



DOI:10.22144/ctujos.2024.294

ỨNG DỤNG PHÂN LOẠI BẰNG MÁY VECTOR HỖ TRỢ TRONG ĐÁNH GIÁ TÍNH TÍCH CỰC HỌC TẬP CỦA SINH VIÊN

Lê Khánh Duy, Nguyễn Ngọc Đăng Khoa, Trương Huỳnh Kỳ, Huỳnh Phước Nghĩa và Lê Phương Thảo*

Bộ môn Sư phạm Toán, Khoa Sư phạm, Trường Đại học Cần Thơ

*Tác giả liên hệ (Corresponding author): lpthao@ctu.edu.vn

Thông tin chung (Article Information)

Nhận bài (Received): 31/12/2023

Sửa bài (Revised): 28/01/2024

Duyệt đăng (Accepted): 20/02/2024

Title: Using support vector machine classification to evaluate positive learning attitude of students

Author(s): Le Khanh Duy, Nguyen Ngoc Dang Khoa, Truong Huynh Ky, Huynh Phuoc Nghia and Le Phuong Thao*

Affiliation(s): Can Tho University

TÓM TẮT

Bài toán phân loại đã được vận dụng trong nhiều lĩnh vực khác nhau, đặc biệt trong y học. Thái độ học tập tích cực là một trong những yếu tố chủ chốt quyết định trực tiếp đến kết quả học tập của sinh viên trong môi trường học tập tự chủ ở bậc đại học. Nghiên cứu này xây dựng bộ dữ liệu đánh giá tính tích cực trong học tập, sau đó áp dụng mô hình phân loại hiệu quả cho bộ dữ liệu này và đề xuất mô hình phân loại để chẩn đoán kịp thời "bệnh chán học" cho cá nhân sinh viên, từ đó có các biện pháp hỗ trợ kịp thời.

Từ khóa: Bệnh chán học, máy vector hỗ trợ, phân loại, thái độ học tập tích cực

ABSTRACT

The classification problem has been used in many fields, especially in medicine. A positive learning attitude is one of the critical factors directly influencing students' academic performance in an autonomous learning environment at the university level. This study builds a data set to evaluate the positive aspects of learning, then applies an effective classification model to this data set and proposes a classification model to diagnose "learning fatigue" promptly in individual students, thereby implementing timely support.

Keywords: Classification, learning fatigue, positive learning attitude, support vector machine

1. GIỚI THIỆU

Mô hình Phân loại sử dụng máy vector hỗ trợ (Support Vector Machine, SVM) được đề xuất bởi Cortes and Vapnik (1995) dựa trên nguyên lý học thống kê và ngày càng nhận được sự quan tâm nghiên cứu của các nhà khoa học trong nhiều lĩnh vực khác nhau như toán học, khoa học dữ liệu, y học, kinh tế tài chính, xã hội (Vapnik, 1998; Tong & Koller, 2002; Huang et al., 2018; Imandoust & Bolandraftar, 2013). Mục tiêu của bài toán phân loại sử dụng SVM là tìm ra siêu phẳng tốt nhất để phân

tách các điểm dữ liệu trong không gian đặc trưng theo các lớp khác nhau đã gán nhãn trước. Siêu phẳng này với vai trò là đường ranh giới, sẽ tối đa hóa khoảng cách đến các điểm dữ liệu gần nhất của các lớp dựa vào các vector hỗ trợ, từ đó giảm thiểu sự sai sót trong phân loại.

Các kết quả thu được của mô hình phân loại SVM đã được ghi nhận, đặc biệt trong lĩnh vực y học (Noble, 2004; Huang et al., 2018). Trong giáo dục, "bệnh chán học" hay còn được gọi là hiện tượng mất động lực học tập, là một trong những vấn

đề đáng lo ngại. Nó thường xuất hiện khi người học mất đi sự hứng thú và định hướng trong quá trình học tập, dẫn đến thái độ học tập không tích cực và không muốn tham gia vào hoạt động học tập. Kết quả học tập của người học phụ thuộc vào nhiều yếu tố chủ quan và khách quan, nhưng then chốt vẫn phụ thuộc vào chủ thể của người học. Điều này thể hiện rõ nét ở thái độ đối với việc học tập của người học. Thái độ học tập (learning attitude) được biểu hiện qua việc chủ động học bài và chuẩn bị bài, tham gia các hoạt động học tập, quản lý và lên kế hoạch học tập, sự quan tâm đến kết quả học tập.

