi hơn (hình 1).

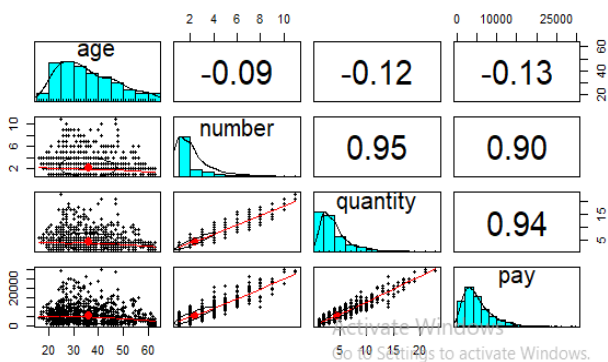
Khách hàng nữ 25 tuổi mua số lượng sản phẩm nhiều nhất năm là 23 sản phẩm, với tổng số tiền mà khách hàng này chi trả là 29.039.000 đồng. Số lượng sản phẩm mà các khách hàng cao tuổi hơn đã mua nhìn chung cũng ít hơn so với các khách hàng trẻ tuổi (hình 2).



Hình 1. Biểu đồ mô tả số tiền đã chi trả theo tuổi của khách hàng



Hình 2. Số sản phẩm khách hàng đã mua theo từng độ tuổi



Hình 3. Mối quan hệ tuyến tính từng đôi một giữa các biến số

Kết quả cho thấy tồn tại mối quan hệ tuyến tính tỷ lệ thuận đáng kể giữa số lần mua hàng (*number*) và số lượng sản phẩm (*quantity*) của khách hàng mua ($r = 0,95$), tiếp đến là tổng số tiền chi cho sản phẩm quần áo thời trang (*pay*) và số lượng sản phẩm (*quantity*) ($r = 0,94$), số sản phẩm (*number*) và mức chi (*pay*) ($r = 0,90$), tuổi của khách hàng (*age*) không thể hiện mối quan hệ đáng kể với các biến còn lại (hình 3). Khi số lần mua và số lượng sản phẩm mà khách hàng mua tăng lên, thì số tiền thanh toán của khách hàng cũng tăng. Điều này hoàn toàn phù hợp với thực tế bán hàng.

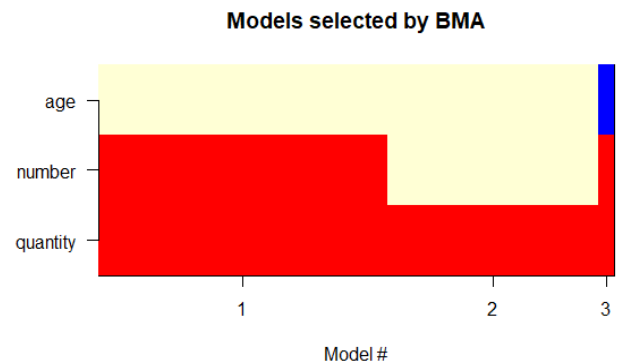
Kết quả phân tích dữ liệu trên phần mềm R với kỹ thuật BMA cho thấy, có 3 mô hình thể hiện mối quan hệ giữa mức chi trả (*pay*) cho các sản phẩm quần áo thời trang của khách hàng và các thông số đầu vào đã được tìm thấy. Từ đó, lựa chọn mô hình tối ưu như sau:

$$\text{pay} = 112,980 - 4,612 \cdot \text{age} + 243,767 \cdot \text{number} + 1177,591 \cdot \text{quantity};$$

$$R^2 = 0,889, \text{BIC} = -2206,86$$

Sự biến thiên của số lượng sản phẩm (*quantity*), số lần mua (*number*) và tuổi (*age*) của khách hàng giải thích được 88,9% sự biến thiên mức chi trả của khách hàng (*pay*) cho các sản phẩm thời trang. Mức chi trả của khách

hàng tỷ lệ thuận với số lần mua hàng, số lượng sản phẩm, tỷ lệ nghịch với tuổi của khách hàng. Số sản phẩm mà khách hàng mua có tác động tới doanh thu nhiều hơn so số lần mua hàng.

