



ISSN 1859-3666  
E-ISSN 2815-5726

# Tạp chí KHOA HỌC THƯƠNG MẠI

TẠP CHÍ CỦA TRƯỜNG ĐẠI HỌC THƯƠNG MẠI



Năm thứ 24 - số 199  
3/2025



# khoa học thương mại

TẠP CHÍ CỦA TRƯỜNG ĐẠI HỌC THƯƠNG MẠI  
BỘ GIÁO DỤC VÀ ĐÀO TẠO

## TỔNG BIÊN TẬP:

**ĐINH VĂN SƠN**

## PHÓ TỔNG BIÊN TẬP:

## THƯ KÝ TÒA SOẠN

## TRƯỞNG BAN TRỊ SỰ

**NGUYỄN THỊ QUỲNH TRANG**

### ☐ Tòa soạn

Phòng 202 nhà T  
Trường Đại học Thương mại  
Số 79 đường Hồ Tùng Mậu  
Mai Dịch, Cầu Giấy, Hà Nội

☐ Điện thoại: 024.37643219 máy lẻ 2102

☐ Fax: 024.37643228

☐ Email: tckhtm@tmu.edu.vn

☐ Website: tckhtm.tmu.edu.vn

### ☐ GP hoạt động báo chí:

Số 195/GP-BTTTT ngày 05/6/2023

### ☐ Chế bản tại: Tòa soạn

Tạp chí Khoa học Thương mại

☐ In tại: Cty TNHH In & TM Hải Nam

☐ Nộp lưu chiểu: 3/2025

## HỘI ĐỒNG KHOA HỌC BIÊN TẬP

**Đinh Văn Sơn** - Đại học Thương mại (Chủ tịch)

**Phạm Vũ Luận** - Đại học Thương mại (Phó Chủ tịch)

**Nguyễn Bách Khoa** - Đại học Thương mại (Phó chủ tịch)

**Phạm Minh Đạt** - Đại học Thương mại (Ủy viên thư ký)

### Các ủy viên

- **Vũ Thành Tự Anh** - ĐH Fulbright Việt Nam (Hoa Kỳ)

- **Lê Xuân Bá** - Viện QLKT TW

- **Hervé B. Boismery** - Đại học Reunion (Pháp)

- **H. Eric Boutin** - Đại học Toulon Var (Pháp)

- **Nguyễn Thị Doan** - Hội Khuyến học Việt Nam

- **Haasis Hans** - Đại học Bremen (Đức)

- **Lê Quốc Hội** - Đại học Kinh tế quốc dân

- **Nguyễn Thị Bích Loan** - Đại học Thương mại

- **Nguyễn Hoàng Long** - Đại học Thương mại

- **Nguyễn Mai** - Chuyên gia kinh tế độc lập

- **Dương Thị Bình Minh** - ĐH Kinh tế Tp Hồ Chí Minh

- **Hee Cheon Moon** - Hội Nghiên cứu TM Hàn Quốc

- **Bùi Xuân Nhàn** - Đại học Thương mại

- **Lương Xuân Quỳ** - Hội Khoa học kinh tế Việt Nam

- **Nguyễn Văn Song** - Học viện Nông nghiệp Việt Nam

- **Nguyễn Thanh Tâm** - Đại học California (Hoa Kỳ)

- **Trương Bá Thanh** - ĐH Kinh tế - Đại học Đà Nẵng

- **Đinh Văn Thành** - Viện Nghiên cứu thương mại

- **Đỗ Minh Thành** - Đại học Thương mại

- **Lê Đình Thắng** - Đại học Québec (Canada)

- **Trần Đình Thiên** - Viện Kinh tế Việt Nam

- **Nguyễn Quang Thuấn** - Viện Hàn lâm KHXH Việt Nam

- **Washio Tomoharu** - ĐH Kwansey Gakuin (Nhật Bản)

- **Lê Như Tuyền** - Grenoble École de Management (Pháp)

- **Zhang Yujie** - Đại học Tsinghua (Trung Quốc)

## MỤC LỤC

### KINH TẾ VÀ QUẢN LÝ

---

- 1. Đinh Văn Sơn và Nguyễn Trần Hưng** - Nghiên cứu các yếu tố tác động đến việc sử dụng công nghệ AI của các doanh nghiệp thương mại điện tử Việt Nam. *Mã số: 199.1TrEm.11* 3

*Research on factors affecting the use of AI technology by Vietnamese e-commerce enterprises*

- 2. Lê Văn Chiến** - Ảnh hưởng của cảm nhận đánh đổi khi sử dụng dịch vụ hỗ trợ bởi AI đến quyết định mua hàng trực tuyến: Vai trò của chất lượng dịch vụ điện tử và hình ảnh thương hiệu. *Mã số: 199.1TrEm.11* 24

*The impact of perceived sacrifice when using AI-powered services on online purchasing decisions: The role of e-service quality and brand image*

### QUẢN TRỊ KINH DOANH

---

- 3. Phạm Hùng Cường, Lâm Nguyễn Như Nguyễn và Trần Thái Khang** - Tác động của marketing người có tầm ảnh hưởng đến ý định mua hàng của người theo dõi: Nghiên cứu thực nghiệm tại Thành phố Hồ Chí Minh. *Mã số: 199.2BMkt.21* 43

*The impact of influencer marketing on followers' purchase intention: Empirical study in Ho Chi Minh City*

- 4. Nguyễn Việt Thái và Bùi Phương Linh** - Vai trò của truyền miệng điện tử trong thúc đẩy ý định mua trực tuyến dịch vụ lưu trú của khách du lịch gen Z Việt Nam. *Mã số: 199.2TRMg.21* 61

*The role of electronic Word-of-Mouth in promoting online purchase intention of accommodation services among Vietnamese Gen Z travelers*

- 5. Đoàn Thị Yên** - Ảnh hưởng của tiền lương đến ý định nghỉ việc của bác sĩ trong các bệnh viện công trên địa bàn Hà Nội. *Mã số: 199.2HRMg.21* 77

*The Impact of Salary on Doctors' Intention to Resign in Public Hospitals in Hanoi*

## **Ý KIẾN TRAO ĐỔI**

---

- 6. Nguyễn Thị Liên và Đoàn Huy Hoàng** - Khám phá các yếu tố ảnh hưởng đến sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên thông qua các mô hình học máy: nghiên cứu tại trường Đại học Thương mại. *Mã số: 199.3OMIs.31* 87

*Exploring Factors Influencing Student Satisfaction With Lecturers' Teaching Performance Using Machine Learning Models: Research at Thuongmai University*

- 7. Trần Việt An, Nguyễn Thị Hiền và Phạm Thị Huyền** - Ảnh hưởng của marketing giác quan tới cảm nhận về chất lượng dịch vụ và sự hài lòng của phụ huynh các trường tiểu học ngoài công lập trên địa bàn thành phố Hà Nội. *Mã số: 199.3OMIs.31* 105

*The Impact of Sensory Marketing on Parents' Satisfaction: A Study at Private Primary Schools in Hanoi*

# Ý KIẾN TRAO ĐỔI

## KHÁM PHÁ CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN SỰ HÀI LÒNG CỦA SINH VIÊN ĐỐI VỚI HOẠT ĐỘNG GIẢNG DẠY CỦA GIÁNG VIÊN THÔNG QUA CÁC MÔ HÌNH HỌC MÁY: NGHIÊN CỨU TẠI TRƯỜNG ĐẠI HỌC THƯƠNG MẠI

**Nguyễn Thị Liên \***

**Email: [liennt@tmu.edu.vn](mailto:liennt@tmu.edu.vn)**

**Đoàn Huy Hoàng \***

**Email: [hoang.dh@tmu.edu.vn](mailto:hoang.dh@tmu.edu.vn)**

**\* Trường Đại học Thương mại**

*Ngày nhận: 19/12/2024*

*Ngày nhận lại: 19/02/2025*

*Ngày duyệt đăng: 24/02/2025*

Sự hài lòng của sinh viên là một thước đo quan trọng để đánh giá chất lượng giảng dạy Strong giáo dục đại học. Các nghiên cứu trước đây cũng đã khẳng định mức độ hài lòng của sinh viên phản ánh trực tiếp hiệu quả giảng dạy và ảnh hưởng đến động lực cũng như kết quả học tập của người học. Do đó, việc lắng nghe và phân tích phản hồi của sinh viên không chỉ cho phép giảng viên tại các trường đại học cải thiện phương pháp giảng dạy mà còn góp phần nâng cao trải nghiệm học tập của người học và chất lượng giáo dục tổng thể. Nghiên cứu này áp dụng các mô hình học máy (machine learning) có giám sát để dự đoán mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên tại Trường Đại học Thương mại và xác định các yếu tố ảnh hưởng quan trọng đến sự hài lòng của sinh viên. Nhóm nghiên cứu đã sử dụng 7 mô hình học máy khác nhau để chạy dữ liệu là các câu trả lời của 3.477 sinh viên về hoạt động giảng dạy của giảng viên Trường Đại học Thương mại năm học 2023-2024, kết quả cho thấy: Mô hình Hồi quy tuyến tính (Linear regression - LR) cho kết quả đảm bảo vượt trội so với các mô hình khác; đồng thời, xác định năm hoạt động của giảng viên ảnh hưởng nhiều nhất đến sự hài lòng của sinh viên bao gồm: phương pháp giảng dạy rõ ràng; khả năng bao quát và kiểm soát học tập của người học; sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả; khuyến khích người học tự học; nghiêm túc, khách quan và công bằng trong việc đánh giá. Kết quả nghiên cứu cung cấp thông tin quan trọng cho giảng viên và nhà quản lý trong việc cải thiện phương pháp giảng dạy, hướng sự tập trung tới các yếu tố ảnh hưởng lớn đến sự hài lòng của sinh viên. Bài báo cũng kì vọng là tài liệu tham khảo hữu ích cho các nhà nghiên cứu trong lựa chọn các mô hình học máy khác nhau phục vụ những điều tra tương đồng.

