

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO (AI) TRONG VIỆC PHÁT HIỆN TRÌNH TỰ HỌC TẬP TIẾNG NHẬT HIỆU QUẢ ĐỐI VỚI HỌC SINH TRUNG HỌC CƠ SỞ

APPLICATION OF ARTIFICIAL INTELLIGENCE (AI) IN IDENTIFYING EFFECTIVE JAPANESE LANGUAGE LEARNING SEQUENCES FOR JUNIOR HIGH SCHOOL STUDENTS

Nguyễn Thị Ngọc Liên^{1*}, Phạm Nguyễn Thanh Tâm²

¹Trường Đại học Ngoại ngữ - Đại học Đà Nẵng, Việt Nam

²Trung học cơ sở Tây Sơn - Đà Nẵng, Việt Nam

*Tác giả liên hệ / Corresponding author: ntnlien@ufl.udn.vn

(Nhận bài / Received: 05/8/2025; Sửa bài / Revised: 16/12/2025; Chấp nhận đăng / Accepted: 10/7/2026)

DOI: 10.31130/ud-jst.2026.24(7A).388

Tóm tắt - Trong nghiên cứu này Trí tuệ nhân tạo (AI) được sử dụng như một công cụ hỗ trợ nhằm xác định chuỗi học tập tối ưu cho học sinh THCS chuyên tiếng Nhật giúp nâng cao hiệu quả tiếp thu và phương pháp học phù hợp. Dữ liệu đầu vào được thu thập gồm mức độ hứng thú, trình tự học tập (TTHT) hiện tại, và thành tích của một số môn khác như toán, văn. Từ kết quả phân tích dữ liệu, thể nhận thấy rằng môn toán có mối liên hệ với kết quả học tập tiếng Nhật. Mối liên hệ này nổi trội hơn so với các môn học còn lại. Giả thuyết đặt ra là kỹ năng lập luận logic, hình dung không gian và mô hình hóa được trau dồi qua môn Toán, giúp học sinh nhanh hiểu bài và ghi nhớ chữ Hán tốt hơn. Khi triển khai TTHT được đề xuất dựa trên mô hình, điểm tăng nhẹ từ 8,70 lên 8,84. Mặc dù sự cải thiện này ở mức khiêm tốn nhưng phát hiện này cho thấy phương pháp tiếp cận mới có thể giúp cải thiện hiệu quả học tập.

Từ khóa - trí tuệ nhân tạo; trình tự học tập tiếng Nhật; dự đoán kết quả học tập; trung học cơ sở

1. Đặt vấn đề

Để hiểu và giao tiếp trôi chảy một ngôn ngữ khác ngoài tiếng mẹ đẻ đã trở thành nền tảng quan trọng giúp người học có nhiều cơ hội du học hay đi làm. Bên cạnh những điểm giống nhau về văn hóa, xã hội, mối liên kết của hai nước, thì tiếng Nhật đang dần khẳng định là một ngôn ngữ có sức hút hấp dẫn đối với người học Việt Nam chúng ta. Đặc biệt, vào tháng 4 năm 2025, sau những lần gặp gỡ và các phiên làm việc giữa những người đứng đầu của hai quốc gia, việc đưa tiếng Nhật vào dạy ở các cấp đã trở thành một nhiệm vụ giáo dục trọng tâm.

Sự kiện này đem lại lợi ích cho cả hai nước Việt Nam và Nhật Bản trong việc gắn kết mối quan hệ và phát triển nguồn nhân lực. Tuy nhiên việc triển khai các lớp tiếng Nhật, sẽ gặp nhiều khó khăn đối với ngành giáo dục Việt Nam, cụ thể nguồn lực giáo viên, thiết kế chương trình đào tạo tiếng Nhật phù hợp và lựa chọn phương pháp giảng dạy hiệu quả giúp học sinh làm quen với ngôn ngữ Nhật một cách có hệ thống và đạt chất lượng.

Hiện nay so với tiếng Nga, tiếng Anh, tiếng Pháp thì bản đồ giảng dạy tiếng Nhật đã có bước chuyển mình mạnh

Abstract - To support junior high school students specializing in Japanese in learning more effectively and appropriately, this study employs Artificial Intelligence (AI) to identify the optimal learning sequence. The dataset includes students' interest levels, current learning sequences, and academic performance in other subjects. A Random Forest machine learning algorithm is used to develop a predictive model of Japanese language achievement. The analysis reveals that Mathematics has the strongest influence on Japanese learning performance. This may be due to the logical reasoning, spatial visualization, and modeling skills cultivated through Mathematics, which help students learn and remember kanji more effectively. Implementation of the proposed learning sequence increases the predicted average score from 8.70 to 8.84. Although this improvement is modest, the findings demonstrate the feasibility of the proposed approach and suggest that it may help improve Japanese learning outcomes.

Key words - Artificial Intelligence; Japanese learning sequence; learning outcome prediction; Junior high school

mẽ về quy mô, phạm vi lan tỏa, từ các giảng đường đại học đến mọi cấp học phổ thông. Theo dữ liệu cập nhật [1], năm 2021, chỉ riêng thành phố Đà Nẵng có 15 trường THCS và PTTH đã đưa ngôn ngữ này vào chương trình chính thức giảng dạy và trong tương lai con số này có thể tăng lên. Điều đó chứng tỏ rằng giữa cha mẹ và học sinh đã có sự thay đổi trong cách nhìn nhận về vai trò và lợi ích của tiếng Nhật đối với định hướng học tập và nghề nghiệp. Tuy nhiên, khi thực tế triển khai chương trình giảng dạy ở một số trường nhiều người học gặp không ít khó khăn, chủ yếu do hạn chế về phương pháp tiếp cận và thiếu các lộ trình hay trình tự học tập (TTHT) chưa phù hợp đối với bản thân. Hiện nay có nhiều chiến lược nhằm phát huy hiệu quả của việc học ngoại ngữ. Trong đó tự học là một hình thức lựa chọn vì mang lại nhiều lợi ích cho học sinh như tiết kiệm tiền học phí, ít tốn quỹ thời gian đưa đón của phụ huynh và bản thân người học cũng chủ động tự sắp xếp được kế hoạch học tập.