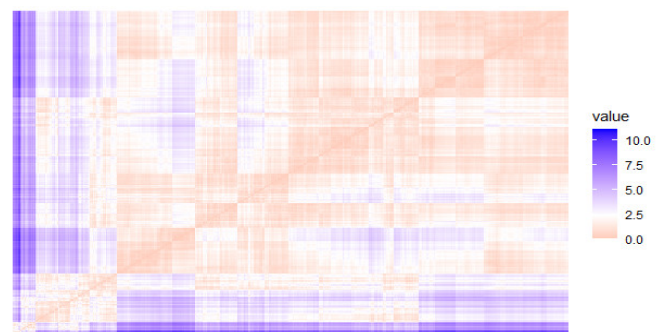


Hình 4. Biểu đồ BMA của các mô hình tối ưu

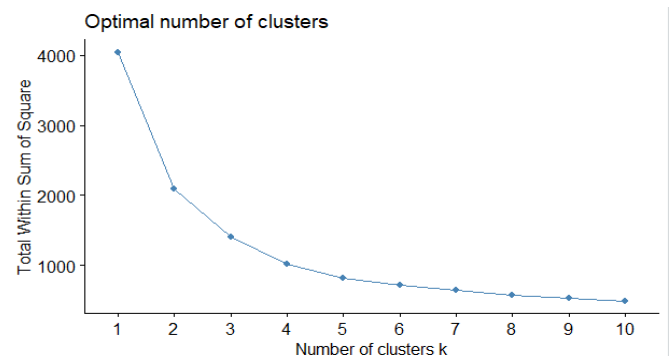
Biểu đồ BMA (hình 4) biểu thị các mô hình tối ưu tìm được, trong đó biến *quantity* xuất hiện trong tất cả các mô hình, số lượng sản phẩm bán được là quan trọng nhất đối với doanh thu năm 2022 của cửa hàng, tiếp đến là số lần mua, độ tuổi có ảnh hưởng không đáng kể.

Kết quả phân nhóm khách hàng ứng dụng K-means clustering:

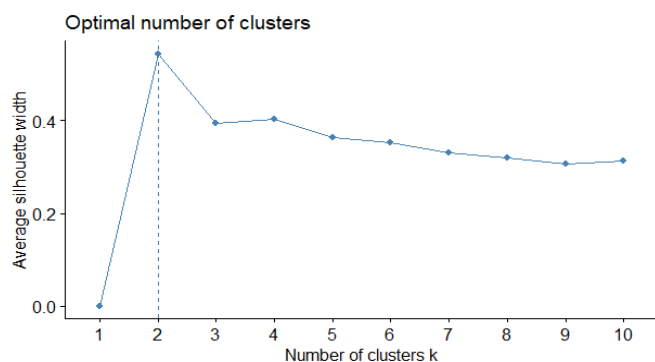
Bộ dữ liệu có chỉ số Hopkins được xác định bằng 0,92 nên có thể phân nhóm bằng K-means clustering. Hình 5 cho thấy khả năng phân nhóm của dữ liệu là khá tốt.