*Từ khóa: Hoạt động giảng dạy của giảng viên, sự hài lòng của sinh viên, mô hình học máy.*

*JEL Classifications: M12, M54, O15.*

*DOI: 10.54404/JTS.2025.199V.06*

### **Giới thiệu**

Nâng cao chất lượng giáo dục và đào tạo, đặc biệt ở bậc đại học là một trong những vấn đề luôn được quan tâm hàng đầu của xã hội hiện nay. Chất lượng đào tạo tại các trường

đại học tốt, sẽ thu hút học sinh tốt nghiệp trung học phổ thông đăng ký vào trường, tạo niềm tin cho sinh viên trong quá trình đào tạo, là điều kiện thuận lợi cho việc tìm kiếm việc làm và thành công trong sự nghiệp của người

*Tác giả liên hệ: [liennt@tmu.edu.vn](mailto:liennt@tmu.edu.vn)*



học sau khi tốt nghiệp. Đánh giá chất lượng đào tạo ở các trường đại học thông qua sự phản hồi của người học về hoạt động giảng dạy của giảng viên cho thấy: sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên tại các trường đại học phụ thuộc vào nhiều yếu tố khác nhau như: sự hiểu biết sâu và rộng của giảng viên, phương pháp giảng dạy hiệu quả, khách quan trong đánh giá người học... Tuy nhiên, các nghiên cứu truyền thống thường chỉ dựa trên thông kê đơn giản, không khai thác đầy đủ giá trị của dữ liệu phản hồi của sinh viên về hiệu quả giảng dạy của giảng viên (Muijtjens et al., 2010) (Macfadyen et al., 2016). Cụ thể như nghiên cứu của Masserini et al., (2019) chỉ ra rằng nhiều trường đại học sử dụng các thông kê mô tả như trung bình và tần suất để đánh giá phản hồi của người học về hoạt động giảng dạy của giảng viên, dẫn đến việc bỏ qua những mối quan hệ phức tạp và tiềm ẩn giữa các yếu tố ảnh hưởng đến sự hài lòng của sinh viên. Cách tiếp cận thông kê mô tả dễ thực hiện nhưng lại thiếu khả năng phát hiện các xu hướng sâu hơn trong dữ liệu phản hồi (Sulis et al., 2019). Điều này cho thấy cần áp dụng các phương pháp phân tích tiên tiến hơn để khai thác toàn diện dữ liệu, làm cơ sở vững chắc trong nâng cao chất lượng giảng dạy và trải nghiệm học tập. Trong khi đó, một số nghiên cứu gần đây đã cho thấy rằng các thuật toán machine learning có thể xử lý dữ liệu phản hồi của sinh viên một cách hiệu quả, giúp phát hiện các mẫu ẩn và mối quan hệ phức tạp giữa các biến số (Yağcı, 2022) (Okoye et al., 2022).

Nghiên cứu này sử dụng các mô hình học máy (machine learning) để dự đoán mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên dựa trên 21 tiêu chí lấy từ hoạt động khảo sát định kỳ của sinh viên Trường Đại học Thương mại năm học 2023-2024 và xác định các hoạt động nào của giảng viên có ảnh hưởng quan trọng nhất đến sự hài lòng của sinh viên. Việc sử dụng đồng thời nhiều mô hình machine learning trong nghiên cứu này cũng kì vọng chứng minh cách thức cải thiện chất lượng dự đoán mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên, khắc phục hạn chế của các phương pháp truyền thống như thông kê mô tả và hồi quy tuyến tính đơn giản.

### 1. Cơ sở lý thuyết về các yếu tố ảnh hưởng đến sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên thông qua các mô hình học máy

#### 1.1. Một số khái niệm cơ bản

*Sự hài lòng của sinh viên* là một khái niệm đa chiều, phản ánh mức độ thỏa mãn của sinh viên đối với các khía cạnh khác nhau của trải nghiệm học tập tại trường đại học. Nhiều nghiên cứu đã chỉ ra rằng sự hài lòng của sinh viên không chỉ đơn thuần là cảm xúc nhất thời mà là kết quả của quá trình đánh giá tổng thể về chất lượng giáo dục, bao gồm cả yếu tố học thuật và phi học thuật (Luis Miguel Olórtgui Alcalde, 2024; Siscan & Moldovan-Batrinac, 2024). Sự hài lòng của sinh viên có thể bị ảnh hưởng bởi nhiều yếu tố, như chất lượng giảng dạy, cơ sở vật chất, dịch vụ hỗ trợ sinh viên, môi trường học tập và các hoạt động ngoại khóa (Wong & Chapman, 2023). Trong nghiên cứu này, sự hài lòng của sinh viên được tập trung vào khía cạnh học thuật, cụ thể là mức độ sinh viên cảm thấy thỏa mãn với hoạt động giảng dạy của giảng viên. Sự hài lòng này không chỉ ảnh hưởng đến kết quả học tập của sinh viên mà còn tác động đến danh tiếng và uy tín của nhà trường (Luis Miguel Olórtgui Alcalde, 2024).

*Hoạt động giảng dạy* là một khái niệm rộng, bao gồm tất cả các hoạt động của giảng viên liên quan đến việc thiết kế, triển khai và đánh giá quá trình dạy học. Theo Rusticus và cộng sự (2023), hoạt động giảng dạy không chỉ đơn thuần là việc truyền đạt kiến thức mà còn bao gồm việc tạo ra môi trường học tập tích cực, khuyến khích sự tham gia của sinh viên và hỗ trợ sinh viên phát triển các kỹ năng cần thiết. Hoạt động giảng dạy hiệu quả đòi hỏi giảng viên phải có kiến thức chuyên môn sâu rộng, kỹ năng sư phạm tốt và khả năng thích ứng với các đối tượng sinh viên khác nhau (Agustin et al., 2024). Hoạt động giảng dạy được xem xét trên các khía cạnh như: chuẩn bị bài giảng (nội dung, phương tiện), phương pháp giảng dạy, tương tác với sinh viên và đánh giá kết quả học tập. Các khía cạnh này được lựa chọn dựa trên các nghiên cứu trước đây về các yếu tố ảnh hưởng đến sự hài lòng của sinh viên (Caldwell et al., 2024; Clements et al., 2025).

*Học máy (machine learning)* là một lĩnh vực của trí tuệ nhân tạo (AI) chuyên nghiên

cứ các thuật toán giúp hệ thống có thể học từ dữ liệu và đưa ra dự đoán mà không cần lập trình trực tiếp (Janiesch et al., 2021). Trong giáo dục, học máy đã được áp dụng rộng rãi để dự đoán kết quả học tập của sinh viên, phát hiện rủi ro bỏ học và tối ưu hóa phương pháp giảng dạy (Alyahyan & Düstegör, 2020). Nghiên cứu của Yağcı (2022) sử dụng các mô hình học máy có giám sát như: Mô hình hồi quy tuyến tính có điều chuẩn (RLR - Regularized Linear Regression), Mô hình hồi quy Máy vector hỗ trợ (SVMR - Support Vector Machine Regression), Mô hình Random Forest (RFR - Random Forest Regression), ... để dự đoán điểm thi cuối kỳ của sinh viên đại học dựa trên điểm thi giữa kỳ. Việc áp dụng học máy không chỉ giúp tăng độ chính xác của dự đoán mà còn cung cấp cái nhìn sâu sắc hơn về các yếu tố quan trọng nhất ảnh hưởng đến sự hài lòng của sinh viên, điều mà các phương pháp thống kê truyền thống khó có thể đạt được.

### 1.2. Tổng quan nghiên cứu

*1.2.1. Các nghiên cứu về sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên*

Trong những năm gần đây, nhiều công trình nghiên cứu đã tập trung khảo sát và đánh giá mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên tại các cơ sở giáo dục đại học. Phần lớn các nghiên cứu này thường áp dụng các phương pháp phân tích thống kê truyền thống như thống kê mô tả và hồi quy tuyến tính để xác định mối quan hệ giữa mức độ hài lòng và các yếu tố ảnh hưởng. Chẳng hạn, một số nghiên cứu đã chỉ ra rằng các yếu tố như chuyên môn giảng viên, phương pháp giảng dạy, khả năng giao tiếp và mức độ tương tác đều có tác động đáng kể đến sự hài lòng của sinh viên (Merola et al., 2022; Sohail & Hasan, 2021). Mặt khác, một số công trình khác chú trọng đánh giá chất lượng giảng dạy thông qua mức độ phù hợp của học liệu, sử dụng cơ sở vật chất hiện đại, hoặc vận dụng các chính sách hỗ trợ sinh viên (Giang & Nguyen, 2019) (Zia et al., 2022).

Mặc dù hướng tiếp cận và kỹ thuật phân tích dữ liệu trong các nghiên cứu này tương đối tương đồng, cụ thể là việc sử dụng mô hình hồi quy để tìm kiếm mối liên hệ giữa các biến, chỉ có sự khác biệt về phương pháp chọn mẫu, kích thước mẫu, phạm vi nghiên

cứ và nội dung cụ thể của các biến độc lập. Một số công trình thực hiện trên phạm vi một trường đại học cụ thể, với kích thước mẫu nhỏ, chủ yếu tập trung vào một học phần hoặc một nhóm giảng viên nhất định. Trong khi đó, các nghiên cứu khác áp dụng phương pháp chọn mẫu ngẫu nhiên, với quy mô mẫu lớn hơn và dữ liệu thu thập từ nhiều ngành học, nhiều học kỳ hoặc thậm chí nhiều cơ sở đào tạo khác nhau (Kalim et al., 2022; Pandita & Kiran, 2023). Sự đa dạng này dẫn đến kết luận rằng, mặc dù phân lớn các nghiên cứu đều sử dụng kỹ thuật phân tích thống kê mô tả và hồi quy tuyến tính, chúng vẫn tồn tại sự khác biệt đáng kể về phương pháp chọn mẫu, số lượng mẫu, cũng như yếu tố ảnh hưởng đến sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên. Kết luận này gợi mở hướng đi cho các nghiên cứu tiếp theo trong việc hoàn thiện khung phân tích và tăng cường tính tổng quát hóa, đặc biệt trong bối cảnh áp dụng các phương pháp phân tích nâng cao như học máy (machine learning) (Martínez-Roget et al., 2020).

*1.2.2. Các nghiên cứu về ứng dụng học máy (machine learning) trong giáo dục và đào tạo*

Trong các nghiên cứu trước đây, các phương pháp truyền thống như thống kê mô tả và hồi quy tuyến tính đã được sử dụng rộng rãi để phân tích dữ liệu liên quan đến phản hồi của người học. Tuy nhiên, những phương pháp này thường chỉ xem xét mối quan hệ tuyến tính và không thể khai thác hết tiềm năng của dữ liệu phức tạp trong giáo dục và đào tạo (Jensen & Artz, 2005). Điều này dẫn đến việc bỏ lỡ các thông tin quan trọng có thể giúp cải thiện chất lượng giảng dạy và trải nghiệm học tập của người học (Miller et al., 2015). Chính vì vậy, gần đây đã có những nghiên cứu sử dụng các thuật toán học máy (machine learning) với kì vọng khắc phục những hạn chế của các phương pháp truyền thống. Cụ thể như:

(Kumar et al., 2024) áp dụng thuật toán Random Forest để dự đoán hành vi bỏ học của học sinh. Nghiên cứu của (Maheshwari et al., 2024) chỉ ra rằng các phương pháp tổng hợp như Random Forest (RF) và Gradient Boosting (GB) vượt trội hơn đáng kể các phương pháp thống kê truyền thống trong việc dự đoán kết quả học tập. Zhang &

Zhang, (2024) đã chứng minh trong nghiên cứu của mình rằng: việc áp dụng mô hình như Support Vector Machine trong giáo dục đã cải thiện đáng kể khả năng dự đoán so với các phương pháp thống kê truyền thống. Nghiên cứu của (Fahd et al., 2022) và (Jaiswal et al., 2020) đã sử dụng thuật toán học máy (machine learning) để dự báo kết quả học tập của sinh viên, giúp nhận diện sớm những người có nguy cơ thất bại hoặc bỏ học. Nghiên cứu của (Musso et al., 2020) ứng dụng học máy (machine learning) để dự đoán điểm số và xác suất hoàn thành khóa học.