Tuy nhiên, hiệu quả của hình thức này phụ thuộc rất lớn vào khả năng duy trì động lực, tính tự giác và kỹ năng điều chỉnh được quá trình học tập của bản thân. Đây là điều

¹ The University of Danang - University of Foreign Language Studies, Vietnam (Nguyen Thi Ngoc Lien)

² Tay Son Junior high school, Da Nang, Vietnam (Pham Nguyen Thanh Tam)

mà không phải tất cả các học sinh đều có thể làm được. Vì vậy hiện nay, nhiều người học vẫn còn gặp nhiều khó khăn từ đó dẫn đến kết quả học tập chưa đạt được như kỳ vọng. Ngược lại, việc học tập ở môi trường giảng dạy uy tín, chuyên nghiệp, chất lượng với đội ngũ thầy cô dày dặn kinh nghiệm lại giúp người học mau hiểu bài, xây dựng lộ trình học tập phù hợp với năng lực của từng người mặc dù chi phí đào tạo có thể cao hơn. Dẫu vậy, kết quả học tập cuối cùng của người học vẫn cần nhiều yếu tố phối hợp: mô hình triển khai phương pháp dạy học, khả năng tiếp thu kiến thức và TTHT phù hợp. Đáng chú ý, các chương trình đào tạo hiện nay đa phần vẫn vận hành theo một khung trình tự cứng nhắc, cụ thể các kỹ năng như: Từ vựng - ngữ pháp - đọc - nghe và nói. (TV, NP, Đọc, Nghe và Nói).

Khung trình tự này được các chuyên gia ngôn ngữ nghiên cứu và đưa vào giảng dạy ở các trường học nhằm đem lại khả năng tiếp thu nhanh cho số đông người học. Tuy nhiên, cách tiếp cận đồng nhất này vô hình chung khiến những người có phong cách học hoặc TTHT khác biệt khó tiếp cận và đạt hiệu quả tối đa. Nhận thấy được vấn đề này, nhóm nghiên cứu [2] đã làm rõ và xác định các nhóm người học có TTHT tương đồng trong chương trình học tiếng Nhật trực tuyến.

Dựa trên kết quả nghiên cứu của nhóm N.T.N. Liên, nghiên cứu này làm rõ phân tích thực tế TTHT của khối học sinh chuyên tiếng Nhật tại trường THCS Tây Sơn, Đà Nẵng. Điểm khác biệt của đề tài nằm ở việc tích hợp AI vào việc nhận diện các TTHT cá nhân hóa của học sinh. Thông qua phương pháp học máy (PPHM) để xây dựng mô hình dự đoán điểm số, dựa trên sự kết hợp giữa TTHT và các thông tin cá nhân khác. Mô hình đạt chỉ số tốt nhất sẽ được sử dụng để đề xuất TTHT hiệu quả cho mỗi học sinh. Kết quả thực nghiệm kỳ vọng sẽ chứng minh rằng việc áp dụng TTHT cá nhân hóa sẽ giúp cải thiện rõ rệt so với kết quả học tập hiện tại.

2. Cơ sở lý thuyết

2.1. Sử dụng AI trong giảng dạy ngoại ngữ

Trong các nghiên cứu quốc tế, việc dùng AI để tổ chức dạy học ngoại ngữ đang rất phổ biến. Tiêu biểu là nghiên cứu [3] đã lấy ý kiến về việc sử dụng chatbot nhằm cải thiện môn nói tiếng Anh ở cấp THCS. Kết quả cho thấy, khi sử dụng chatbot học sinh không chỉ nói tốt hơn mà còn thấy chủ động hơn và tự tin rất nhiều so với cách học cũ.

Về lĩnh vực viết thì nghiên cứu [4] khám phá cách học sinh chỉnh sửa văn bản được tạo bởi AI. Họ nhận diện hai hình thức chỉnh sửa chính: chỉnh sửa tập trung ở đầu đoạn văn trước khi chuyển sang các phần sau, hoặc nhanh chóng tập trung vào nội dung cốt lõi. Đáng chú ý, nghiên cứu cũng chỉ ra rằng việc giữ lại phần lớn nội dung do AI tạo ra có mối tương quan tích cực với điểm số bài viết, trong khi tần suất chỉnh sửa không có thay đổi đến chất lượng. Bên cạnh đó, nghiên cứu [5] đã chứng minh rằng việc khai thác dịch máy và áp dụng trí tuệ công cụ AI tạo sinh trong các hoạt động "nói" và "viết" rất hữu ích cho học sinh THCS trong việc luyện nói, luyện viết tăng sự tự tin khi bản thân điều chỉnh được bài viết.

Ngoài ra, nghiên cứu [6] ghi nhận sự quan tâm từ giáo viên và học sinh trong việc ứng dụng AI vào hoạt động dạy và học. Công nghệ này mang lại tiềm năng lớn để cá nhân hóa nội dung và phát triển khả năng tự học, tuy nhiên cũng có nhiều khó khăn về việc áp dụng kỹ năng công nghệ mới, độ uy tín của AI và vai trò của thầy cô giảng dạy.

Mặc dù, các công trình đi trước đã xác nhận tiềm năng của AI trong việc dạy và học ngoại ngữ, nhưng chưa có nghiên cứu nào đề cập đến ứng dụng AI để phát hiện TTHT tiếng Nhật hiệu quả dành cho đối tượng học sinh THCS.

2.2. Phương pháp phát hiện nhóm người học có trình tự học tập tương đồng

Theo nghiên cứu [2] đã tiếp cận phương pháp phát hiện nhóm người học có TTHT giống nhau bằng cách sử dụng kỹ thuật phân tích nhóm. Bài toán được đặt ra, nhằm xác định các nhóm học sinh có đặc điểm và TTHT thống nhất, dựa trên dữ liệu ghi nhận quá trình học:

"Nghiên cứu đã cho một nhóm gồm n người học là sinh viên các khóa đang theo học tiếng Nhật tại Đà Nẵng [7], trong đó yêu cầu mỗi cá nhân người học $i \in [1, n]$ phải thực hiện đầy đủ khối lượng bài học $l \in [1, ti]$, tại mỗi đơn vị bài học l , hệ thống phần mềm sẽ cung cấp các phương thức giảng dạy $mil \in [1, c]$, với c là tổng số phương pháp học để người học dễ tiếp cận. Một đặc điểm quan trọng của phương pháp này là cơ chế theo dõi, đánh giá chất lượng chỉ khi người học nghiêm túc học và thực hiện bài kiểm tra năng lực. Trong nhóm gồm n người học sẽ có trường hợp người học đạt được ngưỡng năng lực θ_l , thì mới được phép chuyển tiếp sang bài học tiếp theo $l+1$. Ngược lại, trường hợp kết quả chưa đạt chuẩn, người học được yêu cầu học lại và tham gia kiểm tra bổ sung cho đến khi đáp ứng thang điểm quy định. Ký hiệu s_{il} là số lần học lại của người i tại bài học l để đạt mức điểm yêu cầu. Nghiên cứu tập trung giải quyết bài toán: Tìm kiếm nhóm người học nào sẽ thích hợp với các trình tự học cụ thể?"