Ở bậc giáo dục phổ thông, giáo viên chủ nhiệm lớp đồng thời cũng là giáo viên bộ môn của lớp có tương tác thực với học sinh trong giờ học bộ môn, có cơ hội nắm bắt được tinh thần thái độ học tập của cá nhân học sinh. Ngoài ra, các thông tin theo dõi tập thể lớp được ghi nhận chi tiết, cập nhật liên tục trong sổ đầu bài quản lý lớp. Cho nên, giáo viên chủ nhiệm theo dõi sát sao, nhắc nhở và thường xuyên đôn đốc học sinh tích cực học tập. Ở bậc đại học, do môi trường học của sinh viên cần tính tự chủ ý thức cao từ phía người học và thoáng hơn về mặt quản lý, nên việc theo dõi sát sao thái độ học tập của sinh viên gặp nhiều khó khăn. Có vấn đề học tập, với vai trò là người tư vấn giám sát việc thực hiện học tập theo kế hoạch, rất ít cơ hội tương tác thật trong giờ học để có thể nắm bắt được tình hình thái độ học của sinh viên trực thuộc lớp mình quản lý. Vượt qua kỳ tuyển sinh để đảm bảo ngưỡng đầu vào và khi mới nhập học, các tân sinh viên thường rất háo hức, đầy nhiệt huyết với môi trường học tập mới. Nhưng không lâu sau đó, tình trạng sinh viên “đánh mất” thái độ học tập tích cực tăng lên làm cho kết quả học tập giảm sút, và đây đang là vấn đề đáng báo động. Số lượng sinh viên vi phạm quy chế học vụ dẫn đến cảnh báo học vụ và bị đuổi học do không đáp ứng quy định là điều trăn trở cho cả phía người học và phía quản lý. Khi rơi vào tình trạng bị cảnh báo học vụ thì hầu hết đã muộn, sinh viên đã xa rời, mất động lực học, không đảm bảo ngưỡng chất lượng và gặp nhiều khó khăn để khắc phục. Vì vậy, việc nắm bắt kịp thời tình hình thái độ học tập của sinh viên, phân loại một cách dễ dàng, hiệu quả thái độ học tập đang tích cực hay chưa đủ đáp ứng chương trình học của sinh viên là cần thiết và có ý nghĩa thiết thực. Từ đó, việc chủ động có các biện pháp hỗ trợ những sinh viên đang mất dần thái độ học tập tích cực là điều đáng được quan tâm và có ý nghĩa.

Nguyên nhân phổ biến gây ra "bệnh chán học" và một số biện pháp để giải quyết vấn đề này đã được đề cập trong các nghiên cứu (Hải & Nguyệt, 2012; Huyền, 2018). Tuy nhiên, trong môi trường

đại học, không dễ để xác định được cá thể sinh viên nào đang trong tình trạng “mất dần” thái độ học tập tích cực, đang cần nhận được sự quan tâm tư vấn hỗ trợ. Mục tiêu của nghiên cứu này là vận dụng mô hình phân loại sử dụng máy vector hỗ trợ để xây dựng công cụ phân loại theo phương pháp học máy nhằm phân loại, đánh giá tính tích cực trong học tập của sinh viên, kịp thời nhận biết và cảnh báo những trường hợp người học đang có thái độ học tập chưa đạt, cần thiết phải hỗ trợ đúng lúc. Để đạt được mục tiêu trên, phương pháp thông kê và phương pháp nghiên cứu lý thuyết được kết hợp sử dụng. Trước hết, mô hình phân loại sử dụng máy vector được nghiên cứu về cả lý thuyết toán học và cả nguồn code máy tính của mô hình với các loại hàm nhân khác nhau. Tiếp theo, các kết quả nghiên cứu đã công bố về vai trò, các yếu tố tác động đến tính tích cực, chủ động trong học tập của sinh viên được nghiên cứu, từ đó xây dựng bộ câu hỏi, tiến hành khảo sát và thống kê dữ liệu từ các sinh viên. Sau đó, các dữ liệu được tiền xử lý và tiến hành thực nghiệm phân loại với mô hình phân loại đã đề xuất. Các kết quả thu được trong nghiên cứu này là hoàn toàn mới, tiên phong vận dụng máy học SVM trong nghiên cứu ứng dụng giải quyết vấn đề thực tiễn trong lĩnh vực giáo dục, khắc phục được trở ngại trong phân loại dữ liệu dạng định tính.

