

# XÂY DỰNG TẬP DỮ LIỆU CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN KẾT QUẢ HỌC TẬP CỦA SINH VIÊN TRƯỜNG ĐẠI HỌC THỦ ĐÔ HÀ NỘI

Nguyễn Thị Kim Sơn, Nguyễn Hồng Hoa,  
Hoàng Thị Thu Trang, Trần Quỳnh Ngân  
Trường Đại học Thủ đô Hà Nội

**Tóm tắt:** Bài báo này trình bày kết quả nghiên cứu về xây dựng tập dữ liệu thuộc lĩnh vực khoa học giáo dục về hoạt động học tập của sinh viên ngành Khoa học tự nhiên, Khoa Sư phạm, Trường Đại học Thủ đô Hà Nội. Bộ dữ liệu không chỉ tập trung vào việc thu thập thông tin mà còn nhấn mạnh đến các yếu tố ảnh hưởng đến quá trình học tập của sinh viên nhằm định hướng hoạt động giáo dục, phân tích hỗ trợ người học trong việc đưa ra quyết định và nâng cao hiệu quả của quá trình học tập. Bộ dữ liệu được thu thập dựa trên 2 hình thức chính: trực tiếp từ cá nhân thông qua phiếu khảo sát và gián tiếp qua đơn vị quản lý đào tạo. Tập dữ liệu sau khi làm sạch và tiền xử lý bao gồm 992 mẫu với 89 trường thông tin được thu thập từ 10 khóa (K14 đến K23). Các thông tin này được chia thành ba nhóm chính: yếu tố cá nhân, các yếu tố ảnh hưởng đến kết quả học tập và kết quả học tập (bao gồm kết quả học tập ở phổ thông và điểm các học phần tại đại học). Đặc biệt, bộ dữ liệu này đã trải qua quá trình xử lý và làm sạch cùng với các công cụ phân tích thống kê sơ bộ. Từ đó, chúng tôi đề xuất một số khuyến nghị về quản lý đào tạo, phương pháp giáo dục và cách thức học tập, cho nhà trường, đội ngũ giảng viên, sinh viên nhằm nâng cao chất lượng học tập. Điều này tạo ra một nền tảng vững chắc để phục vụ cho các nhiệm vụ nghiên cứu về khoa học giáo dục, ứng dụng các phương pháp học máy và học sâu để dự đoán kết quả học tập của người học.

**Từ khóa:** Học máy, Kết quả học tập, Khoa học dữ liệu, Khoa học giáo dục, Phân tích thống kê, Tập dữ liệu, Yếu tố ảnh hưởng.

Nhận bài ngày 28.12.2023; gửi phản biện, chỉnh sửa, duyệt đăng ngày 28.03.2024  
Liên hệ tác giả: Nguyễn Hồng Hoa; Email: nhhoa@daihocthudo.edu.vn

## 1. MỞ ĐẦU

Trong những năm gần đây, sự phát triển của khoa học kỹ thuật, trí tuệ nhân tạo, học máy học sâu trong phân tích dữ liệu hỗ trợ ra quyết định ở các ngành nghề nói chung, trong đó có khoa học giáo dục nói riêng. Khác với các ngành nghề khác, khi chuyển đổi số có thể dễ dàng thực hiện trong khoảng 1 thập kỷ gần đây: ngân hàng, kinh tế, dự báo thủy văn, dự báo ảnh,... bởi những bài toán đó có sẵn tập dữ liệu do quá trình chuyển đổi số tốt, thì chuyển đổi số trong lĩnh vực giáo dục hiện tại thực hiện vẫn chưa được đồng bộ, công tác chuyển đổi số trên thế giới nói chung, và tại Việt Nam nói riêng thì vẫn còn ở những bước ban đầu. Do vậy, khi muốn ứng dụng trí tuệ nhân tạo, ứng dụng các công cụ phân

tích hiện đại như học máy học sâu vào phân tích giáo dục, rất nhiều trường hợp gặp phải dữ liệu trống (không có dữ liệu để phân tích) mà trong khoa học dữ liệu thì dữ liệu là quan trọng nhất. Thủ tướng Chính phủ đã ký Quyết định số 749/QĐ-TTg vào ngày 03/6/2020 phê duyệt “Chương trình chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030” [1], chương trình đã xác định Giáo dục là một trong 8 lĩnh vực cần ưu tiên chuyển đổi số quốc gia.