Để phát hiện sự tương đồng về phong cách học, nghiên cứu tiền đề đã áp dụng phương pháp Dendrogram. Đây là thuật toán phân cụm thứ bậc được sử dụng để xây dựng sơ đồ dạng cây, làm rõ trình tự sắp xếp các cụm dữ liệu theo từng tầng mức độ tương đồng. Thuật toán Dendrogram bao gồm các bước chính: khởi tạo, lặp gộp nhóm gần nhất dựa trên ma trận khoảng cách D và cập nhật khoảng cách giữa các nhóm. Ma trận khoảng cách D giữa các nhóm học tập được phân loại và xác lập dựa trên hệ số tương quan r_{ij} giữa trình tự học của hai người học i và j , với $r_{ij} \in [-1, 1]$.

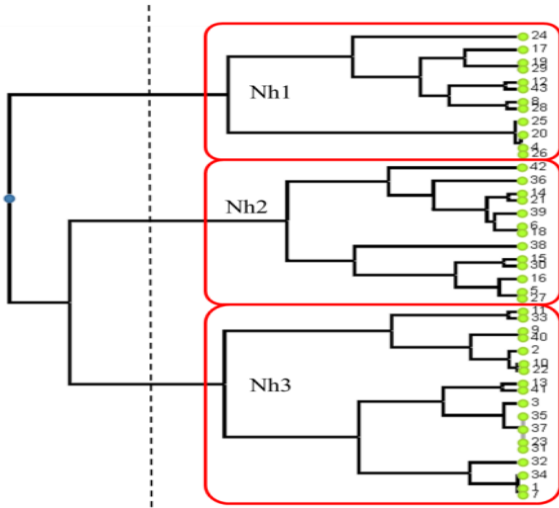
$$D = \begin{bmatrix} 1 & \cdots & 1 - r_{1n} \\ \vdots & \ddots & \vdots \\ 1 - r_{n1} & \cdots & 1 \end{bmatrix} \quad (1)$$

Trong đó, hệ số tương quan $r_{ij} = \frac{cov_{ij}}{\sigma_i \sigma_j}$ được xác định bằng tỷ lệ giữa hiệp phương sai cov_{ij} của TTHT của người học i và j , và tích độ lệch chuẩn $\sigma_i \sigma_j$ của chúng.

Nhóm nghiên cứu [2] đã triển khai trên nhóm đối tượng 43 sinh viên của Khoa Nhật – Hàn – Thái và thiết lập 4 TTHT chính (từ TTHT1 đến TTHT4) và xoay quanh các kỹ năng cốt lõi bao gồm:

- TTHT1: TV, NP, đọc, nghe và nói.
- TTHT2: NP, TV, nói, nghe và đọc.
- TTHT3: TV, nghe, NP, nói và đọc.
- TTHT4: NP, TV, nghe, đọc và nói.

Kết quả thực nghiệm cho thấy TTHT2 và TTHT4 là các trình tự được sinh viên ngành ngôn ngữ Nhật được áp dụng thường xuyên. Kết quả phân cụm ở Hình 1 xác định được ba nhóm người học tương đồng (Nh1, Nh2, Nh3).



Hình 1. Đồ thị Dendrogram trình bày kết quả phân tích nhóm người học [2]

Theo phân tích chuyên sâu chỉ ra: Nhóm Nh1 phù hợp với TTHT4; Nhóm Nh3 chỉ phù hợp với TTHT1; và Nhóm Nh2 có khả năng thích nghi tốt hơn, phù hợp với TTHT2. Từ đây, nhóm nghiên cứu [2] cho biết: Hiệu quả của các TTHT không đồng đều giữa các nhóm người học. Cụ thể, một số TTHT như TTHT2 và TTHT4 cho thấy kết quả nổi bật đối với những nhóm có số lượng người học quy mô lớn, tuy nhiên các TTHT này chưa chắc đem lại hiệu hứng tốt cho các nhóm còn lại. Đây là lý do để tác giả tập trung ứng dụng các PPHM nhằm xây dựng mô hình dự đoán kết quả học tập, từ đó đề xuất TTHT phù hợp cho học sinh chuyên tiếng Nhật tại Trường THCS Tây Sơn, Đà Nẵng.

2.3. Các phương pháp học máy

Hiện nay, có nhiều PPHM, để triển khai và xây dựng mô hình dự đoán kết quả học tập của học sinh chuyên tiếng Nhật tại Trường THCS Tây Sơn, Đà Nẵng, bài báo này chỉ sử dụng 3 PPHM chính gồm:

2.3.1. Hồi quy tuyến tính (Linear Regression)

Linear Regression [7, 8, 9] là thuật toán cơ bản trong học máy giám sát, được dùng để dự báo các giá trị liên tục. Đồng thời, quy tắc hoạt động này tập trung vào việc thiết lập mối tương quan hàm số giữa một hoặc nhiều biến độc lập (X) (đầu vào) và một biến phụ thuộc (y) (đầu ra). Thiết lập mối tương quan giữa tập hợp các biến và biến độc lập (X) được gọi là biến đầu vào và biến phụ thuộc được gọi là biến đầu ra (Y). Mục tiêu của thuật toán là tìm cách tối ưu hóa hàm tuyến tính để biểu diễn dữ liệu.