2. BÀI TOÁN PHÂN LOẠI BẰNG MÁY VECTOR HỖ TRỢ

Hiện nay, có rất nhiều phương pháp phân loại được đề xuất và áp dụng. Theo các phương pháp thống kê truyền thống, có phương pháp Fisher, Logistic, Bayes,... (Vovan et al., 2017). Với kỹ thuật học máy và học sâu, có nhiều phương pháp phân loại khác như: Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Artificial Neural Network (ANN),... (Ma et al., 2019).

Trong bài viết này, phân loại máy học SVM xây dựng mô hình phân loại được vận dụng nhằm đánh giá tính tích cực học tập của sinh viên. Do đánh giá sự tích cực mang yếu tố định tính, thang đo sự tích cực được lượng hóa theo các thành tố có ảnh hưởng đến mức độ tích cực trong thái độ học tập của người học. Việc xây dựng tập dữ liệu mẫu huấn luyện có sự chọn lọc, tham vấn từ các bên liên quan, kết hợp với lợi thế kỹ thuật máy học SVM, đem lại hiệu quả trong việc xây dựng mô hình đánh giá.

Trước hết, ta nhắc lại mô hình bài toán phân loại hai lớp sử dụng các vector hỗ trợ. Với m mẫu ban đầu của tập các mẫu $X = \{x_1, x_2, \dots, x_m\}$, mỗi mẫu được đặc trưng bởi n thông số, ta cần gán mỗi mẫu với một trong hai lớp (nhãn) trong tập nhãn $Y =$

$\{-1; +1\}$. Các mẫu dữ liệu được gán vào lớp cụ thể thông qua một hàm quyết định. Mục đích của bài toán là sử dụng các hàm toán học nhằm xây dựng một siêu phẳng làm cực tiểu hóa độ phân lớp sai của một đối tượng dữ liệu mới. Độ phân lớp sai của một siêu phẳng được đặc trưng bởi khoảng cách bé nhất tới siêu phẳng đó.

Ta xét trường hợp bài toán khả tách tuyến tính hoàn toàn trong không gian đặc trưng. Như vậy, với tập dữ liệu huấn luyện

$$T = \{(x_1, y_1), \dots, (x_m, y_m)\},$$

với $x_i \in \mathbb{R}^n$, $y_i = \{-1, +1\}$, $i = 1, \dots, m$, bằng một hàm tuyến tính các mẫu đều được phân vào đúng lớp. Dạng tuyến tính của siêu phẳng (H) cần tìm có công thức $h(x) = w \cdot x + b = 0$. Khi đó, ta có hai bất phương trình mô tả tính chất của hai lớp phân loại dữ liệu như sau. Bất phương trình

$$w \cdot x_i + b \leq -1,$$

với $y_i = -1$ và bất phương trình

$$w \cdot x_i + b \geq +1,$$

với $y_i = +1$.

Kết hợp hai bất phương trình trên ta có dạng chung cho tất cả các mẫu là

$$y_i(w \cdot x_i + b) \geq 1.$$

Hai siêu phẳng hỗ trợ (H_1), (H_2) song song với (H) là siêu phẳng (H_1): $w \cdot x_i + b = -1$ và siêu phẳng (H_2): $w \cdot x_i + b = 1$. Ta nhận thấy rằng các vector nằm trên hai siêu phẳng hỗ trợ này là các vector gần (H) nhất và được gọi là các vector hỗ trợ. Chúng ta cũng thấy rằng hàm quyết định chỉ phụ thuộc vào các vector này. Như vậy, ta cần tìm w và b sao cho tất cả các mẫu được gán đúng lớp và tối đa hóa khoảng cách lề, $\frac{2}{\|w\|}$, khoảng cách giữa hai siêu phẳng hỗ trợ. Ở đây $\| \cdot \|$ được hiểu là chuẩn Euclide.

