

PHÂN TÍCH THỐNG KÊ TÌNH HÌNH ĐỌC SÁCH CỦA SINH VIÊN KHOA KHOA HỌC TỰ NHIÊN, TRƯỜNG ĐẠI HỌC CẦN THƠ

Nguyễn Thị Hồng Dân*, Võ Văn Tài,

Châu Quốc Việt và Nguyễn Trần Quốc Dũng

Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ, Việt Nam

*Tác giả liên hệ: nthdan@ctu.edu.vn

Lịch sử bài báo

Ngày nhận: 11/01/2022; Ngày nhận chỉnh sửa: 19/5/2022; Ngày duyệt đăng: 07/7/2022

Tóm tắt

Từ số liệu thực tế khảo sát được tại Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ, bằng các phương pháp phân tích thống kê đơn biến và đa biến, bài viết đã xác định được các yếu tố và nhóm các yếu tố ảnh hưởng đến tình hình học sách, cũng như mối liên hệ giữa kết quả học tập và tình hình đọc sách của sinh viên tại Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ. Dựa trên các phương pháp phân loại thống kê, bài viết cũng xây dựng được mô hình phù hợp để phân loại học lực sinh viên thông qua các thông tin về tình hình đọc sách của sinh viên. Kết quả nghiên cứu là thông tin hữu ích cho Khoa Khoa học Tự nhiên cũng như Trường Đại học Cần Thơ nhằm đẩy mạnh phong trào đọc sách của sinh viên, từ đó nâng cao kết quả học tập. Cách làm trong nghiên cứu này có thể áp dụng cho những nơi khác và nhiều vấn đề tương tự trong thực tế.

Từ khóa: Đọc sách, kết quả học tập, Khoa Khoa học Tự nhiên, phân tích thống kê, tình hình đọc sách của sinh viên.

STATISTICAL ANALYSIS ON STUDENT READING PRACTICE AT THE COLLEGE OF NATURAL SCIENCES, CAN THO UNIVERSITY

Nguyen Thi Hong Dan*, Vo Van Tai,

Chau Quoc Viet, and Nguyen Tran Quoc Dung

College of Natural Sciences, Can Tho University, Vietnam

*Corresponding author: nthdan@ctu.edu.vn

Article history

Received: 11/01/2022; Received in revised from: 19/5/2022; Accepted: 07/7/2022

Abstract

Upon the survey at the College of Natural Sciences, Can Tho University and univariate and multivariate statistical analysis, the article has identified the factors and groups of factors impacting students reading books, as well as correlations between the learning results and time for reading of students at the College of Natural Sciences, Can Tho University. Based on statistical classification methods, the article also builds a suitable model to classify students' learning ability through information about students' reading time. The research results are useful information for the College of Natural Sciences as well as Can Tho University to promote student reading. The method in this study can be applied to other institutions and practical issues.

Keywords: College of natural sciences, learning results, reading books, student reading, statistical analysis.

DOI: <https://doi.org/10.52714/dthu.12.4.2023.1055>

Trích dẫn: Nguyễn, T. H. D., Võ, V. T., Châu, Q. V., & Nguyễn, T. Q. D. (2023). Phân tích thống kê tình hình đọc sách của sinh viên Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ. *Tạp chí Khoa học Đại học Đồng Tháp*, 12(4), 25-35. <https://doi.org/10.52714/dthu.12.4.2023.1055>.

1. Đặt vấn đề

Sách là người bạn không thể thiếu trên con đường học tập, làm việc và phát triển của mỗi chúng ta. Đọc sách mang lại rất nhiều lợi ích cho con người như mở rộng kiến thức, nâng cao khả năng giao tiếp và giải trí ở nhiều lĩnh vực trong cuộc sống, đặc biệt là đối với sinh viên (SV), những người mới và chuẩn bị bước vào cuộc sống tự lập và giao tiếp xã hội. Nhưng có một thực tế xảy ra là không phải ai cũng có sở thích và thói quen đọc sách. Theo số liệu do Tổng Cục Thống kê Việt Nam (Niên giám thống kê 2020, tr. 852), thống kê trong ba năm gần đây (2018-2020) cho thấy, bình quân mỗi năm Việt Nam xuất bản khoảng 400 triệu bản sách, trong số này có trên 300 triệu bản là sách giáo khoa, giáo trình phục vụ cho nhu cầu học tập, nghiên cứu, tham khảo kiến thức cho 2,8 triệu học sinh, SV trên cả nước. Như vậy, chỉ còn khoảng 100 triệu bản sách chia gần 100 triệu dân. Điều này đồng nghĩa với việc mỗi năm trung bình một người Việt đọc chừng một quyển sách. Chưa dừng lại ở đó, một khảo sát mới đây của báo Dân Trí (Gia, 2019; Hoài, 2019) nhắm vào đối tượng là giới trẻ cho thấy kết quả đáng quan ngại, 98% giới trẻ cho biết họ không hề đọc quyển sách nào trong tuần qua; 80% bạn trẻ không đụng đến sách suốt một năm qua và chỉ 12% bạn trẻ trong nhóm 20-30 tuổi cho biết bản thân có đọc sách, truyện khác ngoài sách chuyên môn. Theo tác giả Võ (2013), văn hóa đọc ở Việt Nam đang dần “nhạt phai”. Người đọc, nhất là giới trẻ, có xu hướng “lười đọc”, đọc ít, đọc nhanh. Tâm lý chung của họ là ngại đọc sách dày, ngại đọc sách in, ngại đọc sách về vấn đề lý luận - ngại đọc vì không có thời gian. Đó là những kết luận được chỉ ra trong các nghiên cứu về văn hóa đọc của người Việt Nam trong thời gian gần đây (Võ, 2019; Đoàn, 2017; Hossain, 2017; Nguyễn, 2021). Trong các nghiên cứu trên đã chỉ ra các yếu tố ảnh hưởng đến thói quen đọc sách của người Việt dựa trên số liệu khảo sát và phân tích thống kê mô tả. Tuy nhiên chưa áp dụng các phương pháp phân tích thống kê hiện đại vào trong nghiên cứu. Trong nghiên cứu của tác giả Nguyễn (2013) đã áp dụng các phân tích thống kê suy diễn vào trong nghiên cứu của mình và căn cứ trên kết quả phân tích thống kê để chỉ ra các yếu tố ảnh hưởng đến thói quen đọc sách chuyên ngành SV tại Trường Đại học Bách khoa Thành phố Hồ Chí