Công thức tổng quát của mô hình được biểu diễn như sau:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2)$$

Trong công thức này giá trị dự báo y được xác lập dựa trên hệ số chặn β_0 kết hợp với các hệ số hồi quy $\beta_1, \beta_2, \dots, \beta_n$ tương ứng với các biến đầu vào $x_1, x_2, \dots, x_n$. Bên cạnh đó, mô hình còn bao hàm thành phần ε nhằm đại diện cho các tham số nhiễu hoặc những sai số ngẫu nhiên không thể giải thích được bằng biến độc lập.

Về mặt ứng dụng, thuật toán này có nhiều điểm tốt như cấu trúc ngắn gọn, khi tính toán sẽ cho ra kết quả nhanh. Hồi quy tuyến tính thường dùng trong các lĩnh vực như đoán giá nhà, phân tích tình hình kinh tế thế giới hay đánh giá hiệu quả của việc chạy quảng cáo. Tuy nhiên, công cụ này vẫn có những điểm yếu nhất định, bao gồm sự thiếu ổn định trước các giá trị ngoại lai (outliers), đồng thời khi xử lý mối quan hệ với hồi quy tuyến tính phải cần có mối liên hệ và yêu cầu khắt khe về giả định phân phối chuẩn của các thành phần sai số.

2.3.2. Random Forest

Random Forest (RF) [10] là một mô hình học máy thuộc nhóm học tập tổ hợp (ensemble learning), sử dụng kỹ thuật bagging.

Để RF hoạt động dễ dàng thì phải dùng nhiều cây quyết định độc lập, trong đó các tập dữ liệu con được lấy dưới dạng không tuân thủ theo kế hoạch mà phải sử dụng mẫu ngẫu nhiên. Sau đó, những kết quả được cho là dự đoán cuối cùng được tổng hợp: đối với hồi quy, và kết quả là giá trị trung bình từ các cây. Điểm mạnh của RF nổi bật với khả năng giải quyết những dữ liệu được cho là phức tạp, đồng thời không yêu cầu chuẩn hóa dữ liệu và giảm thiểu nguy cơ quá khớp (overfitting) nhờ tính ngẫu nhiên. Mặc dù là một mô hình mạnh mẽ, RF có nhược điểm là tính phức tạp, khó giải thích (black-box), tiêu tốn tài nguyên và có thể giảm hiệu suất khi dữ liệu có nhiều cao.

2.3.3. Gradient Boosting

Gradient Boosting (GB) [11] cũng là một thuật toán học tập tổ hợp, nhưng sử dụng kỹ thuật boosting. Thay vì xây dựng các cây độc lập, GB xây dựng một chuỗi cây, và mỗi cây mới được thêm vào để khắc phục những lỗi sai (residuals) còn lại trước đây chưa được giải quyết. GB tối ưu hóa mô hình qua từng bước lặp bằng cách giảm thiểu hàm mất mát (Loss Function) thông qua phương pháp hạ gradient (gradient descent). GB có hiệu suất cao trong cả phân loại và hồi quy, và giải quyết tốt dữ liệu mất cân bằng. Tuy nhiên, GB dễ bị quá khớp nếu cấu trúc cây quá sâu hoặc số lượng cây quá lớn, đồng thời chậm hơn RF do tính tuần tự trong việc xây dựng mô hình và dễ bị tác động bởi các thành phần sai số.

3. Phân tích dữ liệu

Bài báo này sử dụng dữ liệu định lượng, thông qua việc sử dụng mô hình RF. Chức năng của mô hình RF là xác định TTHT tối ưu, dự đoán điểm trung bình và các yếu tố tác động đầu vào của học sinh chuyên Nhật tại Trường THCS Tây Sơn, Đà Nẵng.

3.1. Đối tượng khảo sát

Đối tượng tham gia khảo sát là những học sinh thuộc lớp chuyên Nhật của khối 7 (2 lớp) và khối 9 (2 lớp) của Trường THCS Tây Sơn, Đà Nẵng. Hoạt động khảo sát đã tiến hành thực hiện vào tháng 10 năm 2024.

Trường THCS Tây Sơn, Đà Nẵng là cơ sở giảng dạy tiếng Nhật từ sớm và hiện tại môn học này vẫn được duy trì đều đặn. Do các yếu tố khách quan nên học sinh khối 6 và khối 8 không tham gia khảo sát. Phiếu khảo sát tại Bảng 1 được thiết kế đơn giản, câu hỏi không quá khó và hoàn toàn thích hợp với đặc điểm lứa tuổi THCS để đảm bảo học sinh hiểu đúng nội dung và cung cấp thông tin đúng thực tế. Phiếu khảo sát được thiết kế như sau:

Tuy nhiên, sau khi loại bỏ các phiếu không đầy đủ thông tin hoặc không đủ điều kiện đưa vào nghiên cứu, bộ dữ liệu được đưa phân tích bao gồm 163 học sinh. Trong đó cấu trúc mẫu cụ thể như sau:

- Khối 7 (2 lớp): 99 học sinh (55 nữ, 45 nam), đã học tiếng Nhật khoảng 1,5 năm.

- Khối 9 (2 lớp): 64 học sinh (35 nữ, 29 nam), đã học tiếng Nhật khoảng 3,5 năm tại trường THCS Tây Sơn. Độ tuổi của học sinh tham gia khảo sát dao động từ 12 đến 14 tuổi.

Bảng 1. Phiếu khảo sát “Ứng dụng AI trong việc phát hiện TTHT tiếng Nhật hiệu quả”