Như vậy, dù học máy (machine learning) trong lĩnh vực giáo dục và đào tạo đã có những bước tiến đáng kể, nhưng hầu hết các nghiên cứu chỉ tập trung vào việc dự đoán điểm số hoặc khả năng hoàn thành khóa học. Điều này tạo ra một khoảng trống quan trọng, cần được lấp đầy bằng cách khai thác học máy (machine learning) để cải thiện chất lượng giảng dạy thông qua việc hiểu rõ hơn về các yếu tố ảnh hưởng đến sự hài lòng của sinh viên.

Việc áp dụng học máy (machine learning) đã mang lại những bước tiến lớn trong việc phân tích dữ liệu giáo dục. Một số nghiên cứu gần đây cho thấy những mô hình học máy này có khả năng xử lý dữ liệu phức tạp và phi tuyến tính giúp nâng cao độ chính xác trong dự đoán mức độ hài lòng của sinh viên (Mutawa & Sruthi, 2023). Không chỉ dừng lại ở khả năng dự đoán, các mô hình này còn hỗ trợ khám phá mối quan hệ giữa các yếu tố ảnh hưởng, từ đó giúp các trường đại học và giảng viên cải thiện chiến lược giảng dạy một cách hiệu quả hơn.

Việc sử dụng các công cụ như SHAP values hoặc Feature importance giúp làm rõ các yếu tố quan trọng nhất trong mô hình machine learning (Zhong et al., 2021). Những phương pháp này đã được chứng minh là hiệu quả trong việc giải thích và đánh giá tầm quan trọng của từng biến trong dự đoán, đặc biệt trong lĩnh vực giáo dục (Hoyle et al., 2015). Nhờ đó, các nhà giáo dục có thể dễ dàng nhận diện chính xác những yếu tố ảnh hưởng lớn nhất đến sự hài lòng của sinh viên, từ đó đưa ra các điều chỉnh phù hợp nhằm tối ưu hóa chất lượng giảng dạy và trải nghiệm học tập.

Ngoài ra, hiện vẫn rất ít nghiên cứu thực hiện so sánh toàn diện hiệu suất của các mô hình học máy (machine learning) khác nhau

trong việc dự đoán và phân tích các yếu tố ảnh hưởng đến sự hài lòng của sinh viên. Một số nghiên cứu đã áp dụng các thuật toán như Random Forest, Boosting Regression hoặc Neural Networks để dự đoán mức độ hài lòng nhưng phân lớn chỉ dừng lại ở việc sử dụng một mô hình cụ thể mà không tiến hành đánh giá chi tiết giữa các mô hình (Simsekler et al., 2021); Điều này dẫn đến việc khó xác định mô hình nào hiệu quả nhất trong cả dự đoán lẫn phân tích yếu tố quan trọng.

Từ đó, nghiên cứu của nhóm tác giả tập trung trả lời hai câu hỏi chính:

*RQ1: Mô hình học máy (machine learning) nào dự đoán chính xác nhất mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên?*

*RQ2: Yếu tố nào có ảnh hưởng lớn nhất đến mức độ hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên theo các mô hình học máy (machine learning)?*

### **2. Quy trình và phương pháp nghiên cứu**

Hình 1 minh họa quy trình nghiên cứu về dự đoán sự hài lòng của sinh viên thông qua việc sử dụng các phương pháp học máy (machine learning) đa dạng. Bảy phương pháp machine learning gồm: BR (Boosting Regression), DCR (Decision Tree Regression), KNNR (K-Nearest Neighbors Regression), LR (Linear Regression), RFR (Random Forest Regression), RLR (Regularized Linear Regression), SVMR (Support Vector Machine Regression) được áp dụng để xử lý dữ liệu về sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên nhằm đảm bảo tính toàn diện và độ chính xác cao.

Cụ thể quy trình nghiên cứu về dự đoán sự hài lòng của sinh viên thông qua việc sử dụng các phương pháp học máy (machine learning) được mô tả theo các giai đoạn sau:

#### **2.1. Giai đoạn tiền xử lý dữ liệu (Data Preprocessing)**

Để các thuật toán machine learning hoạt động hiệu quả, giai đoạn tiền xử lý dữ liệu là một bước quan trọng trong quy trình nghiên cứu (Tanasa & Trousse, 2004). Giai đoạn này được thực hiện thông qua các bước công việc (hình 2) như sau:

##### *(1) Thu thập dữ liệu*

Dữ liệu được thu thập từ 3.477 phiếu hỏi online thông qua link điều tra hoặc phiếu hỏi



(bản cứng) gửi trực tiếp vào buổi học cuối cùng của mỗi lớp học phân học kì 1 và 2 năm học 2023-2024 tại Trường Đại học Thương mại. Trước khi thực hiện khảo sát, sinh viên đã được thông báo về mục đích của khảo sát, quyền lợi và nghĩa vụ của người tham gia. Các lớp học phân lấy mẫu điều tra được mỗi bộ môn chọn ngẫu nhiên cho từng giảng viên trong từng học kì. Việc lựa chọn ngẫu nhiên này nhằm đảm bảo tính đa dạng của mẫu nghiên cứu, bao gồm sinh viên từ các lĩnh vực đào tạo khác nhau. Mẫu phiếu hỏi được cung cấp từ phòng Khảo thí và đảm bảo chất lượng giáo dục (KT&ĐBCLGD), vì vậy nội dung phiếu hỏi là như nhau cho tất cả các lớp học phân. Các dữ liệu thu thập sẽ được tổng hợp thành một file dữ liệu chung.

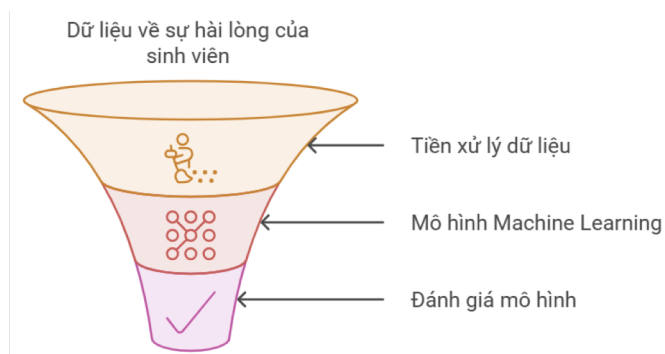
### (2) Chuẩn hoá dữ liệu

Để chuẩn hoá dữ liệu, đầu tiên các giá trị thiếu trong các tiêu chí định lượng được thay

thế bằng giá trị trung bình tương ứng, giúp duy trì tính toàn vẹn của dữ liệu mà không làm sai lệch phân phối ban đầu. Tiếp theo, các biến thông thường (ordinal) được chuẩn hóa để cải thiện hiệu suất mô hình, việc này giúp cân bằng thang đo giữa các biến và giảm thiểu ảnh hưởng của các giá trị ngoại lệ (outliers) (Manuel Vargas et al., 2022). Cuối cùng, các dữ liệu bất thường được phát hiện và loại bỏ, đảm bảo rằng mô hình không bị ảnh hưởng tiêu cực bởi những giá trị không đại diện cho xu hướng chung của dữ liệu.

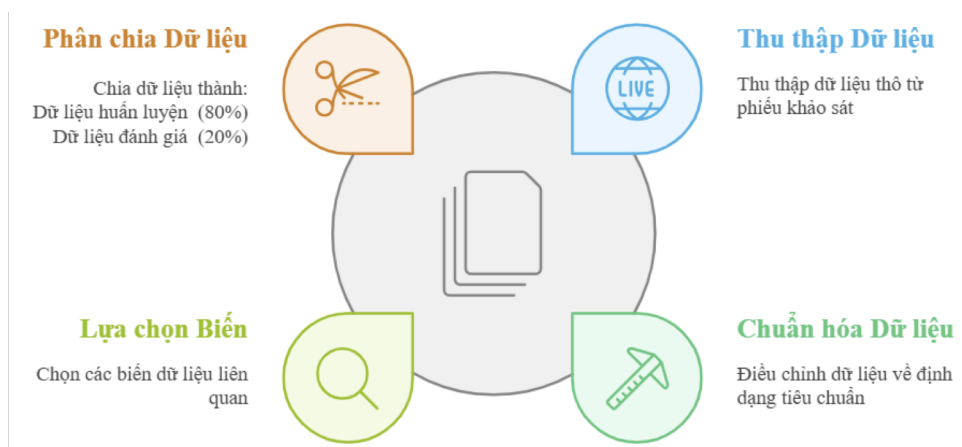
### (3) Lựa chọn biến

Trong nghiên cứu này, nhóm tác giả sử dụng các tiêu chí từ câu 1 đến câu 21 làm biến đầu vào (X) và câu 22 (Đánh giá mức độ hài lòng chung của sinh viên) làm biến đầu ra (Y) (bảng 1). Việc lựa chọn này nhằm tập trung vào việc dự đoán mức độ hài lòng dựa trên các yếu tố định lượng đã thu thập cho phép mô



(Nguồn: Nhóm tác giả)

**Hình 1:** Quy trình nghiên cứu đề xuất



(Nguồn: Theo (Tanasa & Trousse, 2004))

**Hình 2:** Các bước xử lý dữ liệu

# Ý KIẾN TRAO ĐỔI

**Bảng 1:** Các tiêu chí định lượng

STT	Các tiêu chí định lượng (X)	Biến mục tiêu (Y)
1	Chuẩn bị nội dung, slide bài giảng	Đánh giá chung mức độ hài lòng về chất lượng và hiệu quả giảng dạy của giảng viên
2	Chuẩn bị phương tiện, học liệu phục vụ giảng dạy	
3	Việc giới thiệu đầy đủ thông tin chi tiết học phần	
4	Công bố rõ ràng (từ đầu) mục tiêu, chuẩn đầu ra, nội dung, kế hoạch giảng dạy, phương pháp đánh giá, tỷ trọng điểm quá trình học tập	
5	Nội dung bài giảng phù hợp với đề cương của học phần	
6	Nội dung giảng dạy thiết thực, cập nhật, giàu tính thực tiễn	
7	Giảng viên có kiến thức chuyên môn sâu, rộng về nội dung giảng dạy	
8	Giao đề tài, giao nhiệm vụ và chia nhóm thảo luận/Thực hành phù hợp	
9	Tổ chức thảo luận/Thực hành đảm bảo đúng quy định	
10	Hướng dẫn, khuyến khích người học trao đổi, tương tác	
11	Phổ biến hình thức và yêu cầu về việc đánh giá một cách rõ ràng	
12	Việc đánh giá phù hợp với hình thức học tập, mục tiêu và chuẩn đầu ra của học phần	
13	Việc đánh giá phù hợp với nội dung giảng dạy học phần và phản ánh đúng năng lực của người học	
14	Nghiêm túc, khách quan và công bằng trong việc đánh giá	
15	Thái độ ứng xử, trang phục lên lớp lịch sự, phù hợp	
16	Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu	
17	Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả	
18	Mức độ tương tác với người học ở trên lớp (đặt câu hỏi, trao đổi hai chiều)	
19	Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu	
20	Duy trì nghiêm túc thời khoá biểu, vào lớp và tan lớp đúng giờ	
21	Khả năng bao quát và kiểm soát học tập của người học trên lớp	

(Nguồn: Phòng KT và ĐBCLGD, Trường Đại học Thương mại (2024))

hình xác định mối quan hệ giữa các tiêu chí cụ thể và sự hài lòng tổng thể của sinh viên.