Minh. Trong nghiên cứu này của chúng tôi đã áp dụng phương pháp phân tích chùm để xác định các yếu tố ảnh hưởng đến thói quen đọc sách và kết quả học tập của SV khoa Khoa học Tự nhiên tại Trường Đại học Cần Thơ.

Theo công bố của Sở Thông tin - Truyền thông Thành phố Hồ Chí Minh (2019) nhấn mạnh “đó chưa phải là con số đáng buồn nhất vì theo Trung tâm Nghiên cứu Việt Nam và Đông Nam Á, có tới 44% người Việt thi thoảng đọc sách và 26% không bao giờ đụng vào sách”. Theo Lê (2016), trong danh sách 61 nước có số lượng người đọc sách cao trên thế giới không có Việt Nam, trong khi ba nước là Singapore, Malaysia và Indonesia thuộc danh sách này với số đầu sách trung bình của người dân mỗi năm là 12 quyển. Cũng theo kết quả thống kê năm 2012, trung bình một tháng người Hàn Quốc đọc khoảng 4,5 quyển (trong đó 3,1 sách giấy và 1,4 sách điện tử). Năm 2012, bình quân mỗi người Trung Quốc đọc 4,39 cuốn sách giấy, con số này tăng ổn định năm thứ 7 liên tiếp; còn tỷ lệ đọc sách điện tử tăng rõ rệt: năm 2012, bình quân mỗi người Trung Quốc đọc 2,35 cuốn sách điện tử, trong khi đó năm 2011 con số này chỉ là 1,42 cuốn, mức tăng đạt 65,5%. Rõ ràng, khi coi văn hóa đọc là nền tảng, động lực phát triển và hội nhập quốc tế thì thực tế đáng buồn về văn hóa đọc ở Việt Nam không khỏi khiến chúng ta phải lo lắng. Vì vậy, việc xây dựng một xã hội đọc ở nước ta hiện nay là yêu cầu cấp thiết, nhìn từ cả hai phía lý luận và thực tiễn.

Theo thống kê của Trung tâm Học liệu Trường Đại học Cần Thơ (2021), trung bình mỗi năm có khoảng 339.000 số lượt bạn đọc vào trung tâm, số lượng của năm 2020 có giảm so với trước. Cũng trong thống kê này cho thấy số lượt mượn tài liệu về nhà giảm mỗi năm, cụ thể năm 2015 có 112.545 lượt mượn thì tới năm 2020 con số này giảm hơn ba lần chỉ còn gần 38.000 lượt mượn. Ngược lại điều đó, thì số lượt truy cập tài liệu nội sinh dạng số tăng lên mỗi năm, tính cho năm 2020 là 48.735 lượt xem. Nhìn tổng quan số liệu thì ta thấy nhu cầu đọc sách điện tử của SV Trường Đại học Cần Thơ tăng lên.

Trước tình hình đó, để đánh giá mức độ thường xuyên đọc sách của SV khoa Khoa học Tự nhiên, xác định các yếu tố thật sự ảnh hưởng đến nhu cầu đọc sách của SV và tìm mối liên hệ giữa mức độ đọc sách,

giáo trình với kết quả học tập của SV, chúng tôi đã tiến hành khảo sát về tình hình đọc sách của SV khoa Khoa học Tự nhiên. Việc nghiên cứu thực hiện hai vấn đề: (i) xác định các yếu tố ảnh hưởng đến mức độ đọc sách của SV và (ii) xác định mối liên hệ giữa kết quả học tập và thời gian đọc sách của SV thông qua mô hình hồi qui và một số mô hình phân loại. Từ việc phân tích thống kê, chúng tôi mong tìm ra chính sách, giải pháp trong việc nâng cao thời gian và chất lượng đọc sách của SV khoa Khoa học Tự nhiên. Kết quả của bài viết là thông tin hữu ích cho các cấp lãnh đạo trong việc nắm bắt thông tin, đáp ứng nhu cầu của SV trong vấn đề đọc sách.

2. Phương pháp nghiên cứu

2.1. Mẫu nghiên cứu

Số liệu được khảo sát thực tế tại Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ trong học kì II và học kì III năm học 2020-2021. Số lượng khảo sát là 383 SV được thu thập ngẫu nhiên theo các khối, ngành của Khoa từ khóa 43 cho đến khóa 46. Do thời gian thực hiện khảo sát rơi vào thời gian thực hiện chỉ thị giãn cách xã hội, nên khảo sát được thực hiện bằng bảng câu hỏi trực tuyến được gửi trực tiếp đến SV. Bảng khảo sát bao gồm 19 câu hỏi với 19 biến, trong đó có hai biến định lượng và 17 biến định tính, cơ cấu mẫu được trình bày trong Bảng 1.