Hình 5. Tương quan giữa các biến và chỉ số c(Hopkins)

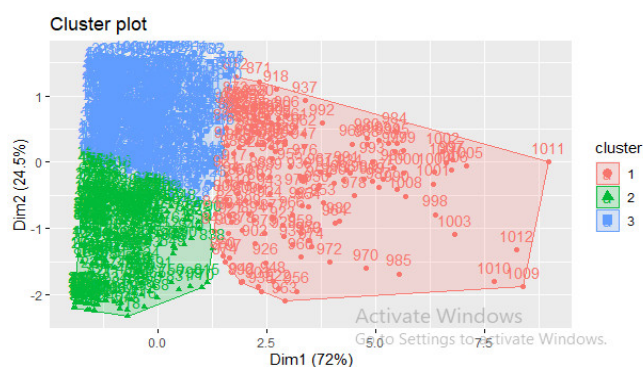


Hình 6a. Kết quả xác định k tối ưu bằng phương pháp Within Sum of Square (WSS)



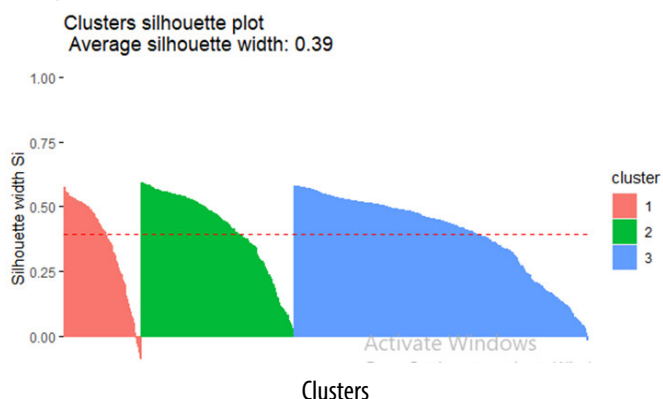
Hình 6b. Kết quả xác định k tối ưu bằng phương pháp Silhouette Width

Từ kết quả xác định số nhóm k tối ưu (hình 6a, 6b), chọn số nhóm $k = 3$. Khách hàng được chia thành 3 nhóm và hiển thị theo đặc trưng thứ nhất và thứ 2 như trên hình 7. Thông số 1 giải thích được 72% và thông số 2 giải thích được 24,5% của tổng số dao động (sự khác biệt) giữa các nhóm. Để đánh giá sự phân nhóm này, chỉ số SW (Silhouette Width) được tính cho các nhóm và thể hiện trên biểu đồ hình 8.



Hình 7. Hiển thị khách hàng theo 3 nhóm

Giá trị trung bình của chỉ số $SW = 0,39$ cho thấy khách hàng được phân thành 3 nhóm là phù hợp. Giá trị chỉ số này ở tâm của 3 nhóm đều cao hơn 0,5 và không có đối tượng nào không phân nhóm tốt (hình 8). Các đặc trưng của từng nhóm khách hàng được xác định như trên bảng 2.



Hình 8. Biểu đồ đánh giá sự phân nhóm của bộ dữ liệu

Bảng 2. Các đặc trưng thống kê của 3 nhóm khách hàng

Tên biến	Nhỏ nhất	Lớn nhất	Trung bình	Trung vị	Độ lệch chuẩn	Khoảng tin cậy 95%
	$N_1 = 51$ (5%)					
age1 (tuổi)	24,00	56,00	33,82	32,00	7,78	(26,04; 41,60)
number1 (số lần mua hàng/năm)	4,00	11,00	7,31	7,00	1,77	(5,54; 9,09)
quantity1 (số lượng sản phẩm/năm)	8,00	23,00	14,33	14,00	3,42	(10,92; 17,75)
pay1 (mức chi = nghìn đồng/năm)	14873	29843	20199	19859	4090,13	(16108,83; 24289,09)
	$N_2 = 683$ (67,5%)					
age2 (tuổi)	16,00	63,00	36,76	34,00	12,01	(24,75; 48,77)
number2 (số lần mua hàng/năm)	1,00	4,00	1,52	1,00	0,68	(0,85; 2,20)
quantity2 (số lượng sản phẩm/năm)	1,00	7,00	2,70	2,00	1,23	(1,47; 3,93)
pay2 (mức chi = nghìn đồng/năm)	399	6165	3210	3132	1501	(1708,59; 4710,60)
	$N_3 = 278$ (27,5%)					
age3 (tuổi)	17,00	61,00	33,68	32,00	10,15	(23,53; 43,83)
number3 (số lần mua hàng/năm)	1,00	8,00	3,48	3,00	1,38	(2,10; 4,86)
quantity3 (số lượng sản phẩm/năm)	2,00	14,00	6,73	6,00	2,34	(4,38; 9,07)
pay3 (mức chi = nghìn đồng/năm)	6179	14604	9141	8570	2301,94	(6838,66; 11442,53)