Ứng dụng phổ biến nhất của khai thác dữ liệu giáo dục là: cải thiện quá trình học tập, cải thiện việc hoàn thành khóa học, hỗ trợ sinh viên lựa chọn khóa học, lập hồ sơ sinh viên, tìm ra các vấn đề dẫn đến bỏ học, xác định mục tiêu của sinh viên, phát triển chương trình giảng dạy, dự đoán kết quả học tập của sinh viên và như một công cụ hỗ trợ đưa ra quyết định khi tuyển sinh. Trong thời đại công nghệ 4.0, việc các trường đại học cần đổi mới mô hình và nâng cao hiệu quả quản lý dựa trên dữ liệu lớn đã trở thành một bài toán cấp bách. Trong quản lý đào tạo, các trường đại học cần chuyển đổi số, tạo ra những hệ thống quản lý dựa trên công nghệ, dựa trên cơ sở dữ liệu lớn để tổ chức quản lý đào tạo, hỗ trợ ra quyết định quản lý một cách nhanh chóng, chính xác [2]. Từ việc phân tích thực trạng học tập của sinh viên, chúng tôi đưa ra sự cần thiết phải xây dựng một tập dữ liệu để phục vụ cho quá trình áp dụng các kỹ thuật phân tích dữ liệu hiện đại như học máy học sâu để hỗ trợ ra quyết định trong triển khai hoạt động giáo dục.

## 2. NỘI DUNG

### 2.1. Sự cần thiết xây dựng tập dữ liệu thông tin về các yếu tố ảnh hưởng đến kết quả học tập của sinh viên

Ngày nay, bối cảnh giáo dục có sự chuyển biến rất lớn khi điều kiện học tập của người học được nâng lên với sự đầu tư cả ở cấp độ quốc gia. Khi công nghệ tiếp tục phát triển với tốc độ chưa từng thấy, tương lai của giáo dục đang trải qua một sự chuyển đổi hay còn được gọi là ứng dụng khoa học trong giáo dục. Do đó, các nghiên cứu khoa học giáo dục đang được chuyển hướng đến nghiên cứu sâu hành vi người học dựa trên dữ liệu để thiết lập các chương trình học cá nhân. Đồng thời, khai phá dữ liệu lớn để sớm dự đoán và định hướng lại quá trình học tập của người học nói riêng, quản lý/ điều hành quá trình giáo dục nói chung [3].

Kết quả học tập đóng một vai trò quan trọng trong đánh giá chất lượng giáo dục, và là đầu ra của giáo dục đại học, đầu vào của nhân sự tuyển dụng của các cơ quan, doanh nghiệp. Thực tế cho thấy, việc đạt được một kết quả như kỳ vọng của bản thân là điều không dễ dàng, bởi kết quả học tập không chỉ đánh giá từ sự nỗ lực, phấn đấu mà còn có các tác nhân bên ngoài tác động làm ảnh hưởng đến kết quả học tập của sinh viên. Chính vì thế, từ trước đến nay, các nhà giáo dục và các nhà nghiên cứu luôn quan tâm đến việc xác định thông tin, các yếu tố ảnh hưởng đến sự thành công hay thất bại của sinh viên như: gia đình, nhà trường, điều kiện kinh tế xã hội, thời gian biểu của mỗi người,... Ví dụ như sinh viên dành thời gian học tập nhiều hơn phần lớn sẽ đạt kết quả học tập tốt hơn do họ có sự đầu tư nhiều hơn cho việc đọc sách, nghiên cứu tài liệu, phát triển tư duy. Những yếu tố bên ngoài thường gắn với gia đình (nền tảng giáo dục của cha mẹ và thu nhập), thường thì trình độ của cha mẹ sẽ ảnh hưởng phần lớn phương pháp nuôi dạy con cái cũng như định hình phẩm chất nhân cách cho con từ khi còn nhỏ. Bên cạnh đó, sự hỗ trợ của các trường đại học cũng được xem là yếu tố môi trường bên ngoài ảnh hưởng đến kết quả học tập của họ, bởi tùy từng trường học sẽ cung cấp cho sinh viên không gian thư viện với đầy đủ các tài liệu tham khảo, cơ sở vật chất thiết yếu, các hỗ trợ trong hoạt động học tập, nghiên cứu khoa học,...