PHIẾU KHẢO SÁT	
Ứng dụng AI trong việc phát hiện trình tự học tiếng Nhật hiệu quả đối với học sinh THCS	
1. Họ và tên: _____	
2. Lớp: _____	
3. Em có thích học tiếng Nhật không?	
☐ Có	☐ Không
4. Mục đích để em học tiếng Nhật là gì? (Chọn một hoặc nhiều đáp án)	
☐ Học theo yêu cầu của bố mẹ	
☐ Muốn giao tiếp bằng tiếng Nhật	
☐ Thỏa mãn niềm yêu thích ngôn ngữ Nhật, văn hóa Nhật Bản	
☐ Khác: _____	
5. Hiện nay, em thường xuyên học tiếng Nhật theo các trình tự nào dưới đây? (Chọn một hoặc nhiều đáp án)	
☐ Từ vựng → Ngữ pháp → Đọc → Nghe → Viết	
☐ Ngữ pháp → Từ vựng → Nói → Nghe	
☐ Từ vựng → Ngữ pháp → Nói → Viết	
☐ Đọc → Nghe → Viết	
☐ Các trình tự học tập khác: _____	
6. Em thường sử dụng phương pháp học tiếng Nhật nào ngoài việc giảng dạy chính khóa trên lớp? (Chọn tất cả các đáp án phù hợp)	
☐ Học thông qua giáo trình Minna no Nihongo và năng lực tiếng Nhật JLPT	
☐ Học thông qua các ứng dụng di động (Duolingo, LingoDeer,...)	
☐ Học ở các lớp học online do các giáo viên người Nhật, người Việt giảng dạy	
☐ Học thông qua YouTube, website,....	
☐ Khác: _____	
7. Em đã cảm thấy phương pháp học của bạn đã tối ưu hay chưa? Vì sao?	

8. Môn học khác mà em giỏi?	Và môn học em chưa giỏi?
(Trên 8.0 trung bình môn)	(Dưới 7.0 trung bình môn)
☐ Ngữ văn	☐ Ngữ văn
☐ Toán	☐ Toán
☐ Tiếng Anh	☐ Tiếng Anh
☐ Lịch sử	☐ Lịch sử
☐ Địa lý	☐ Địa lý
☐ Khoa học tự nhiên	☐ Khoa học tự nhiên
☐ Giáo dục công dân	☐ Giáo dục công dân
☐ Tin học	☐ Tin học
☐ Âm nhạc	☐ Âm nhạc
☐ Mỹ thuật	☐ Mỹ thuật
Cảm ơn em đã tham gia khảo sát!	

3.2. Xây dựng mô hình dự đoán điểm tiếng Nhật

Để có được mô hình dự đoán điểm học tiếng Nhật, tác giả đã tạo 8 câu hỏi, nhằm khai thác các thông tin về mức độ yêu thích, mục đích học tiếng Nhật, trình tự học và năng lực học tập ở các môn học khác của học sinh. Trước khi phát phiếu khảo sát ý nghĩa của các khái niệm cụm từ như TTHT (là cách học sinh sắp xếp kỹ năng theo trình tự mà bản thân thấy phù hợp và năng lực học tập (kết quả học tập ở các môn văn hóa như Toán, Ngữ văn, Tiếng Anh,...) đã được giải thích rõ ràng. Hệ thống ba mô hình học máy tại mục 2.3 khai thác các đặc điểm riêng từ dữ liệu khảo sát để xác lập mô hình dự đoán điểm số trung bình cho môn tiếng Nhật. Dữ liệu đầu vào bao gồm các đặc điểm: "Trình tự học tập" (ký hiệu TT), "Lớp", "Môn học mà bạn giỏi" (MHT) và "Môn học mà bạn chưa giỏi" (MHKT). Do các đặc trưng này mang tính chất phân loại

(categorical data) và không có mối quan hệ thứ bậc, tác giả đã áp dụng phương pháp mã hóa One-Hot Encoding [12] để chuyển đổi chúng thành dạng số (numerical data) mà mô hình học máy có thể xử lý. Đồng thời ở Bảng 2 đã đưa ra ví dụ minh họa về quá trình mã hóa dữ liệu của một học sinh.

Bảng 2. Ví dụ minh họa mã hóa dữ liệu học tập của một học sinh bằng phương pháp One-Hot

Lớp7	Lớp9	TT1	TT2	...	TT5	MHT1 (N.Văn)	MHT2 (Toán)	...	MHKT1 (N.Văn)	MHKT2 (Toán)	MHKT3 (T. Anh)	...
0	1	0	1	0	...	0	1	...	1	0	1	...

Sau khi chuẩn bị dữ liệu, tác giả bắt đầu xây dựng mô hình dự đoán điểm số tiếng Nhật, $\hat{y}_j$ dựa trên ba thuật toán Linear Regression, RF và GB đã trình bày ở Mục 2.3. Mô hình này được công thức hóa như sau:

$$\hat{y}_j = f_j(\mathbf{x}_i) \quad (3)$$

Trong đó, f_j là mô hình dự đoán điểm học tập của PPHM thứ j và $\mathbf{x}_i$ là tập dữ liệu đầu vào của học sinh i .

Hiệu quả của từng mô hình được xem xét thông qua ba tiêu chí đánh giá cụ thể sau đây được sử dụng:

3.2.1. Mean Absolute Error (MAE): Sai số tuyệt đối trung bình

Mean Absolute Error (MAE): Chỉ số của MAE được sử dụng để đo lường độ sai lệch trung bình (sai số tuyệt đối) giữa giá trị trong thực tiễn y_i và giá trị trong dự đoán $\hat{y}_j$

$$MAE_j = \frac{1}{n} \sum_{i=1}^n |y_i - f_j(\mathbf{x}_i)|, \quad (4)$$

Mô hình có MAE càng thấp càng chính xác. MAE là chỉ số dễ hiểu, phản ánh khoảng cách trung bình giữa giá trị trong thực tiễn và giá trị trong dự đoán.

3.2.2. Root Mean Squared Error (RMSE): Sai số bình phương trung bình căn bậc hai

RMSE được xác lập để đo lường căn bậc hai của trung bình bình phương các độ lệch giữa giá trị trong thực tiễn và giá trị trong dự đoán.

$$RMSE_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_j(\mathbf{x}_i))^2}, \quad (5)$$

Điểm đặc trưng của tham số này là khả năng xảy ra các lỗi lớn (outliers) so với MAE do cơ chế bình phương sai số. Điều này khiến RMSE dễ nhạy hơn so với các giá trị ngoại lai khác, giúp nghiên cứu này đánh giá sát sao độ ổn định của mô hình này.