Nhóm 1 có tuổi trung bình là 33,8; gồm 51 khách hàng chiếm 67,5% tổng số khách hàng trong năm, gồm 373 hóa đơn mua hàng (15,7%); với 731 sản phẩm và doanh thu là 1.030.147.000 đồng (17,9% tổng doanh thu), số lần mua trung bình là 7,31; mức chi trung bình cho năm 2022 là 20.199.000 đồng. Nhóm 2 có tuổi trung bình là 36,8; gồm 683 khách hàng chiếm 5% tổng số khách hàng trong năm, gồm 1041 hóa đơn mua hàng; với 1845 sản phẩm và doanh thu là 2.192.155.000 đồng (38% tổng doanh thu); số lần mua trung bình là 1,52; mức chi trung bình cho năm 2022 là 3.210.000 đồng. Nhóm 3 có tuổi trung bình là 33,7; gồm 278 khách hàng chiếm 27,5% tổng số khách hàng trong năm, gồm 967 hóa đơn mua hàng; với 1870 sản phẩm và doanh thu là 2.541.085.000 đồng (44,1% tổng doanh thu); số lần mua trung bình là 3,48; mức chi trung bình cho năm 2022 là 9.141.000 đồng.

Như vậy, có thể chia khách hàng tại cửa hàng này của công ty thành 3 nhóm có đặc điểm về tuổi, số lần mua hàng, số sản phẩm đã mua và số tiền đã chi trả ở các mức khác nhau. Đây là cơ sở ban đầu để công ty có kế hoạch, chiến lược sản phẩm và chăm sóc khách hàng phù hợp với nhu cầu của từng nhóm.

4. KẾT LUẬN

Tồn tại mối quan hệ tuyến tính tỷ lệ thuận đáng kể giữa số lần mua hàng và số lượng sản phẩm của khách hàng mua ($r = 0,95$), tiếp đến là tổng số tiền chi trả cho sản phẩm thời trang của công ty trong một năm ($r = 0,90$), tuổi của khách hàng có mức độ liên quan ít hơn so với các yếu tố còn lại. Tồn tại mối quan hệ tuyến tính tỷ lệ thuận đáng kể giữa số lượng người mua sản phẩm và số lượng sản phẩm bán được, doanh thu và số lượng sản phẩm bán được trong ngày của công ty. Số lượng sản phẩm bán được có ảnh hưởng tới giá trị của các hóa đơn bán hàng hơn là giá sản phẩm của công ty.

Mô hình tối ưu thể hiện mối quan hệ giữa mức chi cho các sản phẩm thời trang và số lần mua, số lượng sản phẩm đã mua của công ty và tuổi của khách hàng đã được xác định. Sự biến thiên của tuổi, số lần mua hàng và số lượng sản phẩm của khách hàng đã mua giải thích được 88,9% sự biến thiên doanh thu của cửa hàng.

Khách hàng tại cửa hàng của công ty có thể chia thành 3. Nhóm 1 chiếm 5% tổng số khách với 731 sản phẩm và tổng doanh thu là 1.030.147.000 đồng chiếm 17,9% trên tổng doanh thu, mức chi trung bình cho các sản phẩm là 20.199.000 đồng. Nhóm 2 chiếm 67,5% tổng số khách hàng với 1845 sản phẩm và tổng doanh thu là 2.192.155.000 đồng chiếm 38% trên tổng doanh thu, mức chi trung bình cho các sản phẩm là 3.210.000 đồng. Nhóm 3 chiếm 27,5% tổng số khách hàng với 1.870 sản phẩm và tổng doanh thu là 2.541.085.000 đồng chiếm 44,1%, mức chi trung bình cho sản phẩm là 9.141.000 đồng. Từ kết quả này là một trong những cơ sở công ty có thể xây dựng chiến lược sản phẩm và sản xuất để đáp ứng phù hợp với từng nhóm khách hàng.