Do vậy, bài toán phân loại sử dụng các vector hỗ trợ chính là bài toán tối ưu có dạng như sau:

$$\max_{w,b} \frac{1}{\|w\|}$$

thỏa mãn

$$(w \cdot x) + b \geq 1, \forall i: y_i = 1,$$

$$(w \cdot x) + b \leq -1, \forall i: y_i = -1.$$

Ta đưa bài toán về bài toán quy hoạch toàn phương như sau:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

thỏa mãn

$$y_i((w \cdot x) + b) - 1 \geq 0, \forall i = 1, \dots, m.$$

Hơn nữa, hàm mục tiêu của bài toán quy hoạch toàn phương này là hàm lồi chặt, nên cực tiểu địa phương cũng là cực tiểu toàn cục của bài toán. Do vậy, nghiệm của bài toán là duy nhất. Việc giải tìm nghiệm có thể thực hiện thuận lợi thông qua thư viện CVXOPT trong Python.

Khi đã giải được w, x ta xác định được siêu phẳng phân tách hai lớp, $h(x) = w \cdot x + b$. Việc phân loại và gán nhãn cho một điểm dữ liệu mới sẽ được thực hiện bằng cách sử dụng hàm dấu của $h(x)$. Cụ thể, với x là mẫu dữ liệu mới, x được phân loại và gán nhãn theo công thức như sau:

$$f(x) = \text{sign}(h(x)) = \begin{cases} +1 & \text{ khi } h(x) \geq 0, \\ -1 & \text{ khi } h(x) < 0. \end{cases}$$

Phân loại SVM có một số ưu điểm như sau:

- Hiệu suất tốt trong các không gian đặc trưng lớn: SVM hoạt động tốt trong các không gian đặc trưng có số chiều lớn. Điều này làm cho SVM phù hợp với việc xử lý dữ liệu có số lượng đặc trưng lớn, như trong các bài toán xử lý hình ảnh hoặc xử lý văn bản.

- Hiệu suất tốt với dữ liệu hạn chế: SVM cũng hoạt động tốt khi dữ liệu huấn luyện có kích thước nhỏ hoặc số lượng mẫu ít. Điều này là do SVM chỉ cần một số lượng nhỏ các vector hỗ trợ để xác định đường ranh giới tối ưu.

- Khả năng xử lý dữ liệu nhiễu: SVM tập trung vào việc tối đa hóa lề (margin), tức là khoảng cách giữa đường ranh giới và các điểm dữ liệu gần nhất của các lớp khác nhau. Điều này giúp SVM có khả năng tốt trong việc xử lý dữ liệu nhiễu và giảm hiện tượng quá khớp (overfitting).

Trong trường hợp bài toán không khả tách tuyến tính hoàn toàn trong không gian đặc trưng, ta cần hỗ trợ từ các hàm nhân (kernel). Sử dụng các hàm nhân để ánh xạ dữ liệu từ không gian đặc trưng ban đầu vào không gian mới mà ở đó bài toán trở thành khả tách tuyến tính. Các hàm nhân phổ biến là:

- Linear kernel: $k(x, y) = x^T y$.

- Polynomial kernel: $k(x, y) = (\gamma x^T y + c_0)^d$.

- Sigmoid kernel: $k(x, y) = \tanh(\gamma x^T y + c_0)$.

- RBF kernel: $k(x, y) = \exp(-\gamma||x - y||^2)$.