3.2.3. Coefficient of Determination (R^2): Hệ số xác định

Hệ số xác định R^2 phản ánh năng lực của mô hình trong việc giải thích mức độ biến thiên của tập dữ liệu so với phương án dự báo trung bình $\bar{y}$. R^2 được xác định bởi công thức:

$$R_j^2 = 1 - \frac{S_{res}}{S_{tot}}, \quad (6)$$

Trong đó, $S_{res} = \sum_{i=1}^n (y_i - f_j(\mathbf{x}_i))^2$ là tổng bình phương độ sai khác (Residual Sum of Squares), và $S_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$ là tổng bình phương tổng thể (Total Sum of Squares). $\bar{y}$ là giá trị trung bình của y_i . Chỉ số $R_j^2 = 1$ đã cho thấy mô hình này giải thích hoàn toàn hợp lý. Tuy nhiên, điều ngược lại khi chỉ số $R_j^2 = 0$ hoặc chỉ số

$R^2_j < 0$ nghĩa là không đem lại hiệu quả hơn so với việc sử dụng giá trị trung bình đơn thuần.

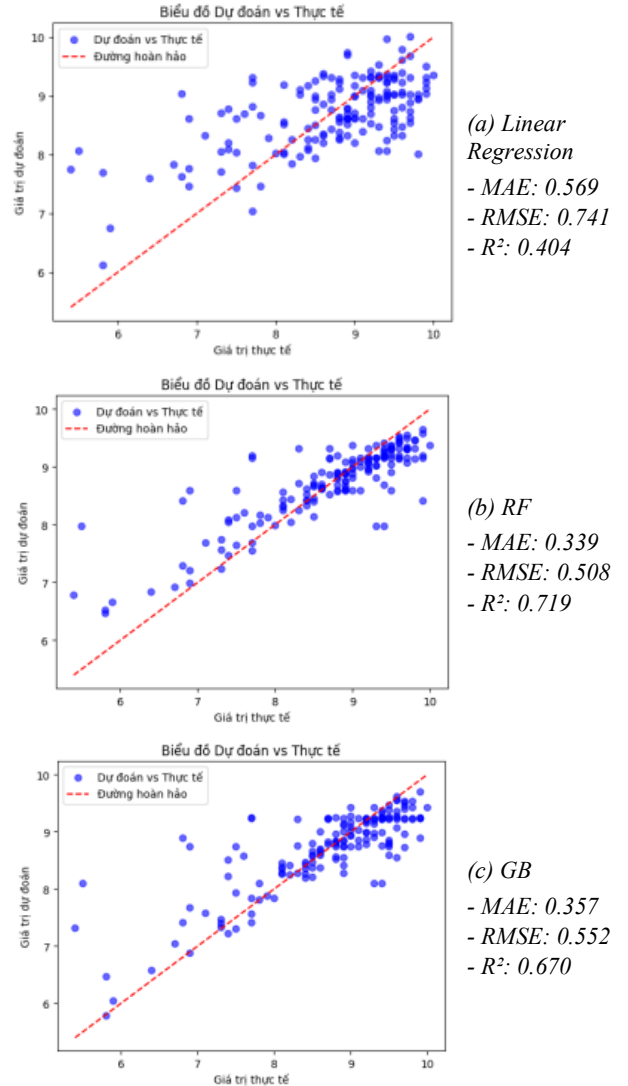
Dựa trên kết quả so sánh giữa ba mô hình gồm: Linear Regression, RF, GB ở Bảng 3 thì mô hình RF cho kết quả vượt trội nhất với $R^2 = 0,719$, MAE thấp nhất (0,339) và RMSE thấp nhất (0,508).

Bảng 3. Kết quả đánh giá thông qua ba mô hình

Tên mô hình	MAE	RMSE	R^2
Linear Regression	0,569	0,741	0,404
RF	0,339	0,508	0,719
GB	0,357	0,552	0,670

Để trực quan hóa độ tin cậy của từng mô hình, Hình 2 trình bày biểu đồ so sánh giữa điểm tiếng Nhật thực tế (y) và điểm dự đoán ($\hat{y}$) của ba mô hình.

Đường chéo màu cam thể hiện trường hợp dự đoán hoàn hảo ($y = \hat{y}$).



Hình 2. Biểu đồ so sánh điểm dự đoán ($\hat{y}$) và điểm thực tế (y) của ba mô hình học máy

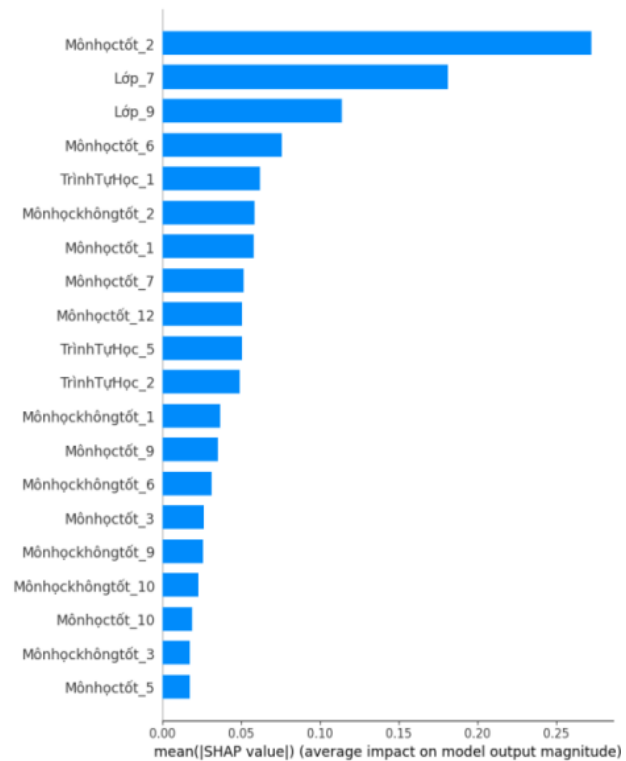
Xét tổng thể ở các hình, ta thấy các điểm dữ liệu được phân bố sát đường chéo thì mô hình được xem là có độ chính xác cao.

Trong các biểu đồ Hình 2, thì biểu đồ 2b cho thấy các điểm dự đoán tập trung gần đường chéo hơn so với Linear Regression và GB, minh chứng cho kết quả đánh giá vượt trội của mô hình này $R^2=0,719$. Do đó, nghiên cứu này tiếp tục được thực hiện với mô hình RF để xem xét mức độ tác động và sự ảnh hưởng của các yếu tố đầu vào.