Bảng 1. Cơ cấu các biến được khảo sát

Biến	Kí hiệu	Thang đo	Tỷ lệ (%)
Ngành	NG	1 = Toán ứng dụng (TUD) 2 = Hóa học (HH) 3 = Hóa dược (HD) 4 = Sinh học (SH) 5 = Vật lý kỹ thuật (VLKT)	27,2 21,4 25,3 18,0 8,1
Khóa học	KH	1 = 43 2 = 44 3 = 45 4 = 46	6,0 31,9 26,1 36,0
Giới tính	GT	1 = Nam 2 = Nữ	30,8 69,2
Điểm trung bình tích lũy	TBTL	Thang đo khoảng từ [0; 4,0] 1 = Kém 2 = Trung bình 3 = Khá 4 = Giỏi 5 = Xuất sắc	7,6 21,6 46,5 15,8 8,5
Phương thức đọc sách	PTĐS	1 = Sách in 2 = Sách Online 3 = Cả 2 phương thức trên	49,1 23,2 27,7
Lý do chọn đọc sách online	LDĐS	1 = Tìm kiếm nhanh 2 = Không có thời gian đến nhà sách 3 = Không tốn tiền mua sách 4 = Sách in không đa dạng về sách ngoại văn 5 = Lý do khác	89,7 27,2 70,3 15,9 3,1
Thể loại sách yêu thích	TL	1 = Chính trị - Pháp luật 2 = Khoa học công nghệ - kỹ thuật 3 = Văn học - Nghệ thuật 4 = Truyện - Tiểu thuyết 5 = Tâm lý - Tâm linh - Tôn giáo 6 = Sách thiếu nhi 7 = Sách kỹ năng 8 = Sách tham khảo 9 = Sách ngoại văn	12,5 25,6 31,3 57,2 30,0 7,8 56,9 40,2 16,4

Thời gian đọc sách trong ngày	TGDSTN	1 = Buổi sáng 2 = Buổi trưa 3 = Buổi chiều 4 = Buổi tối 5 = Tùy thời gian rảnh	1,6 0,8 1,8 20,1 75,7
Thời gian đọc sách trung bình	TGDSTB	1 = '< 0.5 h/ngày' 2 = '0.5 – 1 h/ngày' 3 = '1 – 2 h/ngày' 4 = '2 – 3 h/ngày' 5 = '> 3 h/ngày'	13,6 42,3 30,8 7,8 5,5
Yếu tố đọc sách	YTĐS	1 = Được người thân bạn bè giới thiệu 2 = Bia đẹp 3 = Nội dung hay hấp dẫn 4 = Đáp ứng nhu cầu cung cấp kiến thức 5 = Yếu tố khác	31,1 17,2 71,5 76,2 2,4
Địa điểm tìm sách	ĐĐTS	1 = Nhà sách 2 = Tiệm sách cũ 3 = Thư viện (Trường/Khoa/huyện, tỉnh) 4 = Shop Online 5 = Sách điện tử (Ebook)	64,5 17,5 59,3 47,3 38,1
Địa điểm đọc sách	ĐĐĐS	1 = Ở nhà 2 = Trong thư viện Trường/Khoa 3 = Thư viện thành phố 4 = Nhà sách 5 = Quán Café	89,8 35,5 2,6 15,9 39,2
Thói quen khi đọc xong một quyển sách	TQĐS	1 = Cất để đọc lại 2 = Cho bạn bè mượn 3 = Bán lại cho người khác 4 = Giới thiệu cho người thân bạn bè 5 = Trả lại sách (sách mượn)	70,8 30,6 6,5 40,5 47,0
Mức độ quan trọng của thư viện	MĐQTTV	1 = Rất không cần thiết 2 = Không cần thiết 3 = Bình thường 4 = Quan trọng 5 = Rất quan trọng	0 0 14,4 47,0 38,6
Mức độ thường xuyên đến thư viện	MĐTXTV	0 = Không 1 = Có	44,1 55,9
Lý do thường đến thư viện (Trường/Khoa)	LDTV	1 = Yên tĩnh 2 = Tiết kiệm chi phí 3 = Cơ sở hạ tầng, trang thiết bị đầy đủ 4 = Đa dạng thể loại tài liệu 5 = Nhân viên thư viện thân thiện, vui vẻ 6 = Lý do khác	81,8 71 71 71,5 17,3 1,9
Lý do không thường xuyên đến thư viện	LDKTV	1 = Tra cứu trên điện thoại thông minh nhanh hơn ở thư viện 2 = Thể loại sách/ tài liệu chưa phong phú và đa dạng 3 = Thủ tục mượn, trả rườm rà 4 = Không gian ở thư viện không thoải mái, chật chội 5 = Nhân viên thư viện không thân thiện, vui vẻ 6 = Lý do khác	79,3 19,5 32,0 15,4 4,7 7,7

Đánh giá về khả năng cung cấp tài liệu của thư viện (Trường/Khoa)	ĐGCCTL	0 = Không 1 = Có	7,3 92,7
Giải pháp thu hút đọc sách	GPĐS	1 = Đưa ra nhiều chương trình giảm giá sách, ưu đãi đặc biệt cho SV 2 = Mở nhiều hội sách, các cuộc thi tìm hiểu sách 3 = Bổ sung thêm một số thể loại sách 4 = Ý kiến khác	58,5 66,6 39,9 1,8

Trong cơ cấu mẫu trên, do số lượng SV khóa 43 còn ở lại trong thời gian học kì II và học kì III năm học 2020-2021 còn rất ít nên số quan sát của nhóm này rất hạn chế, do đó trong một số phân tích thống kê, chúng tôi đã loại những quan sát này nên còn lại 342 quan sát trong các phân tích thống kê ở phần sau.

2.2. Phương pháp phân tích số liệu

2.2.1. Các phân tích số liệu

Để làm rõ các yếu tố và nhóm yếu tố có ý nghĩa thống kê ảnh hưởng đến kết quả học tập, các phương pháp phân tích thống kê đơn biến và đa biến đã được sử dụng. Ngoài thống kê mô tả, bài toán ước lượng để đánh giá số liệu ban đầu, các phân tích thống kê sau đã được sử dụng:

Kiểm định Chi bình phương (Roxy & cs., 2008): Kiểm định sự ảnh hưởng của các yếu tố đến kết quả học tập.