3. XÂY DỰNG MÔ HÌNH PHÂN LOẠI ĐÁNH GIÁ TÍCH TÍCH CỰC HỌC TẬP CỦA SINH VIÊN

3.1. Vai trò của thái độ học tập

Tích tích cực học tập đóng vai trò quan trọng trong quá trình học tập và tác động đến các yếu tố sau:

- Tạo động lực học tập: Thái độ tích cực và đam mê đối với việc học giúp tạo ra động lực mạnh mẽ để tiếp thu kiến thức. Khi có thái độ đúng đắn và quan tâm đến học tập, người học sẽ cảm thấy hứng thú và muốn khám phá về các chủ đề mới.

- Khả năng quản lý học tập và giải trí: Thái độ học tập tích cực giúp người học phát triển kỹ năng quản lý thời gian và tạo ra sự cân bằng giữa công việc học tập và hoạt động giải trí. Thay vì lười biếng và trì trệ, một thái độ học tập tích cực khuyến khích người học tổ chức công việc và tận dụng thời gian hiệu quả.

- Tương tác xã hội tốt: Thái độ học tập tích cực cũng tạo điều kiện cho việc tương tác xã hội tích cực trong quá trình học. Người học sẽ muốn tham gia vào các hoạt động nhóm, thảo luận và chia sẻ ý kiến với người khác. Điều này không chỉ cung cấp cơ hội học hỏi từ người khác mà còn giúp xây dựng mối quan hệ và mạng lưới xã hội trong lĩnh vực học tập.

- Tạo điều kiện học hỏi: Thái độ học tập tích cực thúc đẩy việc tạo ra một môi trường học tập tích cực. Người học sẽ tìm kiếm cơ hội học hỏi bằng cách đặt câu hỏi, tham gia vào các hoạt động ngoại khóa, tìm kiếm nguồn tư liệu và tìm hiểu thêm về chủ đề mà họ quan tâm. Thái độ này khuyến khích việc khám phá và sáng tạo, giúp người học phát triển khả năng tự học và mở rộng kiến thức.

3.2. Nguyên nhân dẫn đến thiếu tích cực học tập

Thực trạng mất động lực học có thể ảnh hưởng đến hiệu suất và khả năng học tập. Dưới đây là một số thực trạng chung và nguyên nhân có thể dẫn đến tình trạng:

- Thiếu mục tiêu rõ ràng: khi không có mục tiêu cụ thể và rõ ràng trong quá trình học, người học có thể cảm thấy mất hứng thú và không biết hướng đi của mình.

- Môi trường học tập không thúc đẩy: môi trường học tập không đủ kích thích hoặc không tạo điều kiện thuận lợi để học, như không có nguồn tài

liệu phong phú, không có sự hỗ trợ từ giáo viên hoặc không có sự thú vị trong quá trình học.

- Áp lực và căng thẳng: áp lực từ gia đình, xã hội hoặc căng thẳng về việc đạt được kết quả cao có thể làm mất đi sự đam mê và động lực học.

- Thiếu sự quan tâm và liên quan: khi người học không thấy giá trị hay ý nghĩa của kiến thức đối với cuộc sống và mục tiêu cá nhân của mình, họ có thể mất đi động lực để học.

- Kiến thức không được truyền đạt một cách thú vị: phương pháp giảng dạy không hấp dẫn hoặc không phù hợp với phong cách học của người học có thể làm mất đi sự hứng thú và động lực.

- Mất tự tin: khi người học không tự tin vào khả năng của mình hoặc đã có những trải nghiệm thất bại, họ có thể mất đi động lực để tiếp tục học tập.