Nghiên cứu này được thực hiện trên dữ liệu bán hàng tại cửa hàng tiêu biểu của một công ty may từ 1/1/2022-31/12/2022. Do đó, kết quả có thể bị ảnh hưởng bởi dịch Covid 19.

[2]. Đặng Thị Kim Thoa, *Nghiên cứu các nhân tố ảnh hưởng đến sự sẵn sàng mua hàng may mặc nội địa của người tiêu dùng Việt Nam ở các thành phố*. Luận án tiến sĩ, Đại học Kinh tế quốc dân, Hà Nội, 2017.

[3]. Nguyễn Thị Mai Hoa, *Nghiên cứu đặc điểm tiêu dùng sản phẩm thời trang công sở nữ trên thị trường TP Hà Nội*. Luận văn cao học, Trường Đại học Sư phạm Kỹ thuật Hưng Yên, 2017.

[4]. Dương Thị Hạnh, *Nghiên cứu một số yếu tố ảnh hưởng tới mức chi cho sản phẩm thời trang của phụ nữ 30-55 tuổi tại Thành phố Hà Nội*. Luận văn cao học, Trường Đại học Công nghiệp Hà Nội, 2021.

[5]. Nguyễn Thị Lệ, Dương Thị Hạnh, "Mức chi cho sản phẩm thời trang của phụ nữ 30-55 tuổi tại thành phố Hà Nội," *Tạp chí Khoa học và Công nghệ, Trường Đại học Công nghiệp Hà Nội*, 59, 5, 2023. doi: <https://doi.org/10.57001/huinh5804.2023.178>

[6]. Rajagopal, "Consumer culture and purchase intentions toward fashion apparel in Mexico," *Journal of Database Marketing & Customer Strategy Management*, 18, 286-307, 2011.

[7]. Samiyah Jamal, Dr Adeel Maqbool, Dr SMTariq Zafar, "Analysis of fashion product of apparels from Consumer lifestyle perspectives: An empirical study," *International Journal of Research in Business and Technology*, 3, 3, 2013.

[8]. Ron Kenett, Shelemyahu Zacks, Daniele Amberti, *Modern Industrial Statistics with applications in R, MINITAB and JMP*. John Wiley & Sons, Ltd., 2021 doi: 10.1002/9781118763667.ch9.

[9]. Alboukadel Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. Statistical tools for high-throughput data analysis (STHDA), 2017.

[10]. Paolo Giordani, Maria Brigida Ferraro, Francesca Martella, *An Introduction to Clustering with R*. Springer, 2020.

[11]. Tibshirani R., Walther G., Hastie T., "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 63(2), 411- 423, 2021.

[12]. Joshi K. D., Nalwade P. S., "Modified K-Means for Better Initial Cluster Centres," *International Journal of Computer Science and Mobile Computing*, 11 7 p 2, 2013.

AUTHORS INFORMATION

Nguyen Thi Le¹, Le Thi Hoai Thu²

¹Hanoi University of Industry, Vietnam

²Food Industrial College, Vietnam

TÀI LIỆU THAM KHẢO

[1]. Đào Thị Hà Anh, "Các yếu tố tác động đến lòng trung thành thương hiệu của người tiêu dùng: Nghiên cứu trường hợp hàng may mặc Việt Nam," *Tạp chí Khoa học và Công nghệ, Trường Đại học Công nghiệp Hà Nội*, 46, 2018.