Thời gian gần đây, tình trạng sinh viên ở các trường học bị cảnh báo học vụ hoặc buộc thôi học đang có chiều hướng gia tăng, điều này đáng báo động đối với sinh viên, giảng viên và nhà quản lý. Trái

ngược với quyết tâm tốt nghiệp loại Khá, Giỏi, giờ đây, nhiều sinh viên chỉ còn vài học kỳ là tốt nghiệp nhưng lại bỏ ngang giữa chừng. Bên cạnh đó, hầu hết các trường đại học đã triển khai phương thức đào tạo theo tín chỉ. Một trong những điểm mạnh của phương thức này là giúp người học chủ động xây dựng kế hoạch học tập cho bản thân, có cơ hội giúp sinh viên được tốt nghiệp sớm, để có thể tham gia ngay vào thị trường lao động. Tuy nhiên, điều này cũng khiến rất nhiều sinh viên bị lúng túng, khó khăn trong việc lựa chọn môn học với nhiều môn được giảng dạy trong một học kỳ, hoặc lập kế hoạch chưa thật sự phù hợp với hoàn cảnh, sức khỏe, hoặc năng lực học tập của bản thân. Điều này dẫn đến việc sinh viên không đảm bảo đủ tiêu chuẩn đầu ra, phải kéo dài thời gian học tập, lãng phí thời gian, tiền bạc. Khi đó, ngoài việc tự tìm hiểu sắp xếp, thì sinh viên sẽ cần đến sự trợ giúp của giảng viên (cố vấn học tập) để lên kế hoạch học tập tối ưu/ xác định chiến lược, mục tiêu học tập dài hạn, ngắn hạn nhằm đạt được kết quả học tập kỳ vọng.

Từ những phân tích nêu trên, việc cải thiện tình hình học tập cho người học giúp các nhà quản lý và các nhà giáo có được những giải pháp kịp thời, phù hợp nhằm nâng cao chất lượng, cải thiện tình hình học tập cho người học là một nhu cầu bức thiết của nhà trường. Thông qua việc nghiên cứu các yếu tố ảnh hưởng đến kết quả học tập và ứng dụng thống kê, chúng tôi xây dựng tập dữ liệu giúp các nhà quản lý có cái nhìn tổng quan để đưa ra những chính sách, kế hoạch thực hiện chương trình đào tạo phù hợp, giảm số lượng sinh viên bỏ học, thôi học, giúp sinh viên tiếp cận chương trình học tập một cách tối ưu và hiệu quả nhất.

## 2.2. Xây dựng phiếu khảo sát

### Bước 1: Xác định mục tiêu và câu hỏi nghiên cứu

Dựa trên các tài liệu về tác nhân ảnh hưởng đến kết quả học tập và thực trạng sinh viên ngành Khoa học tự nhiên, Khoa Sư phạm, Trường Đại học Thủ đô Hà Nội, 2 nhóm câu hỏi chính theo Farooq (2011) đã được đưa ra, bao gồm nhóm yếu tố cá nhân và môi trường [3].

### Bước 2: Xác định đối tượng khảo sát và mẫu khảo sát dự kiến

Đối với bài báo này, đối tượng khảo sát là sinh viên ngành Khoa học tự nhiên, Khoa Sư phạm Trường Đại học thủ đô Hà Nội từ khóa 2014 đến nay.

### Bước 3: Xác định các cách thức khảo sát và thu thập dữ liệu

Bài nghiên cứu sử dụng hình thức thu thập dữ liệu online thông qua phiếu khảo sát Google Form đối với sinh viên đang học tập tại trường và hỏi trực tiếp kết hợp trực tuyến đối với sinh viên đã tốt nghiệp. Bên cạnh đó, kết quả học tập các học phần tại đại học được cung cấp từ đơn vị quản lý đào tạo.