#### (4) Phân chia dữ liệu

Sau khi lựa chọn biến, các dữ liệu được chia thành hai tập: tập huấn luyện (training set), chiếm 80% và tập kiểm tra (testing set) chiếm 20% (Bảng 2). Tập huấn luyện được sử dụng để xây dựng và tối ưu hóa mô hình

machine learning, trong khi tập kiểm tra được dùng để đánh giá hiệu suất và khả năng tổng quát hóa của mô hình trên dữ liệu chưa từng thấy trước đó (Acampora et al., 2018). Việc phân chia này đảm bảo rằng mô hình học máy sẽ tìm được các đặc trưng quan trọng từ dữ liệu huấn luyện và có thể áp dụng hiệu quả trên dữ liệu thực tế.

**Bảng 2:** Phân chia dữ liệu

	Dữ liệu đầu vào cho việc huấn luyện	Dữ liệu kiểm tra
Kích thước dữ liệu	2.781	696

(Nguồn: Tổng hợp của nhóm tác giả từ mô hình học máy)

## 2.2. Giai đoạn sử dụng các mô hình học máy (machine learning)

Sử dụng các mô hình học máy (machine learning) là một kỹ thuật đang phát triển mạnh mẽ trong lĩnh vực trí tuệ nhân tạo, các tập dữ liệu lớn được các mô hình học máy (machine learning) sử dụng để dự đoán kết quả và hỗ trợ quá trình ra quyết định dựa trên dữ liệu (Ray, 2019). Machine learning bao gồm nhiều phương pháp: học máy có giám sát, học máy không giám sát, học máy bán giám sát và học máy tăng cường (Ahuja et al., 2020). Mỗi phương pháp học máy (machine learning) phục vụ các mục đích khác nhau; trong nghiên cứu này, các tác giả tập trung vào mô hình học máy có giám sát vì nó phù hợp với bộ dữ liệu thu thập là bộ dữ liệu về giáo dục và khoa học xã hội. Nghiên cứu sử dụng phần mềm mã nguồn mở JASP (phiên bản 0.18.1) để triển khai bảy mô hình machine learning có giám sát. Phần mềm này hiện được sử dụng tại 342 trường đại học trên toàn thế giới, trải dài trên 72 quốc gia minh chứng cho tính ứng dụng và độ tin cậy trong các nghiên cứu giáo dục và khoa học xã hội. Nhiều nghiên cứu đã ứng dụng JASP trong phân tích dữ liệu giáo dục và các lĩnh vực liên quan (Li, 2021; Urbano-Cano et al., 2024; Vasavada, 2023). Dưới đây là giới thiệu sơ lược về bảy mô hình machine learning được áp dụng để xử lý dữ liệu nghiên cứu của nhóm tác giả:

### (1) Mô hình Hồi quy tăng cường (Boosting Regression - BR)

Mô hình Hồi quy tăng cường (BR) được giới thiệu bởi Elith và cộng sự (2008) sử dụng các mô hình thống kê sinh thái để giải thích và dự đoán các biến của mô hình. Mô hình BR có các đặc điểm như: biến phi tuyến, giải thích sự tương tác giữa các biến và dự đoán mối quan hệ giữa các biến một cách hiệu quả. Mô hình Hồi quy tăng cường hoạt động bằng cách liên tiếp khớp một loạt các cây quyết định vào dữ liệu, trong đó mỗi cây cố gắng sửa lỗi của cây trước đó. Quá trình hồi quy tăng cường được thực hiện cho đến khi mô hình đạt được mức độ chính xác đã xác định. Khi kiểm tra dữ liệu, mô hình Hồi quy tăng cường tinh chỉnh dự đoán bằng cách đặt trọng số lớn hơn cho các điểm dữ liệu trước đó bị phân loại sai hoặc dự đoán kém.

### (2) Mô hình K-Nearest Neighbors Regression (KNNR)

Mô hình K-Nearest Neighbors Regression (KNNR) lưu trữ tất cả các mẫu dữ liệu được đưa vào (tạm gọi là: dữ liệu huấn luyện) trong không gian “n” chiều. Mô hình phân loại các điểm dữ liệu mới bằng cách đánh giá sự tương đồng, thường sử dụng các thước đo như khoảng cách hình học “Euclid”. Phân loại dữ liệu huấn luyện dựa trên biểu quyết đa số từ k-neighbor gần nhất của mỗi điểm dữ liệu. Mô hình KNNR phân biệt và xử lý được “dữ liệu huấn luyện” nhiều, với độ chính xác phụ thuộc vào chất lượng dữ liệu. Mô hình KNNR phục vụ cho cả nhiệm vụ phân loại và hồi quy (Sarker, 2021). Khi kiểm tra dữ liệu, thuật toán KNNR tính khoảng cách giữa điểm dữ liệu kiểm tra và tất cả các điểm trong tập huấn luyện, sau đó xác định ‘k’ neighbors gần nhất để dự đoán.

### (3) Mô hình Decision Tree Regression (DTR)

Mô hình Decision Tree Regression (DTR) là một kỹ thuật học máy có giám sát phi tham số được sử dụng rộng rãi cho các nhiệm vụ phân loại và hồi quy (Song & Ying, 2015). Mô hình này được ví như việc ra quyết định theo mô hình của cây: Khi dữ liệu kiểm tra được đưa vào, mỗi mẫu được phân loại bằng cách bắt đầu từ “nút” gốc cây, nơi một thuộc tính cụ thể được đánh giá. Giá trị thuộc tính của dữ liệu kiểm tra xác định nhánh nào của cây sẽ theo “nút” gốc. Quá trình này tiếp tục với mỗi “nút” tiếp theo đánh giá một thuộc tính khác, cho đến khi đạt đến “nút” lá. “Nút” lá đại diện cho lớp hoặc giá trị được dự đoán cho mẫu đó.

### (4) Mô hình Random Forest Regression (RFR)

Mô hình Random Forest Regression (RFR) sử dụng “tập hợp song song”, đồng thời khớp nhiều cây quyết định trên các mẫu con khác nhau của tập dữ liệu. Khi dữ liệu kiểm tra được đưa vào, mỗi cây quyết định trong rừng độc lập phân loại các mẫu kiểm tra hoặc dự đoán giá trị của chúng. Kết quả được xác định thông qua phần mềm để “biểu quyết” đa số hoặc trung bình, giảm thiểu vấn đề quá khớp và nâng cao độ chính xác dự đoán. Mô hình này bao gồm nhiều cây quyết định, thường vượt trội so với mô hình dựa trên một cây quyết định đơn lẻ. Mô hình RFR đa dụng, phù hợp cho cả nhiệm vụ phân loại

và hồi quy, và xử lý tốt cả giá trị phân loại và liên tục (Ziegler & König, 2014).

(5) *Mô hình Support Vector Machine Regression (SVMR)*

Mô hình Support Vector Machine (SVMR) là một kỹ thuật học máy (machine learning) phổ biến cho các nhiệm vụ như phân loại và hồi quy (Cervantes et al., 2020). Trong không gian nhiều chiều, mô hình SVMR xây dựng các siêu phẳng đạt được sự phân tách tốt bằng cách tối đa hóa khoảng cách từ các điểm dữ liệu huấn luyện gần nhất của bất kỳ lớp nào. Nó hiệu quả trong nhiều tình huống và có thể thể hiện hành vi khác nhau dựa trên các hàm toán học khác nhau, được gọi là kernel (hạt nhân). Các kernel phổ biến bao gồm tuyến tính, đa thức, hàm cơ sở xuyên tâm (RBF),.... Trong các nhiệm vụ hồi quy, mô hình SVMR dự đoán các giá trị liên tục bằng cách đặt dữ liệu kiểm tra trong các biên được xác định bởi các véc-tơ hỗ trợ.

(6) *Mô hình Hồi quy tuyến tính có điều kiện chuẩn (Regularized Linear Regression - RLR)*

Mô hình Hồi quy tuyến tính có điều kiện chuẩn (Regularized Linear Regression - RLR) là một kỹ thuật chia các biên thành nhóm và cải thiện độ chính xác dự đoán bằng cách khuyến khích sự thưa thớt trong mỗi nhóm và sự đa dạng giữa chúng (Christidis et al., 2020). Trong quá trình huấn luyện, việc điều chỉnh các hệ số lớn để ngăn chặn sự quá khớp. Khi kiểm tra, mô hình sử dụng các trọng số đã được học (máy học) để dự đoán kết quả cho dữ liệu mới, đảm bảo rằng nó vẫn tổng quát và tránh nhạy cảm quá mức với các biên đôi nhỏ. Quá trình này cải thiện hiệu suất của mô hình trên dữ liệu kiểm tra bằng cách giảm sự quá khớp và cung cấp dự đoán chính xác hơn (Reddy et al., 2020).

(7) *Mô hình Hồi quy tuyến tính (Linear Regression - LR)*

Mô hình Hồi quy tuyến tính (Linear Regression - LR) là một kỹ thuật thông kê cơ bản để mô hình hóa mối quan hệ giữa biến phụ thuộc và một hoặc nhiều biến độc lập (Akritas & Bershad, 1996). Mục tiêu của hồi quy tuyến tính là tìm ra đường thẳng tốt nhất khớp với dữ liệu bằng cách tối thiểu hóa tổng bình phương sai số giữa giá trị dự đoán và giá trị thực tế. Trong quá trình kiểm tra, mô hình sử dụng các tham số đã được ước lượng từ dữ liệu huấn luyện để dự đoán giá trị biến phụ thuộc

cho các mẫu mới. Mặc dù đơn giản, mô hình hồi quy tuyến tính cung cấp một cơ sở quan trọng để so sánh với các mô hình phức tạp hơn và giúp hiểu rõ mối quan hệ tuyến tính giữa các biến trong dữ liệu (Lichti et al., 2021).

**2.3. Giai đoạn đánh giá các mô hình học máy (machine learning)**

Nhóm tác giả sử dụng năm thước đo đánh giá phổ biến gồm MSE, RMSE, MAE, MAPE và  $R^2$  để đánh giá hiệu quả của kết quả dự đoán và so sánh các kỹ thuật machine learning khác nhau. Theo nghiên cứu của Chai & Draxler (2014) và Willmott & Matsuura (2005) giá trị sai số của trung bình bình phương lỗi (MSE) và căn bậc hai trung bình bình phương lỗi (RMSE) phản ánh mức độ chênh lệch giữa giá trị thực tế và giá trị dự đoán. Lỗi tuyệt đối trung bình (MAE) đo lường sai số trung bình không phụ thuộc vào độ lớn của giá trị, trong khi lỗi phần trăm tuyệt đối trung bình (MAPE) cung cấp tỷ lệ phần trăm sai số so với giá trị thực tế. Ngoài ra, hệ số xác định ( $R^2$ ) đánh giá khả năng giải thích phương sai của mô hình. Các nghiên cứu chỉ ra rằng các giá trị lỗi nhỏ hơn và giá trị  $R^2$  cao hơn biểu thị hiệu suất mô hình tốt hơn.