So sánh hai trung bình và phân tích phương sai đơn biến (Roxy & cs., 2008; Scott, 1992): So sánh thời gian đọc sách trung bình của hai và nhiều hơn hai nhóm đối tượng.

Bài toán phân loại (Scott, 1992; Webb, 2000; Tai, 2016; Tai & Thao, 2016): Phân loại các đối tượng theo các nhóm về trung bình tích lũy dựa trên các biến quan sát.

2.2.2. Mô hình phân loại kết quả học tập dựa trên các biến quan sát

Theo quy chế học vụ của Trường Đại học Cần Thơ (2021) thì điểm trung bình tích lũy của SV chia thành năm nhóm:

Bảng 2. Các mức xếp loại học lực

Xếp loại	Điểm TBTL
Xuất sắc	3,60 - 4,00
Giỏi	3,20 - 3,59
Khá	2,50 - 3,19
Trung bình	2,00 - 2,49
Yếu, kém	< 2,00

Tuy nhiên trong phân tích phân loại thống kê với năm nhóm phân loại như trên và cỡ mẫu 342 quan sát kết hợp với nhiều biến phân tích sẽ tạo ra nhiều nhóm có tần số bằng 0. Do đó chúng tôi chia điểm trung bình tích lũy của SV thành ba nhóm như trong Bảng 3.

Bảng 3. Chia nhóm xếp loại học lực

Xếp loại	Nhóm	Điểm TBTL
Giỏi + Xuất sắc	1	3,20 - 4,00
Trung bình + Khá	2	2,00 - 3,19
Yếu, kém	3	< 2,00

Để xây dựng mô hình phân loại kết quả học tập của SV chúng tôi sử dụng các phương pháp thống kê sau đây:

a. Mô hình Fisher (Scott, 1992; Webb, 2000)

Xét tổng thể là các bạn SV có biến quan sát X là kết quả học tập của các bạn được phân loại theo Bảng 3. Các giá trị quan sát của biến X trên đối tượng thứ j là một vec-tơ cột n chiều. Gọi $x_j, S_j, (j = 1, 2, 3)$ lần lượt là vec-tơ cột trung bình và ma trận hiệp phương sai của nhóm xếp loại học lực 1, 2, 3. Ta có:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

Khi đó, ta có hàm phân biệt Fisher của ba nhóm như sau:

$$\hat{d}_i(x) = \bar{x}_i^T S^{-1} x - \frac{1}{2} \bar{x}_i^T S^{-1} \bar{x}_i, i = 1, 2, 3$$

Trong đó S là ma trận hiệp phương sai gộp của ba nhóm tổng thể, được tính như sau:

$$S = \frac{\sum_{i=1}^k (n_i - 1) S_i}{\sum_{i=1}^k n_i - k}$$

Lúc này phần tử mới x sẽ được xếp vào w_j nếu $h_i(x)$ nhỏ nhất, hay nói cách khác xếp x vào w_j nếu:

$$d_i(x) = \max_j \{d_j(x)\}, i = 1, 2, 3$$

b. Mô hình Bayes (Tai, 2016; Tai và Thao, 2016)

Giả sử các nhóm xếp loại học lực với biến quan sát x lần lượt có hàm mật độ xác suất là $f_1(x), f_2(x), f_3(x)$ với xác suất tiên nghiệm lần lượt là q_1, q_2, q_3 . Khi đó, một SV có giá trị quan sát đầu vào là x_0 thì sẽ được xếp vào nhóm học lực i nếu:

$$q_i f_i(x) > q_j f_j(x) \Leftrightarrow \frac{f_i(x)}{f_j(x)} > \frac{q_j}{q_i}, (i, j = 1, 2, 3)$$

Xác suất sai lầm trong bài toán phân loại này được xác định như sau:

$$Pe_{1,2}^{(q)} = 1 - \int_{\mathbb{R}^n} \max_{1 \leq i \leq k} \{q_i f_i\} dx, \quad i = 1, 2, 3$$

2.3. Vấn đề tính toán

- Chúng tôi sử dụng phần mềm SPSS phiên bản 22 để tính toán các thống kê đơn biến và đa biến và bài toán phân loại Fisher cho biến xếp loại học lực của SV.

- Đối với việc thực hiện bài toán phân loại Bayes, chúng ta phải thực hiện bài toán ước lượng hàm mật độ xác suất từ dữ liệu rời rạc, chúng tôi đã sử dụng phương pháp ước lượng hàm hạt nhân, đây là một phương pháp được sử dụng nhiều hiện nay và được đánh giá là độ tin cậy cao. Trong việc sử dụng phương pháp này, chúng tôi chọn hàm hạt nhân dạng chuẩn và tham số tron theo Scott (1992).

- Trong mô hình phân loại Bayes chúng ta cần có thông tin về xác suất tiên nghiệm, chúng tôi đã sử dụng phương pháp Laplace và phân tích chùm mờ (Thao & Tai, 2016).

- Các tính toán của phương pháp Bayes, được sử dụng code viết riêng bao gồm: ước lượng hàm mật độ xác suất n chiều, phân loại phần tử mới và tính sai số Bayes theo phương pháp Monte - Carlo (Tai, 2016) trên phần mềm Matlab 2017.