### Bước 4: Xác định các câu hỏi trong phiếu khảo sát

Bài nghiên cứu đưa ra 36 câu hỏi có liên quan đến 3 yếu tố chính: thông tin cá nhân, các yếu tố ảnh hưởng đến quá trình học tập và kết quả học tập trước đây.

#### (A) Câu hỏi về thông tin cá nhân (4 câu hỏi)

Phần này bao gồm mã sinh viên, lớp, giới tính, ngành học nhằm xác định danh tính sinh viên, đồng thời là cơ sở so sánh kết quả dự đoán với kết quả thực tế.

#### (B) Các yếu tố ảnh hưởng đến kết quả học tập

##### Yếu tố cá nhân (10 câu hỏi)

**Trình độ cha mẹ [2, 3]:** Theo báo cáo nghiên cứu của Cục thống kê quốc gia Anh cho thấy trình độ học vấn của bố mẹ là nhân tố quan trọng ảnh hưởng đến việc sau này con trẻ có thành công trong học tập hay không. Kết quả báo cáo cũng cho thấy tỷ lệ thành công trong học tập của những trẻ có mẹ có trình độ đại học cao gấp 3 lần so với các bạn khác.

**Công việc làm thêm [2]:** Hiện nay, làm thêm trở thành xu hướng khi sinh viên muốn được tích lũy kinh nghiệm và được tiếp cận gần hơn với ngành nghề mình theo đuổi, và cũng có sinh viên đi làm những công việc bán thời gian chỉ phục vụ cho mục đích trang trải kinh tế cho quá trình sinh sống và học tập. Do vậy, sẽ có những ảnh hưởng, tác động khác nhau ứng với mục đích riêng của mỗi người.

**Thời gian sử dụng mạng xã hội:** Theo số liệu thống kê của Viện Chiến lược thông tin và truyền thông, Bộ Thông tin và Truyền thông, thời lượng sử dụng mạng xã hội của sinh viên trung bình là 5 giờ/ngày. Có những sinh viên sử dụng mọi lúc, mọi nơi, mọi thời điểm đến mức gây ra tình trạng “nghiện” mạng xã hội. Mặc dù, mạng xã hội không hoàn toàn là những mặt xấu nhưng có thể là yếu tố gây xao nhãng, khiến sinh viên mất tập trung, lãng phí nhiều thời gian và gây ra kết quả học tập giảm sút.

**Thời gian học tập [2]:** Theo quy định về học chế tín chỉ, sinh viên phải dành nhiều thời gian cho việc tự học, khối lượng kiến thức rất rộng, nếu không dành thời gian tự học thì kết quả nhận lại sẽ không tốt thậm chí là không đạt. Thời gian tự học thường đòi hỏi phải nhiều hơn thời gian lên lớp và được duy trì một cách thường xuyên.

**Tình trạng sức khỏe:** Việc được điều trị các vấn đề về sức khỏe tinh thần trước khi bắt đầu học đại học có liên quan đến nguy cơ bỏ học đại học cao hơn. Sự khác biệt chưa điều chỉnh về tỷ lệ bỏ học là 8,3 điểm phần trăm (13,9% so với 22,2%). Chính vì vậy, trường dữ liệu tình trạng sức khỏe là một trong những yếu tố ảnh hưởng đến kết quả học tập của sinh viên.

**Nguồn kinh phí hỗ trợ học tập [2]:** Sinh viên thường phải làm thêm việc ngoài giờ để trang trải các chi phí sinh hoạt và học phí. Nếu có nguồn tài trợ đủ cho các kinh phí này, họ có thể dành thời gian nhiều hơn cho việc học.

**Học bổng [3]:** Học bổng khuyến khích học tập tạo động lực cho sinh viên cố gắng học tập, thi đua để đạt kết quả tốt.

Ngoài ra còn một số câu hỏi liên quan đến tổ hợp, phương thức xét tuyển, thứ tự nguyện vọng... để phục vụ cho việc khảo sát về điểm số sau này.

#### **Các yếu tố môi trường (10 câu hỏi)**

Các yếu tố môi trường được đánh giá dựa trên thang đo Likert 5 mức độ từ “Rất không đồng ý” đến “Hoàn toàn đồng ý”.

