TÓM TẮT

Thời gian gần đây, hầu hết các trường đại học đều có kế hoạch mở hoặc mở các khóa học trực tuyến, tuy nhiên vấn đề bỏ học trực tuyến đang diễn ra vẫn là một vấn đề đối với các trường đại học. Học trực tuyến có ưu điểm là có thể tiếp thu kiến thức mọi lúc, mọi nơi nhưng thực tế rằng tỷ lệ bỏ học cao hơn so với các lớp học ngoại tuyến vì bạn phải tự quản lý và kiểm soát thời gian học của mình mà không có sự trợ giúp của giáo viên. giáo sư hoặc quản lý. Do đó, điều rất quan trọng là các giáo sư và nhà quản lý phải hỗ trợ sinh viên hành động kịp thời để tránh rủi ro. bỏ học các lớp học trực tuyến của trường đại học. Nghiên cứu này đã sử dụng dữ liệu nhật ký truy cập được ghi trong Hệ thống quản lý học tập (LMS) và thông tin thống kê và dữ liệu tính toán của người học, nhằm mục đích đưa ra các thuật toán dự đoán phù hợp cho việc bỏ học trực tuyến hệ thống dự đoán sớm tại các trường đại học. Nghiên cứu này có dữ liệu nhật ký lịch sử học tập trực tuyến trong 7 năm được ghi lại tại Đại học Điện tử Hệ thống LMS để khắc phục những hạn chế về số lượng dữ liệu của các nghiên cứu hiện tại và dự đoán nguy cơ bỏ học trong thời gian học tập. Các các đặc điểm của dữ liệu bạn sử dụng đã được sử dụng để xác thực tính khả dụng của các mô hình dự đoán bằng cách áp dụng thống kê người học thông tin, số lượng kết nối hệ thống, số lượng bài giảng, dữ liệu điểm của học kỳ trước, cây quyết định dựa trên học máy, rừng tùy ý (RF), máy vectơ hỗ trợ (SVM) và học sâu (DNN). Các nghiên cứu cho thấy thuật toán rừng ngẫu nhiên (RF) có dự đoán và hiệu suất tốt nhất, đồng thời các thuật toán học sâu cũng áp dụng cho các hệ thống quản lý học tập (LMS).