2.4. Các bước thực hiện

Sau khi có số liệu khảo sát, các phân tích lần lượt được thực hiện là:

- Nhập số liệu, hiệu chỉnh dữ liệu và mã hóa số liệu trên phần mềm SPSS.22;

- Tính các thống kê mô tả để đánh giá đặc trưng của số liệu theo từng nhóm đối tượng và tương quan giữa hai nhóm đối tượng;

- Kiểm định các yếu tố ảnh hưởng đến thời gian đọc sách trung bình và mối quan hệ của xếp loại học lực với các yếu tố độc sách. Tìm các yếu tố có ý nghĩa thống kê ảnh hưởng đến xếp loại học lực qua mô hình hồi qui;

- Đánh giá xếp loại học lực của SV qua một số mô hình phân loại. Từ đó tìm mô hình cho kết quả phân loại tối ưu.

Các phân tích thống kê trên được thực hiện với độ tin cậy 95%.

3. Kết quả nghiên cứu

3.1. Phân tích thống kê về các vấn đề liên quan đến thời gian đọc sách trung bình và xếp loại học lực của sinh viên Khoa Khoa học Tự nhiên

3.1.1. Phân tích thống kê biến điểm trung bình tích lũy theo từng biến và từng nhóm đối tượng

Trong phân tích phân loại thống kê với năm nhóm phân loại của biến *Thời gian đọc sách trung bình* và cỡ mẫu 342 quan sát kết hợp với nhiều biến phân tích sẽ tạo ra nhiều nhóm có tần số bằng 0. Do đó chúng tôi hiệu chỉnh biến *Thời gian đọc sách trung bình* của SV thành bốn nhóm, trong đó nhóm có thời gian đọc sách trung bình ‘>2 giờ/ngày’ là kết quả gộp của hai nhóm ‘2 -3 h/ngày’ và ‘>3h/ngày’. Đối với biến *Mức độ quan trọng của thư viện* do không có quan sát nào lựa chọn mức quan trọng nhóm ‘1’ và nhóm ‘2’, nên đối với biến *Mức độ quan trọng của thư viện* chỉ còn ba nhóm phân loại: Rất quan trọng, Quan trọng và Bình thường.

Kết quả của phương pháp kiểm định sự độc lập Chi bình phương và bài toán ước lượng trung bình các yếu tố ảnh hưởng đến xếp loại học lực được tổng kết trong Bảng 4.

Bảng 4. Phân tích thống kê biến điểm trung bình tích lũy theo từng biến và từng nhóm đối tượng

Biến	Đối tượng	Điểm TBTL	Sig.	Kết luận
PTĐS	Sách in	(2,6755; 2,8681)	0,033	Không độc lập
	Sách Online	(2,5129; 2,8199)		
	Cả 2	(2,6690; 2,8604)		
TGĐSN	Sáng - Trưa - Chiều	(2,4951; 3,0406)	0,558	Độc lập
	Tối	(2,6819; 2,9694)		
	Tùy thời gian rảnh	(2,6479; 2,7990)		

TGĐSTB	< 0,5 giờ/ngày	(2,5880; 2,9745)	0,237	Độc lập
	0,5 - 1 giờ/ngày	(2,6490; 2,8405)		
	1 - 2 giờ/ngày	(2,5355; 2,7676)		
	> 2 giờ/ngày	(2,7553; 3,1309)		
MĐTXTV	Không	(2,5093; 2,7153)	0,000	Không độc lập
	Có	(2,7773; 2,9359)		
MĐQTTV	Rất quan trọng	(2,7399; 2,9576)	0,015	Không độc lập
	Quan trọng	(2,6050; 2,7871)		
	Bình thường	(2,4656; 2,8054)		
NG	TUD	(2,5202; 2,7510)	0,000	Không độc lập
	HH, HD	(2,8503; 3,0320)		
	SH	(2,3228; 2,5950)		
	VLKT	(2,2448; 2,7456)		

Theo kết quả của Bảng 4 ta thấy, hầu hết các biến đều có ảnh hưởng đến xếp loại học lực, đó là phương thức đọc sách, mức độ thường xuyên đến thư viện, mức độ quan trọng của thư viện, khóa học và ngành.

3.1.2. Phân tích thống kê biến thời gian đọc sách trung bình theo từng biến và từng nhóm đối tượng

a. Kiểm định sự độc lập của biến thời gian đọc sách trung bình với các biến khác

Kết quả của phương pháp kiểm định sự độc lập Chi bình phương và bài toán ước lượng trung bình các yếu tố ảnh hưởng đến Thời gian đọc sách trung bình được tổng kết trong Bảng 5.

Bảng 5. Phân tích thống kê biến Thời gian đọc sách trung bình theo từng biến và từng nhóm đối tượng

Biến	Đối tượng	Khoảng TGĐSTB	Sig.	Kết luận
XLHL	Yếu, kém	(2,0045; 2,6878)	0,237	Độc lập
	Trung bình	(2,0312; 2,4012)		
	Khá	(2,3805; 2,6509)		
	Giỏi, xuất sắc	(2,1875; 2,6076)		
PTĐS	Sách in	(2,1452; 2,3982)	0,002	Không độc lập
	Sách Online	(2,1579; 2,5344)		
	Cả 2	(2,5363; 2,9143)		
TGĐSN	Sáng - Trưa - Chiều	(1,9116; 3,0884)	0,099	Độc lập
	Tối	(2,3762; 2,7832)		
	Tùy thời gian rảnh	(2,2520; 2,4662)		
MĐTXTV	Không	(2,0088; 2,2622)	0,000	Không độc lập
	Có	(2,5097; 2,7630)		
NG	TUD	(2,1909; 2,5564)	0,977	Độc lập
	HH, HD	(2,3231; 2,5936)		
	SH	(2,0441; 2,5114)		
	VLKT	(2,1515; 2,8141)		
KH	44	(2,2894; 2,5866)	0,234	Độc lập
	45	(2,3218; 2,6677)		
	46	(2,1511; 2,4838)		