MSE và RMSE cung cấp thông tin về mức độ sai lệch tổng quát và được sử dụng rộng rãi trong các bài toán hồi quy. MAE và MAPE là các thước đo bổ sung giúp đánh giá mức độ chính xác trong các bối cảnh có các giá trị thực khác biệt về quy mô. Hệ số  $R^2$  đóng vai trò quan trọng trong việc xác định mức độ giải thích phương sai của biến phụ thuộc bởi biến độc lập. Việc sử dụng đồng thời các thước đo này mang lại cái nhìn toàn diện hơn về hiệu quả của mô hình, giúp lựa chọn phương pháp học máy phù hợp nhất.

**3. Kết quả nghiên cứu và thảo luận**

**3.1. Đặc điểm mẫu nghiên cứu**

Dữ liệu được tổng hợp từ khảo sát định kỳ hàng năm đối với 3.477 sinh viên về chất lượng giảng dạy của 38 giảng viên giảng dạy 20 học phần được chọn ngẫu nhiên từ khoa Quản trị nhân lực, Viện Quản trị kinh doanh và Khoa Toán Kinh tế - Trường Đại học Thương mại, năm học 2023-2024. Dữ liệu bao gồm 21 tiêu chí đánh giá chi tiết về chất lượng giảng dạy, cung cấp cái nhìn tổng thể và đa chiều về sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên (Bảng 1). 21 câu hỏi sử dụng thang đo Likert 5 điểm (1 = Rất không hài lòng,



2 = Không hài lòng, 3 = Bình thường, 4 = Hài lòng, 5 = Rất hài lòng) để đánh giá các khía cạnh khác nhau của hoạt động giảng dạy. Câu hỏi thứ 22 là câu hỏi mở, yêu cầu sinh viên đánh giá chung mức độ hài lòng của mình đối với hoạt động giảng dạy của giảng viên trong học phần.

**3.2. Đánh giá mô hình đo lường**

Để đảm bảo tính khoa học và độ tin cậy của bộ thang đo về mức độ hài lòng của sinh viên, nghiên cứu đã tiến hành kiểm định hệ số tin cậy Cronbach's Alpha và phân tích nhân tố khám phá (EFA). Phân tích EFA được thực hiện trên 21 biến tiêu chí đánh giá hoạt động giảng dạy, sử dụng phương pháp trích xuất Principal Component Analysis. Kết quả kiểm định KMO và Bartlett's Test, cho thấy giá trị KMO là 0.978, một giá trị rất cao (> 0.9) và cho thấy mẫu có tính đại diện rất tốt (Nkansah, 2018). Kiểm định Bartlett's Test cũng có ý nghĩa thống kê (df = 210, p < 0.001), bác bỏ giả thuyết không có mối tương quan giữa các biến quan sát (Knapp & Swoyer, 1967). Các kết quả này khẳng định dữ liệu hoàn toàn phù hợp để thực hiện phân tích nhân tố (Bảng 3).

“Sự hài lòng của sinh viên đối với hoạt động giảng dạy của giảng viên” và thang đo có độ tin cậy nội bộ tốt. Điều này khẳng định tính khoa học và khách quan của thang đo. Dữ liệu thu thập từ phiếu khảo sát này là phù hợp để sử dụng trong nghiên cứu và sử dụng trong các bước phân tích tiếp theo như áp dụng các mô hình học máy để dự đoán sự hài lòng của sinh viên.

**3.3. Hiệu suất của các mô hình (RQ1)**

Các kết quả được đánh giá dựa trên năm thước đo chính: MSE, RMSE, MAE/MAD, MAPE và R<sup>2</sup>. Mô hình hồi quy tuyến tính (LR) thể hiện hiệu suất tốt nhất với MSE thấp nhất (0.103), RMSE là 0.321, và MAE là 0.159. MAPE của LR (4.30%) cũng thuộc nhóm thấp nhất, cho thấy sai số dự đoán chỉ chiếm một phần nhỏ so với giá trị thực tế. R<sup>2</sup> của LR đạt 0.687, cao nhất trong các mô hình, chứng minh khả năng giải thích 68.7% biến thiên trong dữ liệu. Hồi quy tuyến tính có điều chuẩn (RLR) và Hồi quy Rừng ngẫu nhiên (RFR) có hiệu suất tương đương, với MSE lần lượt là 0.104 và 0.106, RMSE khoảng 0.322-0.326 và MAE là 0.159. MAPE

**Bảng 3:** Kiểm định KMO và Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.978
Bartlett's Test of Sphericity	Approx. Chi-Square	59746.800
	df	210
	Sig.	.000

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả)

Bảng 4 cho thấy tất cả 21 biến đều có tải nhân tố (factor loading) rất cao lên nhân tố duy nhất này (tất cả đều lớn hơn 0.689), Điều này một lần nữa khẳng định rằng các tiêu chí trong phiếu khảo sát đều đo lường cùng một khái niệm chung.

Hệ số Cronbach's Alpha được sử dụng để đánh giá độ tin cậy nội bộ của thang đo. Kết quả phân tích cho thấy hệ số Cronbach's Alpha của thang đo tổng thể là 0,969 (Bảng 5). Giá trị này lớn hơn 0.9, cho thấy thang đo có độ tin cậy nội bộ rất tốt (Kilic, 2016).

Dựa trên kết quả phân tích EFA và Cronbach's Alpha, thang đo trong phiếu khảo sát có tính giá trị và độ tin cậy cao. Các tiêu chí trong phiếu khảo sát đều đo lường một khái niệm chung là

của RLR (4.31%) thấp hơn so với RFR (4.36%). R<sup>2</sup> của RLR (0.685) và RF (0.678) cũng khá cạnh tranh, khẳng định rằng hai mô hình này phù hợp cho việc dự đoán mức độ hài lòng của sinh viên (Bảng 6).

Hồi quy tăng cường (BR) cũng cho thấy hiệu suất cạnh tranh, với MSE là 0.105, RMSE là 0.324, và MAE là 0.17. MAPE của BR đạt 4.58%, tương đối thấp, và R<sup>2</sup> là 0.683, chỉ kém nhẹ so với LR. Những chỉ số này khẳng định rằng BR là một mô hình mạnh, phù hợp với dữ liệu này. Hồi quy k-láng giềng gần nhất (KNNR) xếp hạng ba với MSE là 0.117, RMSE là 0.342, và MAE là 0.143. MAPE của KNN (4.09%) thấp nhất trong tất cả các mô hình, chứng minh khả năng dự

## Ý KIẾN TRAO ĐỔI

**Bảng 4:** Hệ số tải

Tiêu chí	Hệ số tải	Tiêu chí	Hệ số tải
12. Việc đánh giá phù hợp với hình thức học tập, mục tiêu và chuẩn đầu ra của học phần	0.840	21. Khả năng bao quát và kiểm soát học tập của người học trên lớp	0.787
13. Việc đánh giá phù hợp với nội dung giảng dạy học phần và phản ánh đúng năng lực của người học	0.835	6. Nội dung giảng dạy thiết thực, cập nhật, giàu tính thực tiễn	0.774
5. Nội dung bài giảng phù hợp với đề cương của học phần	0.821	4. Công bố rõ ràng (từ đầu) mục tiêu, chuẩn đầu ra, nội dung, kế hoạch giảng dạy, phương pháp đánh giá, tỷ trọng điểm quá trình học tập	0.774
11. Phổ biến hình thức và yêu cầu về việc đánh giá một cách rõ ràng	0.819	18. Mức độ tương tác với người học ở trên lớp (đặt câu hỏi, trao đổi hai chiều)	0.773
17. Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả	0.813	8. Giao đề tài, giao nhiệm vụ và chia nhóm thảo luận/Thực hành phù hợp	0.770
14. Nghiêm túc, khách quan và công bằng trong việc đánh giá	0.810	3. Việc giới thiệu đầy đủ thông tin chi tiết học phần	0.770
19. Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu	0.802	2. Chuẩn bị phương tiện, học liệu phục vụ giảng dạy	0.770
16. Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu	0.801	7. Giảng viên có kiến thức chuyên môn sâu, rộng về nội dung giảng dạy	0.766
10. Hướng dẫn, khuyến khích người học trao đổi, tương tác	0.797	1. Chuẩn bị nội dung, slide bài giảng	0.758
9. Tổ chức thảo luận/Thực hành đảm bảo đúng quy định	0.796	15. Thái độ ứng xử, trang phục lên lớp lịch sự, phù hợp	0.706
		20. Duy trì nghiêm túc thời khoá biểu, vào lớp và tan lớp đúng giờ	0.689

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả)

đoán chính xác. Tuy nhiên,  $R^2$  của KNN (0.655) thấp hơn một chút so với các mô hình hàng đầu. Hồi quy máy vector hỗ trợ (SVMR) đạt MSE là 0.112, RMSE là 0.335, và MAE là 0.161. MAPE của SVM (4.39%) tương đối thấp, nhưng  $R^2$  (0.661) cho thấy mô hình này chưa giải thích được phương sai tốt như LR,

BR, hoặc RLR. Hồi quy cây quyết định (DTR) có MSE là 0.14, RMSE là 0.374 và  $R^2$  là 0.599, thấp nhất trong các mô hình. Các chỉ số như MAPE (4.73%) và MAE (0.171) cho thấy mô hình này phù hợp ở mức trung bình nhưng không phải là lựa chọn tốt nhất.

**Bảng 5:** Hệ số Cronbach's Alpha

Reliability Statistics	
Cronbach's Alpha	N of Items
.969	21

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả)

**Bảng 6:** Kết quả hiệu suất của các mô hình

	Mô hình BR	Mô hình DTR	Mô hình KNNR	Mô hình LR	Mô hình RFR	Mô hình RLR	Mô hình SVMR
MSE	0.105	0.14	0.117	<b>0.103</b>	0.106	0.104	0.112
RMSE	0.324	0.374	0.342	<b>0.321</b>	0.326	0.322	0.335
MAE/ MAD	0.17	0.171	0.143	<b>0.159</b>	0.159	0.159	0.161
MAPE	4.58%	4.73%	<b>4.09%</b>	4.30%	4.36%	4.31%	4.39%
R2	0.683	0.599	0.655	<b>0.687</b>	0.678	0.685	0.661

(Nguồn: Tổng hợp của nhóm tác giả từ mô hình học máy)

Dựa trên kết quả phân tích, Hồi quy tuyến tính (LR) được xác định là mô hình dự đoán tốt nhất và cho thấy hiệu suất vượt trội trong việc dự đoán mức độ hài lòng của sinh viên đối với hoạt động giảng dạy tại Trường Đại học Thương mại. Với MSE thấp nhất (0.103), R<sup>2</sup> cao nhất (0.687) và các chỉ số sai số rất thấp. Những kết quả này cho thấy mô hình có khả năng dự đoán chính xác và giải thích tốt nhất phương sai của dữ liệu... Những chỉ số này cho thấy mô hình không chỉ có độ chính xác dự đoán cao mà còn giải thích được 68.7% biến thiên của dữ liệu. So với các mô hình khác, mô hình Hồi quy tuyến tính (LR) kết hợp sự đơn giản và tính hiệu quả làm cho nó trở thành lựa chọn tối ưu để dự đoán mức độ hài lòng của sinh viên trong nghiên cứu này.