3.3. Mô hình phân loại

- Thu thập dữ liệu: do nghiên cứu phân loại về thái độ học tập của sinh viên là chưa sẵn có nên cần thiết phải xây dựng dữ liệu huấn luyện cho mô hình phân loại. Hơn nữa, do đặc điểm của kỹ thuật máy học, các dữ liệu dùng để huấn luyện mô hình phải chuẩn xác, đáng tin cậy. Vì vậy, nguồn dữ liệu dùng cho nghiên cứu này đã được xây dựng. Bộ câu hỏi dùng trong thu thập dữ liệu được đề xuất một cách chọn lọc và có tham vấn của các bên liên quan bao gồm cố vấn học tập, giảng viên trực tiếp giảng dạy, tham vấn giảng viên tâm lý giáo dục, đồng thời cũng có tham khảo đánh giá ngang hàng từ người học trong cùng khóa học. Cụ thể, bộ dữ liệu gồm 143 mẫu được thu thập thông qua việc khảo sát đối với sinh viên sư phạm toán học thuộc Khoa Sư phạm, Trường Đại học Cần Thơ. Cơ cấu mẫu theo các khóa học như sau: khóa 45 gồm 12 sinh viên, khóa 46 gồm 19 sinh viên, khóa 47 gồm 64 sinh viên, khóa 48 gồm 17 sinh viên, khóa 49 gồm 31 sinh viên.

Sở dĩ cơ cấu mẫu không đồng đều giữa các khóa là do việc lấy mẫu trên tinh thần tự nguyện và đóng góp thông tin tin cậy của sinh viên, đồng thời cần có sự phối hợp của cố vấn học tập và giảng viên có trực tiếp giảng dạy trong vai trò xác nhận mẫu có đạt yêu cầu để làm mẫu dữ liệu trong tập huấn luyện hay không.

Mỗi mẫu dữ liệu trong bộ dữ liệu được xem như là một vector gồm 40 thông số được phân theo 9 nhóm tiêu chí và được đo theo 5 mức độ. Cụ thể, các nhóm tiêu chí như sau:

- 1) chủ động học bài và chuẩn bị bài;
- 2) tham gia các hoạt động trong khi học;

3) thực hiện việc quản lý và lên kế hoạch học tập;

4) quan tâm các vấn đề tạo hứng thú trong học tập;

5) hứng thú với bài học, hoạt động trong học tập;

6) ảnh hưởng của những yếu tố làm gián đoạn việc học, tự học;

7) khắc phục trở ngại trong học tập;

8) ý thức về những lợi ích của học tập cho tương lai;

9) tự đặt ra những yêu cầu về thành tích học tập của bản thân qua những mục tiêu.

– Xử lý dữ liệu: Tiền xử lý dữ liệu là một bước quan trọng để chuẩn bị dữ liệu cho việc huấn luyện mô hình. Điều này bao gồm việc xử lý dữ liệu bị thiếu, loại bỏ nhiễu, chuẩn hóa dữ liệu và chuyển đổi các biến đầu vào thành các vector đặc trưng phù hợp cho mô hình phân loại dùng máy vector hỗ trợ.

– Xây dựng mô hình phân loại: sử dụng thư viện `sklearn.svm.SVC` trong Python hoàn thiện mô hình phân loại đa lớp, với đầu ra phân thành 4 lớp như sau: lớp 0: không tích cực, lớp 1: nguy cơ, lớp 2: bình thường, lớp 3: tích cực.

– Dữ liệu sạch đã tiền xử lý (80%) được sử dụng làm tập huấn luyện, ta huấn luyện mô hình phân loại SVM. Mô hình sẽ học từ các điểm dữ liệu đã được gán nhãn với các thông tin về thái độ học tập của sinh viên.

– Tập kiểm tra: 20% dữ liệu sạch được dùng trong vai trò tập kiểm tra. Kết quả cho thấy mô hình hoạt động hiệu quả, với độ chính xác 89%. Trong nghiên cứu tiên phong này, số lượng mẫu còn hạn chế. Để nâng cao hiệu quả của mô hình và phạm vi áp dụng, bộ dữ liệu sẽ được tiếp tục bổ sung các mẫu dữ liệu của sinh viên các khóa mới cũng như sinh viên thuộc các chuyên ngành khác. Việc thực hiện cần từng bước, cẩn thận vì tính quan trọng của tập dữ liệu huấn luyện trong mô hình phân loại máy học.

– Phân loại mẫu mới: với một mẫu dữ liệu mới, mô hình sẽ phân loại và trả ra kết quả tương ứng là một trong bốn lớp đầu ra như trên.