Đối với biến *Thời gian đọc sách trung bình*, qua Bảng 5 ta thấy có hai biến tác động đến thời gian đọc sách trung bình của SV khoa Khoa học Tự nhiên là phương thức đọc sách và biến *Mức độ thường xuyên đến thư viện*. Qua số liệu khảo sát chúng tôi tính toán được tỷ lệ SV có mức độ đọc sách từ '1 - 2h' và '> 2h' của nhóm 'có' thường xuyên đến thư viện cao hơn gấp đôi so với nhóm 'không' thường xuyên đến thư viện. Trong đó phương thức đọc sách được nhiều bạn lựa chọn là kết hợp cả hai hình thức sách in và sách điện tử. Điều này cho thấy chúng ta không nên chỉ cung cấp cho SV nguồn sách in mà còn đẩy mạnh nguồn sách điện tử. Điều này cũng là đòi hỏi tất yếu khi mô hình giảng dạy trực tuyến đang được đẩy mạnh, đáp ứng nhu cầu tìm hiểu tài liệu trực tuyến.

b. Mô hình hồi quy đơn của biến điểm trung bình tích lũy theo các biến có mối liên hệ phụ thuộc

Tiếp tục sử dụng mô hình hồi quy tuyến tính để xây dựng phương trình đánh giá sự ảnh hưởng

của các biến này đến xếp loại học lực của SV khoa Khoa học Tự nhiên, ta thu được kết quả được trình bày trong Bảng 6.

Bảng 6. Mô hình hồi quy tuyến tính giữa điểm trung bình tích lũy với các biến có mối liên hệ phụ thuộc

Mô hình hồi quy	Hệ số R ²	Sig.
2,905 – 0,077 NG	0,012	0,041
2,482+ 0,088 KH	0,015	0,023
2,667 + 0,032 TGĐSTB	0,003	0,609
2,246 + 0,118 MĐQTTV	0,018	0,014
2,612 + 0,244 MĐTXTV	0,040	0,000

c. Mô hình hồi quy bội của biến điểm trung bình tích lũy với các biến độc lập

Ta xác định hệ số tương quan giữa các cặp biến độc lập để kiểm tra hiện tượng đa cộng tuyến xảy ra trong mô hình hồi quy bội.

Bảng 7. Hệ số tương quan các cặp biến độc lập

Biến	NG	KH	GT	PTĐS	TGĐSTN	TGĐSTB	MĐQTTV	MĐTXTV
NG	1							
KH	0,04	1						
GT	- 0,007	- 0,052	1					
PTĐS	0,109	- 0,183	- 0,057	1				
TGĐSTN	- 0,029	- 0,011	0,031	0,018	1			
TGĐSTB	0,006	- 0,054	- 0,011	0,169	- 0,096	1		
MĐQTTV	-0,011	0,027	-0,057	0,009	-0,102	0,072	1	
MĐTXTV	0,016	- 0,136	- 0,046	0,169	- 0,048	0,263	0,288	1

Dựa trên kết quả của Bảng 7, ta thấy không có cặp biến nào có hệ số tương quan lớn hơn 0,6, nên không xảy ra hiện tượng đa cộng tuyến khi đưa vào mô hình hồi quy bội.

Đưa các biến độc lập vào mô hình hồi quy bội và sử dụng giải thuật Backward để loại bỏ biến không có ý nghĩa trong mô hình.

Bảng 8. Các mô hình hồi quy bội qua các bước loại bỏ biến độc lập

Mô hình hồi quy	Hệ số R ²	Sig.
2,232 – 0,073*NG + 0,109*KH + 0,157*GT + 0,001*PTĐS – 0,060 TGĐSTN – 0,023*TGĐSTB + 0,059*MĐQTTV + 0,264*MĐTXTV	0,097	0,000
2,234 – 0,073*NG + 0,109*KH + 0,157*GT – 0,060*TGĐSTN – 0,022*TGĐSTB + 0,059*MĐQTTV + 0,265*MĐTXTV	0,097	0,000
2.174 – 0,073*NG + 0,109*KH + 0,157*GT – 0,057*TGĐSTN + 0,059*MĐQTTV + 0,253*MĐTXTV	0,096	0,000
2,007 – 0,072*NG + 0,109*KH + 0,154*GT + 0,063*MĐQTTV + 0,254*MĐTXTV	0,093	0,000
2,258 – 0,072*NG + 0,113*KH + 0,151*GT + 0,281*MĐTXTV	0,089	0,000

Ta được mô hình hồi quy bội tối ưu là:

$$TBTL = 2,258 - 0,072*NG + 0,113*KH + 0,151*GT + 0,281*MĐTXTV.$$

Qua các mô hình hồi qui xây dựng được từ Bảng 6 và Bảng 8, ta thấy tuy mô hình thu được có ý nghĩa thống kê nhưng mức độ phù hợp thấp, hiệu quả dự báo chưa cao. Qua đó ta thấy để đánh giá sự ảnh hưởng của các yếu tố đến điểm trung bình tích lũy thông qua mô hình hồi qui chưa thu được kết quả cao. Tuy nhiên, cũng nhìn lại các biến đóng góp trong mô hình trên ta thấy sát với tình hình thực tế: khi các bạn SV có Mức độ thường xuyên đến thư viện thì điểm trung bình tích lũy có sự tăng về tỷ lệ thuận. Và ta cũng thấy giá trị của điểm trung bình tích lũy cũng có quan hệ thuận với biến khóa học và giới tính. Nữ có sự vượt trội hơn một chút so với nam. Các khóa mới có điểm trung bình tích lũy cao hơn các khóa cũ, điều này cũng có thể là do các bạn khóa mới chưa học nhiều môn chuyên ngành.

3.2. Đánh giá điểm trung bình tích lũy của sinh viên Khoa Khoa học Tự nhiên qua bài toán phân loại

3.2.1. Chia dữ liệu

Ta chia dữ liệu gốc thành hai tập là tập kiểm tra và tập huấn luyện với tỷ lệ lần lượt là 20%:80%. Kết quả cụ thể được trình bày trong Bảng 9.