GIỚI THIỆU

Giáo dục trực tuyến là một cách tốt để giúp người học tiếp thu kiến thức mới kiến thức và nhận được một mức độ mà không có bất kỳ thời gian hoặc không gian hạn chế. Mặc dù giáo dục trực tuyến đang nhận được nhiều sự chú ý trong các trường đại học bởi vì hầu hết các trường đại học có gần đây đã mở hoặc có kế hoạch mở các khóa học giáo dục trực tuyến, có là một vấn đề mà tỷ lệ bỏ học cao hơn ngoại tuyến giáo dục vì quản lý và giám sát thời gian của người học các hoạt động giáo dục không phải là vốn có trong giáo dục trực tuyến. bên trong trường hợp người học tham gia giáo dục trực tuyến với mục tiêu chẳng hạn như phát triển bản thân và đạt được bằng cấp, giảm ra sẽ mất thời gian và sức mạnh kinh tế, mà cuối cùng dẫn đến ngừng phát triển bản thân và không đạt được một mức độ. Vì vậy, việc hạ thấp tỷ lệ học sinh bỏ học đòi hỏi phải tích cực sự can thiệp của giáo viên hoặc người quản lý nếu người học có khả năng bỏ học và phát triển các dịch vụ dự đoán bỏ học sớm là một thách thức quan trọng đối với các trường đại học. Vì vậy, nó là sự thật rằng không có nhiều nghiên cứu đã được tiến hành về tính khả thi của các mô hình dự đoán để dự đoán việc bỏ học sớm như vậy.. Nghiên cứu này nhằm phát triển một mô hình dự báo hiệu quả để áp dụng cho các hệ thống dự đoán bỏ học hiệu quả mà dự đoán khả năng bỏ học trực tuyến này trước và thông báo cho giáo viên hoặc cán bộ quản lý. Đã có nhiều nghiên cứu liên quan đến tình trạng bỏ học của học trực tuyến. Tuy nhiên, nghiên cứu này có những điểm sau sự khác biệt so với các nghiên cứu trước đây. Thứ nhất, nó có ý nghĩa trong việc nghiên cứu các mô hình phù hợp với ứng dụng học trực tuyến tại các trường đại học. Gần đây, phân tích dự đoán bỏ học bằng cách sử dụng dữ liệu lịch sử học tập trực tuyến đã được tiến hành tích cực, nhưng hầu hết trong số họ được dựa trên Các khóa học trực tuyến MOOCs, gây khó khăn cho việc đăng ký trực tuyến học tập tại các trường đại học. Thứ hai, nghiên cứu này được đặc trưng bởi nghiên cứu một mô hình có thể dự đoán những người bỏ học sớm trên cơ sở hàng tuần. tiên đoán nhất nghiên cứu về việc bỏ học trực tuyến có vấn đề không thể dự đoán bỏ học cho người học đang tiến hành khi họ sử dụng người học khảo sát sau khi học xong và dữ liệu được lưu trữ trong hệ thống học tập gần đây. Trong công việc này, chúng tôi có thể thông báo các trường đại học và giáo sư sớm bỏ học sớm giai đoạn học tập, dưới dạng phân tích dữ liệu hàng tuần thay vì dataJOIV : Int. J. Thông báo. Visualization, 5(4) - Tháng 12 năm 2021 347-353347 sau khi học xong cho phép bỏ học sớm đặc trưng. Thứ ba, nghiên cứu này đã khắc phục được khó khăn trong việc khái quát hóa do lượng dữ liệu nhỏ. Nó sử dụng lịch sử học tập thực tế dữ liệu được thu thập trong hệ thống quản lý học tập (LMS) của các trường đại học trực tuyến ở Hàn Quốc trong hơn một thập kỷ. Các nghiên cứu trước đây liên quan đến nghiên cứu này có thể được chia thành lĩnh vực phân tích nhân tố có tác động đáng kể đến học tập trực tuyến và nghiên cứu dự đoán để ngăn chặn sớm bỏ học như sau A. Nghiên cứu về tính năng Bỏ học của Học trực tuyến Theo các nghiên cứu trước đây, kể từ khi học trực tuyến diễn ra trong các tình huống không gian và không gian tách biệt, nó đã nhấn mạnh khả năng học tập tự định hướng, trong số những thứ khác, tầm quan trọng của dữ liệu phân tích học tập, mà được tính từ hồ sơ học tập và học tập của người học [1][2]. Có những nghiên cứu giải thích rằng dữ liệu hoạt động học tập (bảng tin miễn phí, bảng bài giảng, bảng tài liệu học tập, bảng thảo luận và tỷ lệ truy cập mô-đun nội dung) có tác động đáng kể đến hiệu suất học tập, và chi tiết đó các yếu tố (chuyên ngành, giới tính, tuổi, lớp, học vấn, nghề nghiệp) cũng có tác động lớn hơn đến hiệu suất học tập [3]. Có các nghiên cứu cho thấy sức mạnh dự đoán tăng nhiều nhất khả năng học tập và dự đoán khi sử dụng dữ liệu từ người học cá nhân, môi trường học tập và quá trình học tập một cách toàn diện