Mặc dù một số nghiên cứu gần đây đã đề xuất áp dụng các thuật toán machine learning tiên tiến (như Random Forest, Boosted Regression, Neural Networks) để xử lý dữ liệu phân loại phức tạp hơn, qua đó phát hiện các mẫu ẩn và mối quan hệ phi tuyến (Yağcı, 2022), kết quả nghiên cứu cho thấy rằng trong bối cảnh cụ thể này, LR vẫn có thể vượt trội về độ chính xác và khả năng giải thích mô hình. Điều này hàm ý rằng không phải lúc nào các mô hình machine learning phức tạp hơn cũng cho hiệu quả cao hơn. Thay vào đó, lựa chọn mô hình tối ưu còn phụ thuộc vào đặc điểm dữ liệu, mục tiêu phân tích và các biên số cụ thể trong bối cảnh nghiên cứu.

### 3.4. Phân tích yếu tố quan trọng (RQ2)

Để trả lời câu hỏi nghiên cứu 2 (Tiêu chí ảnh hưởng lớn nhất đến mức độ hài lòng của sinh viên?), tác giả phân tích Feature Importance từ hai mô hình phù hợp nhất: Hồi quy tuyến tính (LR) và Hồi quy tuyến tính có điều chuẩn (RLR) (bảng 7). Cả hai mô hình

được chọn vì chúng đạt hiệu suất dự đoán cao nhất trong đánh giá mô hình (bảng 6). Các giá trị Tổng thất trung bình khi loại bỏ được sử dụng để xác định các tiêu chí quan trọng, với giá trị cao hơn thể hiện tầm quan trọng lớn hơn trong việc dự đoán mức độ hài lòng của sinh viên.

Bảng 7 cho ta thấy kết quả 5 tiêu chí quan trọng nhất của từng mô hình. Cả hai mô hình đều cho thấy sự đồng nhất trong việc xác định các yếu tố quan trọng. Trong đó, bốn tiêu chí xuất hiện đồng thời ở cả hai mô hình với giá trị Tổng thất trung bình khi loại bỏ cao gồm:

- (16) Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu: Đây là yếu tố quan trọng nhất trong cả hai mô hình, với giá trị 0.365 (LR) và 0.358 (RLR).

- (19) Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu: Yếu tố này đứng thứ hai ở cả hai mô hình, với giá trị 0.328 (LR) và 0.335 (RLR).

- (14) Nghiêm túc, khách quan và công bằng trong việc đánh giá với giá trị 0.325 (LR) và 0.324 (RLR).

- (17) Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả với giá trị 0.327 (LR) và 0.328 (RLR).

Ngoài ra, một số tiêu chí quan trọng khác chỉ xuất hiện trong mô hình: Hồi quy tuyến tính (LR): (21) Khả năng bao quát và kiểm soát học tập của người học trên lớp (0.327). Hồi quy tuyến tính có điều chuẩn (RLR): (7) Giảng viên có kiến thức chuyên môn sâu, rộng về nội dung giảng dạy (0.324).

Các kết quả này cung cấp một góc nhìn toàn diện hơn về các yếu tố ảnh hưởng. Dựa trên phân tích từ hai mô hình, tác giả đưa ra top 5 yếu tố ảnh hưởng lớn nhất đến sự hài

**Bảng 7:** Các yếu tố quan trọng ảnh hưởng nhất đến mô hình

Mô hình hồi quy tuyến tính (LR - Linear Regression)		Mô hình hồi quy tuyến tính có điều chuẩn (RLR - Regularized Linear Regression)	
Chỉ số quan trọng của đặc trưng	Tồn thất trung bình khi loại bỏ	Chỉ số quan trọng của đặc trưng	Tồn thất trung bình khi loại bỏ
(16) Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu	0.365	(16) Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu	0.358
(19) Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu	0.328	(19) Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu	0.335
(21) Khả năng bao quát và kiểm soát học tập của người học trên lớp	0.327	(17) Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả	0.328
(17) Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả	0.327	(14) Nghiêm túc, khách quan và công bằng trong việc đánh giá	0.325
(14) Nghiêm túc, khách quan và công bằng trong việc đánh giá	0.325	(7) Giảng viên có kiến thức chuyên môn sâu, rộng về nội dung giảng dạy	0.324

(Nguồn: Tổng hợp của nhóm tác giả từ mô hình học máy)

**Bảng 8:** Yếu tố ảnh hưởng lớn nhất đến sự hài lòng của sinh viên

Xếp hạng	Yếu tố
1	Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu
2	Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu
3	Khả năng bao quát và kiểm soát học tập của người học trên lớp
4	Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả
5	Nghiêm túc, khách quan và công bằng trong việc đánh giá

(Nguồn: Tổng hợp của nhóm tác giả từ mô hình học máy)

lòng của sinh viên (bảng 8). Các yếu tố này có thể được coi là trọng tâm để cải thiện chất lượng giảng dạy, dựa trên sự đồng thuận từ các mô hình machine learning.

### 3.5. Thảo luận

Kết quả nghiên cứu cho thấy các mô hình machine learning có khả năng dự đoán chính xác mức độ hài lòng của sinh viên. Đặc biệt, mô hình Hồi quy tuyến tính (LR) đạt hiệu suất tốt nhất với MSE thấp nhất (0.103),  $R^2$  cao nhất (0.687) và các chỉ số sai số rất thấp. Điều này khẳng định hiệu quả của việc áp dụng

machine learning trong phân tích dữ liệu khảo sát sinh viên vượt trội so với các phương pháp thống kê truyền thống. Hơn nữa, mô hình còn cho phép định lượng mức độ ảnh hưởng của từng yếu tố đến sự hài lòng. Điều này hỗ trợ các nhà giáo dục hiểu sâu hơn về nhu cầu và mong muốn của sinh viên, từ đó điều chỉnh và cải thiện phương pháp giảng dạy một cách hiệu quả. Việc áp dụng các mô hình machine learning như vậy góp phần nâng cao chất lượng giáo dục và trải nghiệm học tập, đáp ứng tốt hơn kỳ vọng của sinh viên.



Lý do mô hình này phù hợp với bối cảnh nghiên cứu tại Trường đại học Thương mại có thể được giải thích bởi bốn yếu tố chính. Thứ nhất, mô hình Hồi quy tuyến tính (LR) giả định có mối quan hệ tuyến tính giữa các tiêu chí đánh giá hoạt động giảng dạy và mức độ hài lòng của sinh viên. Kết quả nghiên cứu cho thấy giả định này là hợp lý trong bối cảnh dữ liệu khảo sát tại Trường đại học Thương mại, tức là sự thay đổi của các tiêu chí giảng dạy (ví dụ: cải thiện phương pháp giảng dạy, tăng cường tương tác với sinh viên) có xu hướng dẫn đến sự thay đổi tương ứng và tỷ lệ thuận trong mức độ hài lòng của sinh viên. Thứ hai, đặc thù của bộ dữ liệu tại Trường đại học Thương mại khiến mô hình Hồi quy tuyến tính (LR) trở thành lựa chọn phù hợp. Phiếu khảo sát sử dụng thang đo Likert 5 điểm, tạo ra dữ liệu có tính gần như liên tục, giúp mô hình Hồi quy tuyến tính (LR) hoạt động hiệu quả hơn so với các mô hình như DTR hoặc RFR. Những mô hình này vốn có lợi thế hơn khi xử lý dữ liệu phân loại hoặc phi tuyến tính. Thứ ba, hồi quy tuyến tính có tính đơn giản và dễ diễn giải, giúp giảng viên và nhà quản lý dễ dàng hiểu được mức độ ảnh hưởng của từng tiêu chí đánh giá đến sự hài lòng của sinh viên. Các hệ số hồi quy cho thấy yếu tố nào có tác động mạnh nhất, từ đó hỗ trợ việc ra quyết định trong cải thiện chất lượng giảng dạy. Đây là một ưu điểm quan trọng so với các mô hình học máy phức tạp hơn, vốn thường khó diễn giải và yêu cầu nhiều tài nguyên tính toán hơn. Thứ tư, đặc điểm văn hóa và môi trường giáo dục tại Trường đại học Thương mại có thể khiến mối quan hệ giữa các yếu tố ảnh hưởng và mức độ hài lòng của sinh viên trở nên tuyến tính hơn so với các trường đại học khác. Sự đồng nhất trong phương pháp giảng dạy của giảng viên và sự tương đồng trong kỳ vọng của sinh viên cũng là yếu tố góp phần vào tính phù hợp của mô hình này.

Việc xác định các yếu tố quan trọng nhất giúp giảng viên tập trung cải thiện phương pháp giảng dạy một cách hiệu quả. Năm yếu tố ảnh hưởng lớn nhất đến sự hài lòng của sinh viên đối với hoạt động của giảng viên được xác định như sau:

*Phương pháp giảng dạy, tiếp cận vấn đề khoa học rõ ràng, mạch lạc, dễ hiểu*

Phương pháp giảng dạy rõ ràng và mạch lạc giúp sinh viên nắm bắt kiến thức một cách

hiệu quả. Khi giảng viên trình bày vấn đề khoa học theo cách dễ hiểu, sinh viên có thể theo kịp nội dung bài giảng và liên kết các khái niệm mới với kiến thức đã có. Điều này vừa nâng cao hiệu quả học tập vừa tăng sự hứng thú và động lực học tập của sinh viên.

*Khuyến khích người học học tập, sẵn sàng lắng nghe, trả lời câu hỏi và hướng dẫn người học tự học, tự nghiên cứu*

Sự tương tác tích cực giữa giảng viên và sinh viên đóng vai trò quan trọng trong việc thúc đẩy quá trình học tập. Khi giảng viên khuyến khích sinh viên đặt câu hỏi, tham gia thảo luận và hỗ trợ họ trong việc tự học, sinh viên sẽ cảm thấy được tôn trọng và quan tâm. Điều này kích thích tư duy độc lập, phát triển kỹ năng nghiên cứu và tạo môi trường học tập chủ động.

*Khả năng bao quát và kiểm soát học tập của người học trên lớp*

Giảng viên cần thể hiện khả năng theo dõi tiến trình học tập của sinh viên, xác định rõ mức độ tiếp thu và tốc độ học tập của lớp. Việc kịp thời nhận diện khó khăn, chậm trễ hay hiểu lầm của sinh viên và điều chỉnh phương pháp giảng dạy phù hợp sẽ giúp đảm bảo mọi sinh viên đều có cơ hội phát triển. Nhờ đó, chất lượng học tập được duy trì đồng đều, giảm bớt sự chênh lệch trong nhận thức và tăng cảm giác hài lòng của tất cả người học.