	1	2	3	4	5
Mức độ thích ứng với môi trường học tập					
Phương pháp học tập					
Mức độ hỗ trợ của trường					
Mức độ hỗ trợ của giảng viên					
Điều kiện cơ sở vật chất					

Chất lượng giảng viên					
Mức độ phù hợp của chương trình đào tạo					
Mức độ cạnh tranh trong học tập					
Ảnh hưởng của bạn bè					
Mức độ yêu thích ngành học					

### (C) Kết quả học tập (12 câu hỏi)

Kết quả học tập trước đây thường tạo ra nền tảng kiến thức và kỹ năng cần thiết cho việc học tập hiện tại. Sinh viên thường dựa vào kiến thức đã học trước đó để tiếp tục học tập và nắm bắt thông tin mới. Những kinh nghiệm học tập trước đây giúp họ phát triển kỹ năng tự học, tư duy logic và phương pháp giải quyết vấn đề.

## 2.3. Phương pháp thu thập và xử lý dữ liệu

### 2.3.1. Phương pháp thu thập

Tiến hành khảo sát trực tuyến trên Google Form và được gửi đến tất cả sinh viên các lớp ngành Khoa học tự nhiên đang theo học tại trường.

Đối với các khóa đã ra trường, chúng tôi lựa chọn phỏng vấn trực tiếp hoặc trực tuyến thông qua các nền tảng mạng xã hội. Dữ liệu toàn bộ các học phần và điểm tốt nghiệp được lấy ở đơn vị quản lý đào tạo. Tổng cộng có hơn 1000 câu trả lời đã được gửi nhưng chỉ có 992 câu trả lời hợp lệ được chấp nhận để phân tích sâu hơn. Như vậy, bộ dữ liệu hoàn chỉnh bao gồm 992 phản hồi đã được phân tích bằng IBM SPSS Phiên bản 27.

### 2.3.2. Xử lý dữ liệu

Xử lý dữ liệu hay làm sạch dữ liệu là quá trình sắp xếp và lọc dữ liệu để loại bỏ những dữ liệu không cần thiết, không chính xác. Quy trình này nhằm tìm ra các lỗi sai như: trùng lặp, tính toán sai hoặc thiếu dữ liệu,... để đảm bảo dữ liệu đủ tiêu chuẩn dùng cho học máy.

Nguyên nhân dẫn đến việc dữ liệu bị trùng lặp có thể do sinh viên gửi phản hồi nhiều lần do bị nhập sai thông tin hoặc do lỗi mạng. Để giải quyết vấn đề trên, chúng tôi sử dụng cột mã sinh viên để lọc các dữ liệu giống nhau.

Trong quá trình thu thập dữ liệu sẽ gặp phải vấn đề dữ liệu khuyết ngẫu nhiên (Missing at Random – MAR) đối với điểm học tập và khuyết không ngẫu nhiên (Missing not at Random – MNAR) đối với phiếu khảo sát sinh viên. Để phát hiện các giá trị trống, chúng tôi sử dụng công cụ lọc dữ liệu trong excel bằng cách chọn Data → Filter → Chọn Blank. Để khắc phục tình trạng này, trước hết chúng tôi chấp nhận xóa đi các trường dữ liệu khuyết nhiều gây ảnh hưởng đến mô hình phân tích. Cách làm này đơn giản nhưng có hạn chế là làm giảm kích thước mẫu dữ liệu. Tuy nhiên với một tập dữ liệu lớn sẽ không gây ra ảnh hưởng nhiều vì có đủ quan sát sau khi loại bỏ các giá trị khuyết. Nhưng phương pháp này chỉ sử dụng khi các đối tượng khảo sát không trả lời toàn bộ bảng câu hỏi nghiên cứu. Còn đối với các sinh viên chỉ bỏ trống một vài câu hỏi, chúng tôi thực hiện thay thế bằng các giá trị trung bình hoặc các cách tính phù hợp với từng trường dữ liệu.

Việc sai dữ liệu dạng số có ảnh hưởng rất lớn đến phân tích và báo cáo, vì kết quả phụ thuộc vào đồng nhất của dữ liệu để đưa ra số liệu chính xác. Nguyên nhân dẫn đến tình trạng này là khi sinh viên

làm phiếu khảo sát không đồng nhất các dấu theo như quy định, và tùy theo mỗi loại máy sẽ có định dạng khác nhau. Để khắc phục tình trạng trên, dữ liệu dạng số sẽ được tự động sắp xếp trong cửa sổ dữ liệu lọc của AutoFilter. Sau đó vào “Format Cells → Number”, dữ liệu sẽ tự động chuyển về định dạng số theo như mong muốn.