A. Nghiên cứu về tính năng Bỏ học của Học trực tuyến

Theo các nghiên cứu trước đây, kể từ khi học trực tuyến diễn ra trong các tình huống không gian và không gian tách biệt, nó đã nhấn mạnh khả năng học tập tự định hướng, trong số những thứ khác, tầm quan trọng của dữ liệu phân tích học tập, mà được tính từ hồ sơ học tập và học tập của người học [1][2]. Có những nghiên cứu giải thích rằng dữ liệu hoạt động học tập (bảng tin miễn phí, bảng bài giảng, bảng tài liệu học tập, bảng thảo luận và tỷ lệ truy cập mô-đun nội dung) có tác động đáng kể đến hiệu suất học tập, và chi tiết đó các yếu tố (chuyên ngành, giới tính, tuổi, lớp, học vấn, nghề nghiệp) cũng có tác động lớn hơn đến hiệu suất học tập [3]. Có các nghiên cứu cho thấy sức mạnh dự đoán tăng nhiều nhất khả năng học tập và dự đoán khi sử dụng dữ liệu từ người học cá nhân, môi trường học tập và quá trình học tập toàn diện [4].

B. Nghiên cứu về dự báo bỏ học trực tuyến

Gần đây, nghiên cứu về các mô hình dự đoán đã được tích cực thực hiện xung quanh các bài giảng MOOC để thông báo nhu cầu can thiệp đối với học sinh có nguy cơ bỏ học. Đây là một nghiên cứu phát triển dự đoán bỏ học tạm thời hàng tuần mô hình và đề xuất một mô hình cung cấp thông tin về can thiệp được cá nhân hóa bằng cách sử dụng xác suất [5]. [6] đã áp dụng các phương pháp học máy để dự đoán hiệu suất học tập của trực tuyến mở quy mô lớn khóa học (MOOCs) và cho thấy rằng có một mối tương quan chặt chẽ giữa hành vi nhấp chuột của người học và kết quả học tập. Chúng tôi cũng giải thích rằng trong số các phương pháp học máy khác nhau, rừng ngẫu nhiên cho thấy hiệu suất tốt nhất để dự đoán các vấn đề. Một nghiên cứu dự đoán bỏ học theo phiên của Các khóa học MOOC định nghĩa các đơn vị học tập và dự đoán là phiên và sử dụng Bộ nhớ dài hạn ngắn hạn (LSTM) và d Thuật toán Gated Current Unit (GRU) cho thấy LSTM các mô hình hoạt động tốt hơn tới 12,2% so với các mô hình GRU (Dựa trên AUC) [7].

II. VẬT LIỆU VÀ PHƯƠNG PHÁP

Để phát triển các mô hình dự đoán, điều quan trọng là để chọn dữ liệu để phân tích và các tính năng của dữ liệu, và các thuật toán được sử dụng trong mô hình dự đoán cũng phải được được chọn cho mục đích. Khi mô hình dự đoán được tạo, cần phải trải qua các bước đánh giá như thế nào chính xác dự đoán sẽ được trong các tình huống thực tế. Những cái này dòng chảy như sau:

Diagram

Description automatically generated1) Thu thập dữ liệu: Sử dụng dữ liệu nhật ký trong Học tập Hệ thống quản lý của trường đại học Cyber ( 2012.03.01~ 2019.12.01)

2) Lựa chọn tính năng: Thông tin tuần, Người dùng thông tin, Thông tin học tập trước đây, Nhật ký truy cập, Hoạt động trong lớp , Tình trạng bỏ học

3) Tiền xử lý dữ liệu: Sử dụng Oversampling, Chuẩn hóa, kỹ thuật mã hóa một lần nóng, v.v.