3.3. Kết quả nghiên cứu

Kết quả học tập là một chỉ số tổng hợp nhằm phản ánh năng lực tiếp cận nội dung bài học, hoàn thiện các kỹ năng còn hạn chế và nhận thức học tập của người học. Để làm rõ hệ số ảnh hưởng của các biến số đầu vào đối với năng lực tiếp thu tiếng Nhật nghiên cứu đã sử dụng giá trị thuật toán (SHapley Additive exPlanations) hay còn gọi SHAP [13]. Đây là hướng giải pháp tối ưu giúp làm rõ cơ chế dự báo của mô hình học máy thông qua việc xác định giá trị đóng góp biên của từng thuộc tính đặc trưng. TTHT và mức độ chi phối trung bình của các yếu tố đã qua xử lý mã hóa One-Hot được chi tiết hóa tại Hình 3.

Từ biểu đồ Hình 3, một phát hiện quan trọng và khá thú vị là Môn học tốt_2 (môn Toán) có mức độ ảnh hưởng đến kết quả dự đoán điểm tiếng Nhật nhất, vượt trội hơn cả môn Tiếng Anh hay các môn học khác. Điều này tác giả đưa ra giả thuyết rằng học sinh giỏi Toán thường sở hữu các kỹ năng lập luận, phân biện, hình học không gian và năng lực mô hình hóa tốt. Những học sinh có đủ kiến thức, kỹ năng này thường có xu hướng tiếp cận và nắm bắt hiệu quả nhanh hơn về hệ hổng chữ viết phức tạp của tiếng Nhật, đặc biệt là chữ Hán (Kanji). Tuy nhiên, đây chỉ là giả thuyết ban đầu và cần được kiểm chứng bằng số liệu định lượng trong các nghiên cứu chuyên sâu tiếp theo.



Hình 3. Mức độ ảnh hưởng trung bình (SHAP Value) của các đặc trưng đến kết quả dự đoán điểm tiếng Nhật

3.4. Thực nghiệm đề xuất trình tự học phù hợp

Nghiên cứu tập trung xem xét khả năng đáp ứng yêu cầu học tập cũng như hiệu quả của các TTHT tập đôi với học sinh chuyên Nhật tại Trường THCS, Tây Sơn. Dựa trên mô hình RF tốt nhất, tác giả đề xuất học sinh chuyển đổi sang TTHT tối ưu hơn. Bảng 4 tóm tắt quy mô học sinh tham gia các TTHT cũ và các TTHT mới được đề xuất.

Bảng 4. Số lượng học sinh chuyên tiếng Nhật tham gia các TTHT (trước và sau đề xuất)

TTHT cũ \ TTHT mới	TTHT 1	TTHT 2	TTNT 3	TTHT 4	TTHT 5
TTHT1	43	8	1	11	20
TTHT2	1	2	0	0	0
TTHT3	0	0	2	0	2
TTHT4	29	4	5	14	21

Trong đó:

- TTHT1: TV→NP→Đọc→Nghe→Nói;
- TTHT2: NP→TV→Nói→Nghe→Đọc;
- TTHT3: TV→Nghe→NP→Nói→Đọc;
- TTHT4: NP→TV→Nghe→Đọc→Nói;
- TTHT5: Các trình tự khác không thuộc 4 trình tự trên.

Ghi chú: Bảng 4 học sinh tham gia các trình tự học tập. Hàng: TTHT mới (TTHT1–TTHT4); cột: TTHT cũ (TTHT1–TTHT5). Ô số: số lượng học sinh; số in đậm: học sinh không cần thay đổi TTHT

Căn cứ ở bảng thống kê, có tất cả 71 học sinh tương ứng với các ô trên đường chéo không cần thay đổi TTHT hiện tại. Cụ thể tương ứng với các ô nằm trên đường chéo: $43 + 2 + 2 + 14 = 71$ học sinh, trong khi đó, 92 học sinh còn lại đã chưa lựa chọn đúng TTHT phù hợp và cần đề xuất chuyển sang các TTHT khác phù hợp hơn. Sau khi đề xuất các TTHT phù hợp cho học sinh, tác giả tiếp tục áp dụng mô hình học máy RF vốn đã được áp dụng trước đó để dự đoán điểm trung bình học tập của học sinh. Dữ liệu cho thấy điểm trung bình đã **tăng từ 8,70 lên 8,84**. Để đánh giá mức độ cải thiện một cách định lượng, tác giả đã áp dụng chỉ số kiểm định t ghép đôi (paired t-test) [14] để so sánh điểm học tập của 163 học sinh trước và sau khi áp dụng hệ thống đề xuất AI như sau:

- Điểm trung bình và độ lệch chuẩn trước khi áp dụng: $M = 8.71$, $Var = 0.45$.

- Điểm trung bình và độ lệch chuẩn sau khi áp dụng: $M = 8.84$, $Var = 0.25$.

Hiệu trung bình giữa hai giai đoạn là -0.124 (95% CI $[-0.174, -0.075]$), với hệ số tương quan cặp rất cao ($r = 0.89$). Kiểm định t ghép đôi xác nhận điểm khác nhau này mang ý nghĩa thống kê ($t(162) = -4.96$, $p < 0.001$). Mặc dù kích thước hiệu ứng (Cohen's $d \approx 0.39$) ở mức nhỏ, vừa, phù hợp với mức tăng khoảng 1.4%, kết quả này chỉ ra rằng việc sử dụng mô hình AI để phát hiện TTHT đã giúp thay đổi, cải thiện điểm số học tập của học sinh một cách rõ rệt.

Mặc dù, mức tăng không quá lớn, song điều này vẫn phản ánh một tín hiệu tích cực và khẳng định rằng hiệu quả của việc đề xuất TTHT rất phù hợp. Việc lựa chọn đúng TTHT không chỉ đem lại kết quả tốt mà còn hỗ trợ học sinh duy trì hiệu quả học tập ổn định và lâu dài.

4. Kết luận và kiến nghị

4.1. Kết luận

Nghiên cứu đã ứng dụng thành công AI để xây dựng mô hình dự đoán kết quả học tập môn tiếng Nhật dựa trên TTHT của học sinh. Trên cơ sở mô hình dự báo đã được tối ưu hóa bởi mô hình RF, tác giả đã đề xuất điều chỉnh TTHT nhằm góp phần hỗ trợ người học thuận lợi trong việc tiếp thu kiến thức và cải thiện kết quả học tập.