*Sử dụng phương pháp, phương tiện dạy học hợp lý, hiệu quả*

Việc áp dụng các phương pháp và phương tiện dạy học phù hợp giúp tối ưu hóa quá trình truyền đạt kiến thức. Sử dụng công nghệ giáo dục, ví dụ như trình chiếu, mô phỏng hoặc phần mềm hỗ trợ, có thể làm cho bài giảng trở nên sinh động và tương tác hơn. Phương pháp giảng dạy linh hoạt, kết hợp giữa lý thuyết và thực hành, giúp sinh viên hiểu sâu hơn và áp dụng kiến thức vào thực tế.

*Nghiêm túc, khách quan và công bằng trong việc đánh giá*

Sự công bằng và minh bạch trong đánh giá tạo nên môi trường học tập đáng tin cậy. Khi sinh viên nhận thấy kết quả của họ được đánh giá một cách nghiêm túc, không thiên vị và phản ánh đúng năng lực cá nhân, họ sẽ có thêm động lực để học tập một cách trung thực, nỗ lực hơn trong việc nâng cao kết quả và chấp nhận phản hồi như một cơ hội để hoàn thiện.

Kết quả về các nhân tố ảnh hưởng quan trọng nhất đến sự hài lòng của sinh viên như trên cho thấy có sự tương đồng đáng kể với những nghiên cứu đi trước của (Merola et al., 2022; Sohail & Hasan, 2021). Tuy nhiên, với mẫu điều tra lớn và sử dụng kết hợp nhiều mô hình học máy như nghiên cứu trên đây đảm bảo tính tổng quát nhất định và kì vọng có thể áp dụng kết quả vào thực tế tại các cơ sở giáo dục đại học khác.

Nghiên cứu không chỉ khẳng định vai trò của học máy trong việc phân tích dữ liệu phản hồi từ sinh viên mà còn mở ra những định hướng quan trọng cho việc nâng cao chất lượng giảng dạy tại Trường đại học Thương mại. Từ những yếu tố then chốt ảnh hưởng đến sự hài lòng của sinh viên, nghiên cứu đề xuất một số giải pháp mang tính chiến lược và cụ thể:

### *Đối với cơ sở đào tạo*

Trường Đại học Thương mại nên đầu tư vào việc phát triển một hệ thống thu thập và phân tích phản hồi của sinh viên một cách tự động, sử dụng các kỹ thuật học máy (như mô hình Hồi quy tuyến tính đã được chứng minh là hiệu quả trong nghiên cứu này). Hệ thống này không chỉ giúp tiết kiệm thời gian và nguồn lực so với các phương pháp truyền thống mà còn cho phép phân tích dữ liệu phản hồi một cách sâu sắc hơn, xác định các xu hướng và vấn đề tiềm ẩn, từ đó đưa ra các quyết định quản lý kịp thời và chính xác.

Dựa trên các yếu tố quan trọng đã được xác định, nhà trường nên tổ chức các khóa đào tạo, bồi dưỡng thường xuyên để nâng cao năng lực cho giảng viên. Các khóa đào tạo này nên tập trung vào kỹ năng thiết kế bài giảng khoa học, rõ ràng, mạch lạc; phương pháp giảng dạy tích cực, khuyến khích sự tham gia của sinh viên; sử dụng công nghệ thông tin và các phương tiện dạy học hiện đại trong giảng dạy; kỹ năng đánh giá kết quả học tập của sinh viên một cách công bằng, khách quan.

### *Đối với giảng viên*

Từ kết quả nghiên cứu, mỗi giảng viên có thể chủ động cải thiện chất lượng giảng dạy của mình thông qua các hành động như: Chuẩn bị bài giảng kỹ lưỡng, câu trúc nội dung rõ ràng, sử dụng ngôn ngữ dễ hiểu, kết hợp các ví dụ minh họa sinh động, gần gũi với thực tiễn. Tạo không khí học tập thoải mái, cởi mở, khuyến khích sinh viên trao đổi, thảo luận, đặt câu hỏi, sẵn sàng lắng nghe và giải

đáp thắc mắc. Tăng cường sử dụng các phương tiện, công nghệ dạy học đa dạng, phù hợp với nội dung bài giảng và đặc điểm của sinh viên. Công bố rõ ràng các tiêu chí đánh giá ngay từ đầu, đảm bảo quá trình đánh giá diễn ra nghiêm túc, khách quan và công bằng với tất cả sinh viên.

### **4. Kết luận**

Nghiên cứu đã chứng minh hiệu quả vượt trội của việc áp dụng machine learning trong dự đoán mức độ hài lòng của sinh viên và xác định các yếu tố quan trọng ảnh hưởng đến sự hài lòng. Hồi quy tuyến tính (LR) cho thấy hiệu suất vượt trội, khẳng định tính ưu việt của phương pháp này so với các mô hình khác. Kết quả không chỉ đóng góp quan trọng về mặt học thuật mà còn mang lại ý nghĩa thực tiễn, hỗ trợ các nhà giáo dục trong việc nâng cao chất lượng giảng dạy và trải nghiệm học tập của sinh viên.

Kết quả nghiên cứu cung cấp cơ sở dữ liệu quan trọng cho giảng viên và nhà quản lý trong việc cải thiện phương pháp giảng dạy. Bằng cách tập trung vào các yếu tố ảnh hưởng lớn nhất đến sự hài lòng của sinh viên, các cơ sở giáo dục có thể điều chỉnh chiến lược giảng dạy và nâng cao chất lượng đào tạo. Việc đầu tư vào nâng cao kiến thức chuyên môn của giảng viên, áp dụng phương pháp giảng dạy hiệu quả, và tạo môi trường học tập tích cực sẽ góp phần tăng cường sự hài lòng và hiệu quả học tập của sinh viên.

Mặc dù nghiên cứu đã đạt được những kết quả khả quan trong việc ứng dụng mô hình học máy để dự đoán và phân tích sự hài lòng của sinh viên, cần thừa nhận rằng những phát hiện này chủ yếu có giá trị trong bối cảnh cụ thể của Trường Đại học Thương mại. Các đặc điểm riêng của Trường Đại học Thương mại về chương trình đào tạo, đội ngũ giảng viên và đối tượng sinh viên có thể ảnh hưởng đến kết quả, do đó, việc suy rộng cho các trường đại học khác cần được tiến hành một cách thận trọng. Nghiên cứu này cũng có một số hạn chế cần được nhìn nhận: Bộ dữ liệu sử dụng giới hạn trong một nhóm giảng viên và học phân cụ thể, có thể không phản ánh đầy đủ toàn bộ bối cảnh giáo dục; Ngoài ra, nghiên cứu chỉ sử dụng dữ liệu định lượng từ khảo sát, chưa kết hợp với các yếu tố khác như phong cách giảng dạy hoặc môi trường học tập.

Trong tương lai, nghiên cứu có thể mở rộng bằng cách thu thập dữ liệu từ nhiều giảng viên, học kỳ hoặc trường đại học khác nhau để tăng tính tổng quát. Việc kết hợp dữ liệu định tính và sử dụng các kỹ thuật xử lý dữ liệu phi tuyến như Deep Learning cũng có thể cải thiện hiệu suất mô hình và cung cấp những hiểu biết sâu sắc hơn. ♦

**Tài liệu tham khảo:**

Acampora, G., Herrera, F., Tortora, G., & Vitiello, A. (2018). A multi-objective evolutionary approach to training set selection for support vector machine. *Knowledge-Based Systems*, 147, 94-108. <https://doi.org/10.1016/j.knsys.2018.02.022>.

Agustin, F., Suherdi, D., & Purnawarman, P. (2024). Effective English teachers in online EFL learning: A post-pandemic reflection on students' perspectives. *Studies in English Language and Education*, 11(1), 188-204. <https://doi.org/10.24815/siele.v11i1.30124>.

Ahuja, R., Chug, A., Gupta, S., Ahuja, P., & Kohli, S. (2020). Classification and Clustering Algorithms of Machine Learning with their Applications. In X.-S. Yang & X.-S. He (Eds.), *Nature-Inspired Computation in Data Mining and Machine Learning* (Vol. 855, pp. 225-248). Springer International Publishing. [https://doi.org/10.1007/978-3-030-28553-1\\_11](https://doi.org/10.1007/978-3-030-28553-1_11).

Akritas, M. G., & Bershady, M. A. (1996). Linear Regression for Astronomical Data with Measurement Errors and Intrinsic Scatter. *The Astrophysical Journal*, 470, 706. <https://doi.org/10.1086/177901>.

Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3. <https://doi.org/10.1186/s41239-020-0177-7>.

Caldwell, D., Johnson, C., Moore, M., Moore, A., Poush, M., & Franks, A. M. (2024). Teaching Through the Student Lens: Qualitative Exploration of Student Evaluations of Teaching. *American Journal of Pharmaceutical Education*, 88(3), 100672. <https://doi.org/10.1016/j.ajpe.2024.100672>.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020).

A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215. <https://doi.org/10.1016/j.neucom.2019.10.118>.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>.

Christidis, A.-A., Lakshmanan, L., Smucler, E., & Zamar, R. (2020). Split Regularized Regression. *Technometrics*, 62(3), 330-338. <https://doi.org/10.1080/00401706.2019.1635533>.

Clements, J. N., McClendon, K., Borchert, J., Ciapala, S. R. L., Courtney, L. A., Cummings, J., Fallon, J. M., Fritz, M. K., Gibson, C. M., Isaacs, A. N., Israel, E. N., Johnson, P., Krauter, A., & Stolte, A. R. (2025). Teaching experiences in pharmacy residency. *JACCP: Journal of The American College of Clinical Pharmacy*, 8(1), 62-69. <https://doi.org/10.1002/jac5.2066>.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.

Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, 27(3), 3743-3775. <https://doi.org/10.1007/s10639-021-10741-7>.

Giang, T., & Nguyen, N. (2019). *Đạy học kết hợp: Một hình thức phù hợp với dạy học đại học ở Việt Nam thời đại kỹ nguyên số (B-learning: A Suitable Learning Modality for Higher Education in Vietnam at Digital Age)*. *Tạp Chí Khoa học trường Đại học Sư phạm Hà Nội, 1*, 2019.

Hoyle, B., Rau, M. M., Zitlau, R., Seitz, S., & Weller, J. (2015). Feature importance for machine learning redshifts applied to SDSS galaxies. *Monthly Notices of the Royal Astronomical Society*, 449(2), 1275-1283. <https://doi.org/10.1093/mnras/stv373>.

Jaiswal, G., Sharma, A., & Sarup, R. (2020). Machine Learning in Higher

Education: Predicting Student Attrition Status Using Educational Data Mining. In A. Solanki, S. Kumar, & A. Nayyar (Eds.), *Advances in Computational Intelligence and Robotics* (pp. 27–46). IGI Global. <https://doi.org/10.4018/978-1-5225-9643-1.ch002>.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>.