4) Lựa chọn thuật toán: Cây quyết định, Ngẫu nhiên- Rừng (RF), Máy Vector hỗ trợ (SVM), Thần kinh sâu Mạng(DNN)

5) Đánh giá: Độ chính xác, Thu hồi (Độ nhạy), Độ chính xác, F-đo, đường cong ROC

1. Thu thập dữ liệu

Graphical user interface

Description automatically generatedNghiên cứu này sử dụng 98.685 thông tin thống kê của sinh viên từ tháng 3 năm 2012 đến tháng 12 năm 2019 và 1.480.275 dữ liệu nhật ký được lưu trữ trong hệ thống quản lý học tập phục vụ học tập trực tuyến. Thông tin thống kê của học sinh, chẳng hạn như số lượng tuyển sinh, tình trạng học bổng, tuổi và đăng ký khóa học, được sưu tầm trong quản lý học hành chính hệ thống (ADS), và hồ sơ truy cập hàng tuần và học tập hồ sơ hoạt động được thu thập trong quản lý học tập hệ thống (LMS).

1. Lựa chọn tính năng

Các điều kiện để học trực tuyến bỏ học sớm khác nhau theo các mục tiêu của mô hình dự đoán, tùy thuộc vào việc khóa học đã được hoàn thành hay chưa, liệu có học lại ghi danh xảy ra trong học kỳ tiếp theo, lớp, vv Nghiên cứu này nhấn mạnh việc bỏ học sớm trong các khóa học trực tuyến của các trường đại học, và vì vậy liệu một người học có đăng ký lại vào trường học kỳ sau được đặt ra như một điều kiện để bỏ học. TRONG để dự đoán học sinh bỏ học như một đơn vị khóa học của các khóa học như MOOC, nó được coi là hợp lý để thiết lập các điều kiện như liệu khóa học đã được hoàn thành và điểm số như điều kiện bỏ học

A picture containing timeline

Description automatically generatedTable

Description automatically generated with low confidenceViệc lựa chọn các tính năng sử dụng được coi là quan trọng đã được sử dụng như cả hai điều kiện cho học sinh bỏ học và dự đoán Phân tích. Trong nghiên cứu này, các nghiên cứu trước đây về phân tích học trực tuyến bỏ học sớm đã được xem xét và các tính năng kiểm tra mà được coi là có một mức độ cao của ảnh hưởng đến học sinh bỏ học và lựa chọn cuối cùng là các tính năng có thể được sử dụng và thu thập từ một trường đại học trực tuyến học hệ thống LMS. Các tính năng được sử dụng trong nghiên cứu này là được hiển thị trong Bảng 1.

Trong công việc này, chúng tôi sử dụng tổng cộng 23 tính năng dữ liệu bằng cách sử dụng tổng số trong số 22 tính năng đầu vào và các mục đăng ký lại để bỏ học sớm bản án. Các tính năng chính đã chọn được mô tả trong tiếp theo:

• Tuần: Là thông tin đơn vị phân tích, dự đoán. Đối với trường đại học, thông tin từ 1 đến 15 tuần đã được sử dụng vì việc học diễn ra trong một học kỳ từ 1 đến 15 tuần

• Học bổng: Sinh viên có nhận được học bổng hay không học bổng tại thời điểm ghi danh cho học kỳ này

• Previous Degree: Thông tin bằng cấp trước khi vào đại học (cao đẳng hoặc trung học)

• Chuyển trường: Dù là sinh viên năm nhất hay chưa, các trường chuyển trường có thể có ý chí học tập mạnh mẽ hoặc có một sự phẫn nộ mạnh mẽ với việc học trực tuyến.

• Nhiều chuyên ngành được chọn: Dù bạn đã hoàn thành chuyên ngành kép hoặc chuyên ngành phụ khác với chuyên ngành chính, và nếu bạn chọn chuyên ngành kép, bạn rất sẵn lòng học.

• Tổng truy cập tính đến tuần hiện tại (lms): Số của các hệ thống học tập trực tuyến cho đến tuần hiện tại trên mà họ dựa vào. Càng nhiều lần, càng ít có khả năng

• Điểm tổng thể của các khóa đào tạo trực tuyến trước đây: Điểm cao hơn cho đến học kỳ trước ít có khả năng đi chệch hướng; lớp thấp hơn có nhiều khả năng đi chệch hướng

• Tình trạng bỏ học: Như các trường đại học thường tính toán bỏ học dựa trên việc rút đăng ký tiếp theo học kỳ, có đăng ký lại học kỳ tiếp theo đã được chọn như một biến phụ thuộc của mô hình.