Mô hình phân loại có thể sử dụng để đánh giá thái độ học tập của họ dựa trên các thông tin đã cho. Kết quả đánh giá có thể được sử dụng để cảnh báo về các sinh viên có nguy cơ thấp hoặc cao về thái độ học tập. Các cảnh báo này có thể được sử dụng để hướng dẫn và hỗ trợ sinh viên trong việc cải thiện thái độ và kết quả học tập của họ. Việc áp dụng mô hình SVM để đánh giá và cảnh báo thái độ học tập của sinh viên có thể cung cấp thông tin hữu ích cho nhà trường, giáo viên, hoặc cố vấn học tập để hỗ trợ sinh viên và đưa ra các biện pháp phù hợp để nâng cao hiệu suất học tập.

4. KẾT LUẬN VÀ KIẾN NGHỊ

Mô hình phân loại hỗ trợ hiệu quả trong phân loại sinh viên theo các mức phân loại xác định trước. Từ đó, giúp sàng lọc nhận biết những sinh viên có thái độ học tập chưa tích cực, nhận biết những sinh viên có nguy cơ cao hoặc thấp về thái độ học tập chưa đáp ứng chương trình học. Từ đó, cung cấp cảnh báo sớm về những sinh viên có thái độ học tập không tốt hoặc có nguy cơ thấp về hiệu suất học tập. Điều này cho phép nhà trường, cố vấn học tập và giảng viên kịp thời can thiệp và có các biện pháp hỗ trợ, tư vấn phù hợp để giúp sinh viên cải thiện thái độ và kết quả học tập.

Việc áp dụng mô hình phân loại SVM để đánh giá và cảnh báo thái độ học tập của sinh viên mang lại lợi ích quan trọng trong việc hỗ trợ sinh viên, tối ưu hóa tài nguyên và nâng cao hiệu suất học tập chung. Các kết quả thu được trong nghiên cứu này là tiền đề cho việc triển khai giải quyết phân loại dữ liệu và dự báo trong lĩnh vực giáo dục. Khi dữ liệu ở dạng định tính, việc phân loại dữ liệu bằng phương pháp máy học sẽ hiệu quả hơn so với phương pháp phân loại thông thường.

LỜI CẢM ƠN

Bài báo được tài trợ bởi đề tài nghiên cứu của Trường Đại học Cần Thơ, mã số: TSV2024-88.

TÀI LIỆU THAM KHẢO

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297. <https://doi.org/10.1007/BF00994018>

Huang, S., N. Cai, P., Pacheco, P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of supportvector machine (SVM) learning in cancer geometrics. *Cancer Genomics-Proteomics*, 15(1), 41–51. <https://doi.org/10.21873/cgp.20063>

Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (KNN) approach for pre-dicting economic events: Theoretical background. *International Journal of Engineering Research and Applications*, 3(5), 605-610.

Ma, Y., Zhang, Q., Li, D., & Tian, Y. (2019). LINEX support vector machine for large-scale

- classification. IEEE Access, 7, 70319-70331.
<https://doi.org/10.1109/ACCESS.2019.2919185>
- Noble, W. S. (2004). Support Vector Machine Applications in Computational Biology. In: Schoelkopf, B., Tsuda, K., Vert, J. P. (eds.) *Kernel Methods in Computational Biology*, pp. 71–92. MIT Press, Cambridge.
<https://doi.org/10.7551/mitpress/4057.003.0005>
- Tong, S., & Koller, D. (2002). Support Vector Machine Active Learning with Applications to Text Classification. *The Journal of Machine Learning Research*, 2(1), 45–66.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience
- Vovan, T., Chengoc, H., & Nguyentrang, T. (2017). Textural features selection for image classification by Bayesian method. In: 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 733–139.
<https://doi.org/10.1109/FSKD.2017.8393365>
- Huyền, N. T. (2018). Thực trạng tính tích cực học tập của sinh viên trường đại học kinh tế quốc dân, *Tạp chí Giáo dục*, 437(1), 23-27.
- Hải, H. T., & Nguyệt, N. T. (2012). Tính tích cực học tập và vấn đề tích cực hóa hoạt động học tập của sinh viên trong đào tạo tin chi. *Journal of social sciences, humanities and education*, 2(4), 112-115.