Dữ liệu thực nghiệm thu được từ học sinh chuyên tiếng Nhật tại Trường THCS Tây Sơn, Đà Nẵng, đã chứng minh rằng việc áp dụng TTHT theo mô hình đề xuất có khả năng tăng điểm trung bình môn học tiếng Nhật của học sinh từ 8.70 lên 8.84. Kết quả này cho thấy việc ứng dụng AI để góp phần tối ưu hóa trong việc giảng dạy và học tập tiếng Nhật ở bậc trung học cơ sở. Hướng nghiên cứu tiếp theo, tác giả dự kiến mở rộng quy mô thu thập dữ liệu, tiếp tục cải tiến mô hình dự đoán và phát triển thêm các TTHT mới, phù hợp với đặc điểm của từng nhóm học sinh, nhằm mục tiêu tăng chất lượng dạy và học tiếng Nhật.

4.2. Kiến nghị

Từ kết quả nghiên cứu đã đạt được, tác giả xin đề xuất một số kiến nghị:

4.2.1. Đối với giáo viên

Bên cạnh việc áp dụng AI để phát hiện TTHT hiệu quả cho học sinh chuyên Nhật tại trường THCS Tây Sơn, giáo viên nên tích cực áp dụng các công cụ AI vào dạy học. Cụ thể, AI có thể giúp giáo viên trong việc phân loại bài tập, thiết kế bài giảng, theo dõi tiến trình học và có thể đưa ra dự đoán sớm kết quả học tập. Điều này cung cấp căn cứ khoa học giúp giáo viên phát hiện kịp thời những học sinh tiếp thu kém, từ đó kiến nghị các hình thức hỗ trợ phù hợp. Ngoài ra, giáo viên cần nghiên cứu và chủ động làm mới cách dạy để tương thích với năng lực và mức độ hiểu bài của từng nhóm học sinh. Việc này giúp học sinh vượt khó có hứng thú trong học tập, tích cực tham gia các hoạt động nhà trường đưa ra, đồng thời giảng viên là người hỗ trợ học sinh tiến bộ theo khả năng cá nhân. Việc phối hợp các hoạt động tương tác, trò chơi học tập, thảo luận nhóm, hoặc luyện tập ngữ pháp và phát âm tiếng Nhật, nên được thực hiện dựa trên các gợi ý từ mô hình AI. Sự kết hợp này không chỉ tạo sự hào hứng mà còn góp phần củng cố kiến thức một cách bền vững cho học sinh.

4.2.2. Đối với Nhà trường

Căn cứ vào định hướng của Bộ GD&ĐT chương trình môn Tiếng Nhật ở phổ thông [15], nhà trường cần xem xét áp dụng mô hình AI như một công cụ quản lý chất lượng dạy học tiếng Nhật.

Để việc triển khai ứng dụng AI trong giảng dạy một cách thuận lợi, nhà trường cần đầu tư vào cơ sở vật chất và hạ tầng công nghệ. Đây là nền tảng cốt yếu giúp mô hình AI phát huy tối đa hiệu quả. Về chiến lược dài hạn, nhà trường nên xem xét và từng bước triển khai ứng dụng AI sang các môn học và ngoại ngữ khác. Đây là hướng đi phù hợp với xu thế chuyển đổi số trong giáo dục, góp phần nâng cao chất lượng đào tạo toàn diện.

TÀI LIỆU THAM KHẢO

- [1] Embassy of Japan in Vietnam, "Implementation status of Japanese language teaching and learning in provinces/cities and high schools in Vietnam", *vn.emb-japan.go.jp*, March 4, 2021. [Online]. Available: https://www.vn.emb-japan.go.jp/itpr_ja/20210304/JapaneseLanguage_1_vn.html [Accessed July 03, 2025].
- [2] N. T. N. Lien and P. M. Tuan, "Detection model for similar learning style groups – its application to online Japanese study", *The University of Danang – Journal of Science and Technology*, vol. 7, no. 128, pp. 80–84, 2018. <https://jst-ud.vn/jst-ud/article/view/862>
- [3] A. Alenezi and A. Alenezi, "Evaluating the effectiveness of chatbot-assisted learning in enhancing English conversational skills among secondary school students", *Education Sciences*, vol. 15, no. 9, article 1136, 2025. <https://doi.org/10.3390/educsci15091136>
- [4] D. J. Woo, Y. Yu, and K. Guo, "Exploring EFL Secondary Students' AI-generated Text Editing While Composition Writing", *arXiv preprint*, 2025. <https://doi.org/10.48550/arXiv.2505.17041>
- [5] M. Shinoka, "Utilizing Machine Translation and Generative AI in Junior High School English Classes: Through Language Activities of 'Speaking' and 'Writing'", *Bulletin of the Tottori University Junior High School*, vol. 56, pp. 91–97, 2025
- [6] L. T. N. Ha and N. L. T. Tien, "The current situation and potential of AI applications in foreign language teaching and learning at some high schools in Da Nang City," *The University of Danang - Journal of Science and Technology*, vol. 23, no. 8C, pp. 55–59, 2025.
- [7] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd edition. New York: Springer, 2021.
- [8] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press, 2009.
- [9] G. A. F. Seber, A. J. Lee, *Linear Regression Analysis*. Hoboken: Wiley-Interscience, 2012.
- [10] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. New York: Springer, 2009.
- [11] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial", *Frontiers in Neuroinformatics*, vol. 7, article 21, 2013. <https://doi.org/10.3389/fnbot.2013.00021>
- [12] P. Rodríguez, M. A. Bautista, J. González, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding", *Image and Vision Computing*, vol. 75, pp. 21–31, 2018. <https://doi.org/10.1016/j.imavis.2018.04.004>
- [13] E. Strumbelj and I. Kononenko, "An efficient explanation of individual classifications using game theory", *Journal of Machine Learning Research*, vol. 11, pp. 1–18, 2010.
- [14] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 6th edition. London: SAGE Publications, 2023.
- [15] Ministry of Education and Training, *General Education Curriculum for Japanese Language*, No. 32/2018/TT-BGDĐT, 2018.