Nhiều nghiên cứu trước đây đã sử dụng liệu có nên đi chệch khỏi điểm của môn học như một tiêu chí để xác định xem có nên đăng ký lại học kỳ tiếp theo, dễ dàng đăng ký các trường đại học, như một tiêu chí để xác định xem có nên học lại ghi danh.

Chart, scatter chart

Description automatically generatedBằng cách kiểm tra mối tương quan giữa các tính năng được chọn, chúng tôi đã xác nhận mối tương quan giữa 22 tính năng được chọn để loại trừ các tính năng không liên quan đến bỏ học khỏi phân tích và giá trị đăng ký lại xác định xem bỏ học.

Trong nghiên cứu này, không có tính năng nào bị loại trừ khỏi phân tích tương quan vì 22 tính năng có tương quan cao đã được lựa chọn thông qua các nghiên cứu trước đây về trích xuất tính năng.

1. Tiền xử lý dữ liệu

Quá trình tiền xử lý dữ liệu để tạo ra dữ liệu được thu thập phù hợp để phân tích với các thuật toán học máy là thiết yếu. Trong nghiên cứu, các phương pháp sau được sử dụng trong Quá trình tiền xử lý dữ liệu:

1) Chuyển đổi dữ liệu thành số: Chuyển đổi dữ liệu chuỗi như vậy như học bổng, giới tính, bằng cấp trước đây, địa chỉ (trong nước/nước ngoài), nhiều chuyên ngành và tình hình tuyển sinh dữ liệu cho học kỳ tiếp theo thành số.

2) Xóa dữ liệu không đầy đủ (thiếu) : Dữ liệu sinh viên những người không có thông tin cho học kỳ trước, chẳng hạn như trở lại trường học hoặc nhập cảnh lại, sẽ bị xóa.

3) Dữ liệu tiếng ồn loại bỏ: Thông tin độ trước đó, thông tin địa chỉ, đôi khi thiếu giá trị, vì vậy dữ liệu là đã xóa

4) Diệt trừ dữ liệu mâu thuẫn: như khi một người nam số an sinh xã hội bắt đầu bằng 2, v.v.

5) Giải quyết mất cân bằng dữ liệu : Oversampling of dropout dữ liệu nhãn

D. Lựa chọn thuật toán

Đối với các dự đoán bỏ học sớm, điều quan trọng nhất là chọn mô hình dự đoán (thuật toán) tốt nhất. Công việc này sử dụng máy học, một thuật toán trí tuệ nhân tạo đó là hiệu quả để xử lý một lượng lớn dữ liệu, tham khảo nghiên cứu trước về dữ liệu lớn để thực hiện dự đoán tối ưu người mẫu. Đặc biệt, các thuật toán học sâu được tiến hành thí nghiệm để tăng hiệu suất và độ chính xác thông qua tối ưu hóa siêu tham số. Mỗi thuật toán có các đặc điểm sau:

1) Cây quyết định: Thuật toán được sử dụng rộng rãi cho vấn đề phân loại và hồi quy, đặc biệt là có lợi thế là hiểu rõ cách các thuật toán được dự đoán bằng trực quan, trong khi có bất lợi của sự suy giảm hiệu suất do lấy mẫu quá mức của dữ liệu huấn luyện.

2) Random-Forest (RF): Thuật toán phổ biến nhất mô hình được sử dụng để phân loại và hồi quy trong quyết định mô hình cây giải quyết sự suy giảm hiệu suất vấn đề gây ra bởi oversampling của dữ liệu đào tạo. Tuy nhiên, Random Forest sử dụng nhiều bộ nhớ hơn các mô hình tuyến tính và có nhược điểm là đào tạo và dự đoán chậm.