## 2.4. Phân tích thống kê

### 2.4.1. Thông tin cá nhân

Dữ liệu khảo sát thu được câu trả lời của 992 sinh viên trong đó có 161 nam sinh chiếm tỷ lệ 16,23% và 830 nữ sinh chiếm 83,67%. Nhóm sinh viên được khảo sát thuộc 2 ngành Sư phạm Toán học với 75% và 25% sinh viên ngành Sư phạm Vật Lý. Chúng tôi thực hiện điều tra trên 10 khóa sinh viên từ K14 đến K23 nhằm hiểu rõ sự phát triển trong giáo dục và chất lượng đầu vào của Trường Đại học Thủ Đô Hà Nội.

### 2.4.2. Các yếu tố ảnh hưởng đến kết quả học tập

#### Về yếu tố cá nhân

Đối với nhóm các câu hỏi mang tính cá nhân hóa, đầu tiên, dữ liệu thống kê cho thấy trình độ học vấn phụ huynh của sinh viên với hơn nửa số phụ huynh (59,2%) đã hoàn thành cấp trung học, trong khi có một phần nhỏ (0,4%) chỉ có trình độ học vấn tiểu học. Về công việc làm thêm, một số lượng sinh viên (68,2%) tham gia làm việc ngoài giờ, tuy nhiên, phần lớn sinh viên nhận hỗ trợ học phí từ gia đình (67,8%) và bản thân chỉ chiếm 29,5% do nguồn thu nhập của sinh viên không ổn định. Dữ liệu cũng cho thấy thời gian học tập, với hơn một nửa sinh viên (55,8%) dành từ 3 đến 6 giờ mỗi ngày cho việc học tập. Khoảng thời gian 3-6 tiếng là thời lượng trung bình truy cập mạng xã hội hàng ngày và hầu hết sinh viên (43,6%) sử dụng từ 2 đến 3 nền tảng mạng xã hội bao gồm Facebook, Zalo và Tiktok. Khảo sát cũng phản ánh tình trạng sức khỏe của sinh viên, 79,6% cho biết họ cảm thấy khỏe mạnh, trong khi đó 7% cho biết họ đang gặp phải vấn đề sức khỏe.

Do đối tượng khảo sát tập trung vào sinh viên ngành Khoa học tự nhiên nên phần lớn sinh viên lựa chọn tổ hợp xét tuyển các môn tự nhiên (58,3%). Đối với hình thức xét tuyển, hình thức xét điểm thi Trung học phổ thông Quốc gia được sử dụng nhiều nhất với 520 sinh viên tương đương 52,4%, tiếp đến là hình thức xét học bạ được với 44,1% và chỉ có 35 sinh viên chọn phương thức xét tuyển khác. Cuối cùng, dữ liệu thu thập tỷ lệ sinh viên nhận được học bổng, với khoảng 14,6% được hỗ trợ thông qua học bổng.

Mặc dù dữ liệu thu được 992 câu trả lời khảo sát nhưng chúng tôi thực hiện phân tích chuyên sâu trên 715 nhãn tốt nghiệp.

**Bảng 1: Mối tương quan giữa các yếu tố cá nhân với GPA**

	Trình độ bố mẹ	Công việc làm thêm	Nguồn kinh phí học tập	Thời gian học	Thời gian sử dụng MXH
Pearson Correlation	.670**	.042	.031	.604**	.379**
Sig. (2-tailed)	.000	.259	.414	.000	.000
N	715	715	715	715	715
	Tình trạng sức khỏe	Tổ hợp xét tuyển	Phương thức xét tuyển	Thủ tự nguyện vọng	Học bổng
Pearson Correlation	.024	.019	.254**	-.031	.452**
Sig. (2-tailed)	.527	.620	.000	.406	.000
N	715	715	715	715	715