Bảng 9. Kết quả chia dữ liệu

Thang đo	Dữ liệu gốc	Tập kiểm tra	Tập huấn luyện
Yếu, kém (Y = 0)	26	3	23
Trung bình, khá (Y = 1)	233	46	187
Giỏi, xuất sắc (Y = 2)	83	20	63
Tổng	342	69	273

Bảng 11. Tổng hợp khả năng phân loại đúng (%) trên tập kiểm tra

Trường hợp	Biến	Khả năng phân loại đúng				
		Fisher	BayesU	BayesP	BayesL	BayesC
1 biến	X_1	29,00	83,94	97,59	97,69	99,97
	X_2	62,30	89,01	98,52	98,58	100,00
	X_3	23,20	88,80	98,52	98,58	100,00
2 biến	X_1, X_2	63,80	94,05	99,18	99,21	95,94
	X_1, X_3	39,10	94,02	99,18	99,21	99,95
	X_2, X_3	65,20	96,17	99,49	99,52	99,68
3 biến	X_1, X_2, X_3	66,70	97,90	99,72	99,73	98,70

Bảng 11 cho ta thấy khả năng phân loại đúng của hai phương pháp Fisher và Bayes. Trong đó với sai số rất nhỏ gần như bằng 0, xác suất tiên nghiệm được tìm bằng phương pháp phân tích chùm mờ BayesC có xác suất phân loại đúng cao nhất. Trong

3.2.2. Xác định biến có ý nghĩa thống kê

Với thang đo chia thành 3 mức, ta sử dụng hồi quy đa thức để xác định biến có ý nghĩa thống kê. Kết quả được trình bày trong Bảng 10.

Bảng 10. Kết quả kiểm tra biến có ý nghĩa thống kê

Biến	Likelihood Ratio Tests		
	Chi - Square	df	Sig.
NG	25,933	6	0,000
KH	37,790	4	0,000
GT	4,591	2	0,101
PTĐS	5,020	4	0,285
TGĐSTN	0,804	4	0,938
TGĐSTB	4,331	6	0,632
MĐQTTV	7,231	4	0,124
MĐTXTV	11,312	2	0,003

Qua kết quả của Bảng 10, ta thấy có ba biến có kết quả kiểm định với Sig. nhỏ hơn 0,05 là NG, KH và MĐTXTV. Điều này có nghĩa đây là ba biến có tác động đến Xếp loại học lực (XLHL) của SV. Ta sẽ thực hiện bài toán phân loại với ba biến này.

Đặt tên biến:

$$X_1: NG \quad X_2: KH \quad X_3: MĐTXTV$$

Đối với phân loại bằng phương pháp Bayes, ta sẽ dựa theo cách tìm các xác suất tiên nghiệm khác nhau. Theo đó, xác suất tiên nghiệm dựa trên phân phối đều, tỷ lệ mẫu, ước lượng Laplace và thuật toán phân tích chùm mờ lần lượt được gọi là BayesU, BayesP, BayesL và BayesC.

3.2.3. Thực hiện phân loại theo phương pháp Fisher và Bayes để tìm mô hình có khả năng phân loại tốt nhất

trường hợp một biến là biến X_2 hoặc biến X_3 ; BayesC cũng là phương pháp có xác suất phân loại đúng cao nhất cho trường hợp hai biến X_1, X_3 ; BayesL là phương pháp có xác suất phân loại đúng cao nhất cho ba biến.

Bảng 12. Tổng hợp khả năng phân loại đúng (%) cao nhất trên tập huấn luyện

Trường hợp	Biến	Khả năng phân loại đúng				
		Fisher	BayesU	BayesP	BayesL	BayesC
1 biến	X_1	23,40	78,86	94,01	94,08	94,83
	X_2	67,00	84,54	95,70	95,57	100,00
	X_3	20,90	87,04	96,11	96,16	100,00
2 biến	X_1, X_2	64,80	89,60	96,96	96,99	92,08
	X_1, X_3	23,10	91,32	97,25	97,28	98,77
	X_2, X_3	56,40	93,60	98,02	98,04	99,50
3 biến	X_1, X_2, X_3	49,50	95,60	98,60	98,62	96,92

Bảng 12 tổng hợp khả năng phân loại đúng trên tập huấn luyện ở trường hợp một, hai và ba biến cho các phương pháp. Nhìn chung, phương pháp BayesC cho kết quả phân loại tốt hơn trong bài toán này.

Qua kết quả thực hiện trên các phương pháp, ta thấy khi phân loại điểm trung bình tích lũy thì phương pháp Bayes có xác suất phân loại đúng cao hơn so với phương pháp Fisher. Phương pháp Bayes với xác suất tiên nghiệm được tìm theo hai phương pháp là ước lượng Laplace và thuật toán phân tích chùm mờ có khả năng phân loại đúng cao nhất, đối với hai biến *Khóa học* (X_2) và *Mức độ thường xuyên đến thư viện* (X_3) là hai biến có khả năng phân loại đúng cao nhất gần như là 100%. Mô hình này cũng giống như mô hình hồi qui ở trên khẳng định lại một lần nữa sự tác động của 2 biến này.