Jensen, J. B., & Artz, N. (2005). Using Quality Management Tools to Enhance Feedback from Student Evaluations. *Decision Sciences Journal of Innovative Education*, 3(1), 47–72. <https://doi.org/10.1111/j.1540-4609.2005.00052.x>.

Kalim, U., Tran, P. Q. B., Bibi, S., & Khamphouvong, L. (2022). Impact of Academic and Non-Academic Service Quality on International Students Satisfaction in Chinese Universities. *Journal of Education and Learning*, 11(6), 86. <https://doi.org/10.5539/jel.v11n6p86>.

Kilic, S. (2016). Cronbach's alpha reliability coefficient. *Journal of Mood Disorders*, 6(1), 47. <https://doi.org/10.5455/jmood.20160307122823>

Knapp, T. R., & Swoyer, V. H. (1967). Some Empirical Results Concerning the Power of Bartlett's Test of the Significance of a Correlation Matrix. *American Educational Research Journal*, 4(1), 13–17. <https://doi.org/10.3102/00028312004001013>.

Kumar, D., Kothiyal, A., Kumar, R., Hemantha, C., & Maranan, R. (2024). Random Forest approach optimized by the Grid Search process for predicting the dropout students. *2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET)*, 1–6. <https://doi.org/10.1109/ICICET59348.2024.10616372>.

Li, R. (2021). The Application of JASP Data Visualization in Blended Foreign Language Teaching. *Advances in Educational Technology and Psychology*, 5(4), 37–45. <https://doi.org/10.23977/aetp.2021.54006>.

Lichti, D. D., Chan, T. O., & Belton, D. (2021). Linear regression with an observation distribution model. *Journal of Geodesy*, 95(2), 23. <https://doi.org/10.1007/s00190-021-01484-x>.

Luis Miguel Olórtégui Alcalde. (2024). The University Student is Seen as a Customer: A Qualitative Analysis to Enhance Academic Performance from Satisfaction. *Evolutionary Studies in Imaginative Culture*, 1553–1566. <https://doi.org/10.70082/esiculture.vi.1551>.

Macfadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (2016). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment & Evaluation in Higher Education*, 41(6), 821–839. <https://doi.org/10.1080/02602938.2015.1044421>.

Maheshwari, A., Malhotra, A., Hada, B. S., Ranka, M., & Basha, M. S. A. (2024). Comparative Analysis of Machine Learning Models in Predicting Academic Outcomes: Insights and Implications for Educational Data Analytics. *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSSES)*, 1–7. <https://doi.org/10.1109/ICSSSES62373.2024.10561260>.

Manuel Vargas, V., Gutiérrez, P. A., & Hervás-Martínez, C. (2022). Unimodal regularisation based on beta distribution for deep ordinal regression. *Pattern Recognition*, 122, 108310. <https://doi.org/10.1016/j.patcog.2021.108310>.

Martínez-Roget, F., Freire Esparis, P., & Vázquez-Rozas, E. (2020). University Student Satisfaction and Skill Acquisition: Evidence from the Undergraduate Dissertation. *Education Sciences*, 10(2), 29. <https://doi.org/10.3390/educsci1002029>.

Masserini, L., Bini, M., & Pratesi, M. (2019). Do Quality of Services and Institutional Image Impact Students' Satisfaction and Loyalty in Higher Education? *Social Indicators Research*, 146(1), 91–115. <https://doi.org/10.1007/s11205-018-1927-y>.

Merola, R. H., Coelen, R. J., & Hofman, W. H. A. (2022). What really matters Factors impacting international student satisfaction: The case of the UK. *International Journal of Management in Education*, 16(1), 83. <https://doi.org/10.1504/IJMIE.2022.119684>.

Miller, L. D., Soh, L.-K., Samal, A., Kupzyk, K., & Nugent, G. (2015). A Comparison of Educational Statistics and Data



Mining Approaches to Identify Characteristics that Impact Online Learning. *Journal of Educational Data Mining*, 7(3), 117–150. <https://doi.org/10.5281/zenodo.3554732>.

Muijtjens, A. M. M., Timmermans, I., Donkers, J., Peperkamp, R., Medema, H., Cohen-Schotanus, J., Thoben, A., Wenink, A. C. G., & Van Der Vleuten, C. P. M. (2010). Flexible electronic feedback using the virtues of progress testing. *Medical Teacher*, 32(6), 491–495. <https://doi.org/10.3109/0142159X.2010.486058>.

Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. *Higher Education*, 80(5), 875–894. <https://doi.org/10.1007/s10734-020-00520-7>.

Mutawa, A. M., & Sruthi, S. (2023). Enhancing Human–Computer Interaction in Online Education: A Machine Learning Approach to Predicting Student Emotion and Satisfaction. *International Journal of Human–Computer Interaction*, 1–17. <https://doi.org/10.1080/10447318.2023.2291611>.

Nkansah, B. K. (2018). On the Kaiser-Meier-Olkin's Measure of Sampling Adequacy. *Mathematical Theory and Modeling*. <https://www.semanticscholar.org/paper/On-the-Kaiser-Meier-Olkin%E2%80%99s-Measure-of-Sampling-Nkansah/e8b787a9bb9266c1f941de1772c4d585cea4fc21>.

Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Achem, J. A. G., Escamilla, J., & Hosseini, S. (2022). Towards teaching analytics: A contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. *Education and Information Technologies*, 27(3), 3891–3933. <https://doi.org/10.1007/s10639-021-10751-5>.

Pandita, A., & Kiran, R. (2023). The Technology Interface and Student Engagement Are Significant Stimuli in Sustainable Student Satisfaction. *Sustainability*, 15(10), 7923. <https://doi.org/10.3390/su15107923>.

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference on Machine Learning, Big Data,*

*Cloud and Parallel Computing (COMITCon)*, 35–39. <https://doi.org/10.1109/COMITCon.2019.8862451>.

Reddy, M. R., Kumar, B. N., Rao, N. M., & Karthikeyan, B. (2020). A New Approach for Bias–Variance Analysis Using Regularized Linear Regression. In L. C. Jain, M. Virvou, V. Piuri, & V. E. Balas (Eds.), *Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals* (Vol. 1064, pp. 35–46). Springer Singapore. [https://doi.org/10.1007/978-981-15-0339-9\\_4](https://doi.org/10.1007/978-981-15-0339-9_4).

Rusticus, S. A., Pashootan, T., & Mah, A. (2023). What are the key elements of a positive learning environment? Perspectives from students and faculty. *Learning Environments Research*, 26(1), 161–175. <https://doi.org/10.1007/s10984-022-09410-4>.

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>.

Simsekler, M. C. E., Alhashmi, N. H., Azar, E., King, N., Luqman, R. A. M. A., & Al Mulla, A. (2021). Exploring drivers of patient satisfaction using a random forest algorithm. *BMC Medical Informatics and Decision Making*, 21(1), 157. <https://doi.org/10.1186/s12911-021-01519-5>.

Siscan, E., & Moldovan-Batrinac, V. (2024). Student Satisfaction Impact on the Economic Higher Education Institution Competitiveness. *Economica*, 1(127), 7–16. <https://doi.org/10.53486/econ.2024.127.007>.

Sohail, M. S., & Hasan, M. (2021). Students' perceptions of service quality in Saudi universities: The SERVPERF model. *Learning and Teaching in Higher Education: Gulf Perspectives*, 17(1), 54–66. <https://doi.org/10.1108/LTHE-08-2020-0016>.

Song, Y.-Y., & Ying, L. U. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130.

Sulis, I., Porcu, M., & Capursi, V. (2019). On the Use of Student Evaluation of Teaching: A Longitudinal Analysis Combining Measurement Issues and Implications of the Exercise. *Social Indicators Research*, 142(3), 1305–1331. <https://doi.org/10.1007/s11205-018-1946-8>.

Tanasa, D., & Trousse, B. (2004). Advanced data preprocessing for intersites Web usage mining. *IEEE Intelligent Systems*, 19(2), 59–65. IEEE Intelligent Systems. <https://doi.org/10.1109/MIS.2004.1274912>.

Urbano-Cano, A. L., López-Mesa, D. J., Alvarez-Rosero, R. E., & Garces-Gomez, Y. A. (2024). Predictive model for acute myocardial infarction in working-age population: A machine learning approach. *International Journal of Electrical and Computer Engineering (IJECE)*, 14(1), 854. <https://doi.org/10.11591/ijece.v14i1.pp854-860>.

Vasavada, B. (2023). *Use of supervised machine learning algorithms in predicting postoperative mortality in gastrointestinal and HPB surgeries*. Surgery. <https://doi.org/10.1101/2023.07.05.23292033>

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82. <https://doi.org/10.3354/cr030079>.

Wong, W. H., & Chapman, E. (2023). Student satisfaction and interaction in higher education. *Higher Education*, 85(5), 957–978. <https://doi.org/10.1007/s10734-022-00874-0>.

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>.

Zhang, P., & Zhang, F. (2024). Enhancing Education and Teaching Management Through Data Mining and Support Vector Machine Algorithm. *International Journal of E-Collaboration (IJeC)*, 20(1), 1–16. <https://doi.org/10.4018/IJeC.357998>.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J.-J., & Zhang, H. (2021). Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environmental Science & Technology*, 55(19), 12741–12754. <https://doi.org/10.1021/acs.est.1c01339>

Zia, T., Khawaja, S., & Qureshi, F. (2022). Student Satisfaction with Admission and Induction in the UK Private Higher Education Sector. *International Journal of Private Higher Education*, 1(1), 47–79. <https://doi.org/10.60166/VNXX6686>.

Ziegler, A., & König, I. R. (2014). Mining data with random forests: Current options for real-world applications. *WIREs Data Mining and Knowledge Discovery*, 4(1), 55–63. <https://doi.org/10.1002/widm.1114>.

## Summary

Student satisfaction is a critical metric for evaluating teaching quality in higher education. Previous studies have also confirmed that student satisfaction directly reflects teaching effectiveness and influences learners' motivation as well as academic. Thus, listening to and analyzing student feedback not only enables university instructors to improve their teaching methods but also contributes to enhancing the overall learning experience and educational quality. This study employs supervised machine learning models to predict student satisfaction with teaching activities at Thuongmai University and to identify key factors influencing their satisfaction. The research utilized seven different machine learning models on data consisting of responses from 3,477 students regarding faculty teaching activities during the 2023–2024 academic year. The results revealed that the Linear Regression model outperformed the other models. Additionally, five key teaching activities influencing student satisfaction were identified: clear teaching methods, the ability to oversee and manage student learning, the effective use of teaching methods and tools, encouraging self-study, and fairness and objectivity in assessment. The study provides critical insights for instructors and administrators to improve teaching methods, focusing on factors that significantly impact student satisfaction. Furthermore, it offers valuable information for researchers in selecting appropriate machine learning models for similar investigations.

## 1. Giới thiệu

Bài báo là sản phẩm của đề tài NCKH cấp Bộ B2025-TMA-06