3) Support Vector Machine (SVM): Thuật toán có lợi thế của hoạt động với một số lượng nhỏ dữ liệu đặc điểm, nhưng nó có nhược điểm là có tốc độ và các vấn đề về bộ nhớ khi tăng kích thước mẫu, và nó rất khó hiểu làm thế nào dự đoán được xác định trong các phân tích khó khăn.

4) Mạng lưới thần kinh sâu (DNN): Thuật toán được thiết kế đối với các lớp ẩn sâu của mạng lưới thần kinh nhân tạo, đó là hiệu quả trong việc phân biệt nội dung hoặc tính năng chính trong vật liệu phức tạp theo cách gần giống nhất tư duy của con người. Đặc biệt, hiệu suất tốt được thể hiện trong phân tích hình ảnh và phân tích ngôn ngữ tự nhiên, và tốt hiệu suất có thể được mong đợi trong khối lượng dữ liệu cao. Tuy nhiên, thành phần lớp ẩn và siêu tham số điều chỉnh là rất quan trọng vì những bất lợi của các vấn đề về trang bị quá mức và tốn thời gian cho việc học

E. Sự đánh giá

Text

Description automatically generatedCác mô hình dự đoán được tạo từ dữ liệu đào tạo nên kiểm tra tính chính xác của các dự đoán so với dữ liệu xác minh và dự đoán từ dữ liệu chưa được đào tạo cần được kiểm chứng bằng cách phản ánh toàn diện không chỉ chính xác mà còn nhớ lại, độ chính xác, v.v. Vì vậy, trong công việc này, chúng tôi chứng minh thông qua các phép đo F của ROC đường cong và các giá trị AUC. Các phương pháp đánh giá được sử dụng trong nghiên cứu này như sau Bảng II

Phương pháp đánh giá trên dựa trên sự nhầm lẫn ma trận như trong Bảng 3, và mỗi chỉ số được sử dụng trong đánh giá như sau:

• True Positive (TP): bỏ học (thực tế) → bỏ học (sự dự đoán)

• False Positive (FP): không bỏ học (thực tế) → bỏ học (dự đoán) Đánh giá sai

• False Negative (FN): bỏ học (thực tế) → không bỏ học (dự đoán) Đánh giá sai

Table

Description automatically generated• True Negative (TN): không bỏ học (thực tế) → không bỏ học (dự đoán)

F. Thiết kế mạng lưới thần kinh sâu

Diagram

Description automatically generatedSự phát triển của mô hình được sử dụng trong công việc này đã được thực hiện sử dụng các thư viện Keras và Tensor Flow, làm cho nó đơn giản để thực hiện ngoại trừ các mô hình học sâu. Tuy nhiên, đối với mô hình deep learning, các bước thiết kế mạng deep learning là cần thiết và sự khác biệt về hiệu suất phát sinh tùy thuộc vào phương pháp thiết kế của họ. Kết quả deep learning deep neural mô hình mạng được sử dụng trong công việc này được hiển thị trong Hình 4.

Trong các mô hình DNN (Deep Neural Network), hàm chi phí đã được sử dụng như binary\_crossentropy, 22 tham số đầu vào, 3 các lớp ẩn và hàm đầu ra giữa các lớp là được sử dụng như Relu. Lớp đầu ra cuối cùng đã sử dụng chức năng sigmodi. Hơn nữa, chức năng tối ưu hóa đã chọn Adam.

III. KẾT QUẢ VÀ THẢO LUẬN

Thí nghiệm đã thu thập và sử dụng 1.390.650 log thực tế hồ sơ và dữ liệu lịch sử từ dữ liệu học tập của 98.685(98.685 \* 15 tuần) học sinh từ tháng 3 năm 2012 đến tháng 12 năm 2019 như sau Hình 5.