4. Kết luận

Dựa trên số liệu khảo sát thực tế và thông qua các phân tích thống kê, bài viết đã xác định được các biến liên quan đến vấn đề đọc sách và mối liên hệ giữa vấn đề đọc sách với kết quả học tập của SV khoa Khoa học Tự nhiên. Từ việc phân tích các biến có ý nghĩa thống kê, chúng tôi đã tìm ra được các mô hình hồi quy và các phương pháp phân loại tối ưu nhất trong việc xác định các biến có ảnh hưởng nhiều nhất đến kết quả học tập của SV khoa Khoa học Tự nhiên. Thông qua cuộc khảo sát, chúng tôi cũng đã nhận được những thông tin hữu ích trong việc nâng cao thời gian đọc sách trung bình, chất lượng đọc sách của SV khoa Khoa học Tự nhiên, đó là: cần cập nhật thêm các loại sách chuyên ngành hiện đại, sách ngoại văn, mở các cuộc thi tìm hiểu sách, và đề khuyến khích các bạn đến đọc sách thì nên bổ sung thêm các sách tâm lý, sách giải trí, ... tạo cho các bạn có thói quen đọc sách từ các loại sách giải trí, tâm lý và các bạn sẽ hình thành dần thói quen đọc sách, kĩ

năng đọc sách và sẽ thúc đẩy các em đọc sách chuyên ngành. Cũng từ các mô hình xây dựng được cho ta thấy khi các em có thời gian đến thư viện nhiều hơn thì kết quả học tập có sự cải thiện.

Lời cảm ơn: Nhóm tác giả xin gửi lời cảm ơn đến toàn thể SV khoa Khoa học Tự nhiên đã cung cấp những số liệu quan trọng để có thể hoàn thành được bài viết này; Xin gửi lời cảm ơn đến Trường Đại học Cần Thơ đã hỗ trợ kinh phí cho các bạn SV thực hiện đề tài nghiên cứu khoa học, mã số: TSV2021-44; Xin chân thành cảm ơn các phản biện đã có những góp ý để bài viết có thể hoàn thiện hơn.

Tài liệu tham khảo

- Đại học Cần Thơ. (2021). *Qui định công tác học vụ năm 2021*, Đại học Cần Thơ (Số 1813/QĐ-ĐHCT). Truy cập từ QD1813_QD_ban_hanh_Quy_dinh_cong_tac_hoc_vu_2021.pdf (ctu.edu.vn).
- Đoàn, T. L. (2017). *Nghiên cứu phát triển văn hóa đọc các tỉnh miền núi phía Bắc Việt Nam*. Trường Đại học Văn hóa Hà Nội.
- Gia, M. (Ngày 22 tháng 4 năm 2019). Bình quân một người Việt đọc 1 quyển sách mỗi năm. *Cổng Thông tin Chính phủ Thành phố Hồ Chí Minh*. Truy cập từ <https://tphcm.chinhphu.vn/binh-quan-mot-nguoi-viet-doc-1-quyen-sach-moi-nam>.
- Hoài, H. (Ngày 15 tháng 4 năm 2019). 5 năm Ngày sách Việt Nam: Người Việt đọc 4 hay 1 cuốn sách mỗi năm?. *Báo Điện tử VOV*. Truy cập từ <https://vov.vn/van-hoa/van-hoc/5-nam-ngay-sach-viet-nam-nguoi-viet-doc-4-hay-1-cuon-sach-moi-nam-897994.vov>.
- Hossain, Z. (December 2017). What are the leisure choice, reading and library habits of Vietnamese

- students in the age of Internet?. *Social Sciences Information Review*, Vol.11, No.4. <https://vjol.info.vn/index.php/ssirev/article/view/38247>.
- Lê, T. (Ngày 17 tháng 11 năm 2016). 61 quốc gia có người đọc sách nhiều nhất, không có Việt Nam. *Tuổi trẻ Online*. Truy cập từ <https://tuoitre.vn/61-quoc-gia-co-nguoi-doc-sach-nhieu-nhat-khong-co-viet-nam-1221061.htm>.
- Nguyễn, T. K. N. (2021). *Thông điệp về văn hóa đọc trên báo điện tử*. Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội.
- Nguyễn, T. N. H. (Ngày 21 tháng 4 năm 2021). Mơ ước có tên mình trong “Quốc gia đọc sách”. *Trang tin Điện tử Đảng bộ Thành phố Hồ Chí Minh*. Truy cập từ <https://www.thanhuytphcm.vn/tin-tuc/mo-co-ten-minh-trong-quoc-gia-doc-sach-1491876894>.
- Roxy, P., Chris, O. & Jay, D. (2008). *Statistics and data analysis*. Thomson. New York.
- Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. Wiley & Son, New York.
- Tai, V.V. (2016). L1-distance and classification problem by Bayesian. *Journal of Applied Statistics*, 44(3), 385-401. <http://dx.doi.org/10.1080/02664763.2016.1174194>.
- Thao, N.T. & Tai, V.V. (2016). A new approach for determining the prior probabilities in the classification problem by Bayesian method. *Advances in Data Analysis and Classification*, 11, 629-643. <https://link.springer.com/article/10.1007/s11634-016-0253-y>.
- Tổng cục Thống kê Việt Nam. (Năm 2021). *Niên giám Thống kê Việt Nam năm 2020 = Statistical yearbook of Vietnam 2020*. Hà Nội: NXB Thống kê. Truy cập từ <https://www.gso.gov.vn/wp-content/uploads/2021/07/Sach-NGTK-2020Ban-quyen.pdf>.
- Trung tâm Học liệu Trường Đại học Cần Thơ (2021). *Số liệu tổng hợp*. Truy cập từ <https://lrc.ctu.edu.vn/index.php/aboutus/statistics>.
- Võ, T. T. H. (2013). Văn hoá đọc ở Việt Nam trong bối cảnh hội nhập và phát triển. *Tạp chí Thư viện Việt Nam*, 2, 20-27. *Thư viện quốc gia Việt Nam*. Truy cập từ <https://nlv.gov.vn/nghep-vu-thu-vien/van-hoa-doc-o-viet-nam-trong-boi-canhhoi-nhap-va-phat-trien.html>.
- Webb, A. (2000). *Statistical pattern recognition*. Wiley & Sons, New York.