Table

Description automatically generatedA picture containing table

Description automatically generatedThí nghiệm thực hiện phân tách 7:3 thu được và dữ liệu được xử lý trước vào tập dữ liệu học tập và tập dữ liệu thử nghiệm, dẫn đến việc học các mô hình cây quyết định, rừng ngẫu nhiên mô hình, mô hình máy vector hỗ trợ và mô hình DNN.

A picture containing graphical user interface

Description automatically generatedChúng tôi xác minh hiệu suất của dự đoán bằng cách áp dụng một người thử nghiệm mô hình đã học. Việc xác minh hiệu suất kết quả cho từng mô hình như sau

Kết quả kiểm định dự báo bằng Quyết định Mô hình cây cho thấy hiệu suất cao hơn đáng kể với 0,91% độ chính xác và AUC là 0,82 như hình trên. Tuy nhiên, các tốc độ học chậm dự kiến sẽ gây ra thời gian xử lý và các vấn đề về bộ nhớ khi áp dụng cho các hệ thống dự đoán thực tế.

Chart, line chart

Description automatically generatedChart

Description automatically generatedCác mô hình Rừng ngẫu nhiên sau đây cho thấy tốt nhất hiệu suất với độ chính xác 0,96%, AUC là 0,95 và tốc độ xử lý nhanh nhất.

Chart

Description automatically generated Các mô hình Máy Vector Hỗ trợ sau đây cho thấy hiệu suất thấp nhất với độ chính xác 0,81% và AUC 0,69 hiệu suất. Nó cũng mất quá nhiều thời gian để tìm hiểu, đó là không phù hợp với hệ thống dự đoán của trường đại học.

Cùng với việc phát triển các mô hình dự báo tốt hiệu suất, công việc này nhằm mục đích xác định tiềm năng cho ứng dụng của các phương pháp học sâu. thí nghiệm cho thấy rằng chúng tôi đã sử dụng thuật toán học sâu cơ bản nhất, nhưng với độ chính xác 0,85% và AUC 0,77, chúng tôi có thể thể hiện hiệu suất tuyệt vời. Kết quả thí nghiệm như bảng 5 dưới.

Chart, bar chart

Description automatically generatedTable

Description automatically generatedNhư thể hiện trong bảng này, các mô hình rừng ngẫu nhiên thực hiện tốt nhất với độ chính xác 96% , F-đo 0,84% và ACU 0,95%, trong khi các mô hình học sâu (DNN) thể hiện độ chính xác giá trị 0,85% và ACU 0,77 với Adam và ba ẩn các lớp tính năng tối ưu hóa.

IV. PHẦN KẾT LUẬN

Nghiên cứu này đã phát triển và xác nhận học máy và các mô hình học sâu để dự đoán việc bỏ học trực tuyến sớm người học. Đặc biệt, chúng tôi đã phát triển một mô hình dự đoán những người bỏ học sớm bằng cách kiểm tra hiệu suất hàng tuần, vì vậy dự đoán hàng tuần có thể được thực hiện để ngăn chặn sớm bỏ học trực tuyến, và trong trường hợp này, nó đã được xác nhận rằng Random-Forest là hiệu quả nhất. Ngoài ra, chúng tôi đã điều tra khả năng sử dụng sớm dự đoán bỏ học của các mô hình DNN thông qua nhiều thí nghiệm thông qua điều chỉnh siêu tham số của độ sâu học hỏi. Kết quả là, độ chính xác cao 85% đã được xác nhận, và với hiệu suất đạt được hơn nữa, người ta mong đợi rằng sâu kỹ thuật học tập có thể được sử dụng để dự đoán việc bỏ học sớm của người học trực tuyến. Chúng tôi dự định cải thiện khả năng ứng dụng và hiệu suất của các mô hình học sâu khác nhau để phân phối dự đoán sử dụng CNN và RNN trong quá trình học sâu kỹ thuật trong tương lai.Table

Description automatically generated