Về tổng quan, khoảng 1 nửa các yếu tố mang tính cá nhân hóa có mối tương quan rất thấp hoặc thậm chí không có tương quan với kết quả GPA như Công việc làm thêm, Kinh phí học tập, Thứ tự nguyện vọng, Tổ hợp xét tuyển và Tình trạng sức khỏe. Mặc dù có thể có quan điểm cho rằng công việc làm thêm có thể ảnh hưởng đến kết quả học tập, tuy nhiên, việc quản lý thời gian và thay đổi công việc làm thêm một cách linh hoạt mà không làm ảnh hưởng đến thời gian học tập. Rõ ràng 2 yếu tố Thứ tự nguyện vọng và Tổ hợp xét tuyển không quyết định đến GPA do kết quả học tập phụ thuộc chủ yếu vào sự chủ động và nỗ lực của học sinh, không chỉ dựa vào thứ tự nguyện vọng hay lựa chọn môn học.

Những yếu tố có mối tương quan lớn nhất bao gồm Trình độ học vấn của cha mẹ và Thời gian học tập với hệ số tương quan lần lượt là 67% và 60,4%. Dựa vào dữ liệu thống kê, trình độ học vấn cha mẹ của sinh viên tương đối cao, trung bình từ 3 trở lên, trong đó K19 là cao nhất với giá trị trung bình đạt 3,44. Có thể thấy, cha mẹ có trình độ học vấn cao thường cung cấp những góc nhìn sâu sắc và kiến thức phong phú từ kinh nghiệm và học vấn của bản thân, từ đó giúp con cái hiểu biết sâu hơn và phát triển khả năng tư duy. Bên cạnh đó, việc sử dụng phương pháp học hiệu quả và quản lý thời gian hợp lý giúp tối ưu hóa hiệu suất học tập và đạt được kết quả mong muốn. Thống kê số liệu cho thấy, thời gian học tập của 7 khóa trung bình từ 3 đến 3,5, trong đó chỉ có K17 là thấp nhất với giá trị trung bình 3,09.

Ba yếu tố còn lại có mối tương quan thấp hơn bao gồm: Thời gian sử dụng mạng xã hội, Học bổng và Phương thức xét tuyển với hệ số Pearson giao động khoảng 20% đến 40%. Trong đó, 2 khóa K18 và K19 có thời gian sử dụng mạng xã hội ít nhất với giá trị trung bình lần lượt là 2.71 và 2.76. Các số liệu thống kê cũng cho thấy, phần lớn sinh viên sử dụng phương thức xét học bạ Trung học phổ thông và điểm thi Trung học phổ thông. Về học bổng, tỷ lệ sinh viên ngành Sư phạm Toán đạt học bổng cao hơn ngành Sư phạm Vật lý là 0,08.

*Về yếu tố môi trường học tập*

Nhìn chung, các yếu tố có mức đánh giá trung bình khá cao cho thấy sự ảnh hưởng hài lòng của môi trường đại học đối với sinh viên. Dựa trên dữ liệu thống kê, có thể thấy ấn tượng tích cực về chất lượng và mức độ hỗ trợ của giảng viên, với mức đánh giá trung bình lần lượt là 3.27 và 3.37. Điều này cho thấy giảng viên đóng vai trò quan trọng trong quá trình tạo ra một môi trường học tập tích cực, khuyến khích sự sáng tạo và phát triển cá nhân của sinh viên.. Tuy nhiên, sinh viên chưa thực sự hài lòng với mức độ hỗ trợ của trường và cơ sở vật chất tại trường, việc này có thể dẫn đến sự giảm sút của tinh thần học tập và có thể ảnh hưởng tiêu cực đến hiệu suất học tập của họ.

**Bảng 2. Mối tương quan giữa các yếu tố môi trường với GPA**

	Mức độ thích ứng với môi trường	Phương pháp học tập	Mức độ hỗ trợ của trường	Mức độ hỗ trợ của giảng viên	Điều kiện cơ sở vật chất
Pearson Correlation	.322**	.677**	-.112**	-.019	-.055
Sig. (2-tailed)	.000	.000	.003	.605	.142
N	715	715	715	715	715
	Chất lượng giảng viên	Chương trình đào tạo	Mức độ cạnh tranh	Ảnh hưởng của bạn bè	Mức độ yêu thích ngành học
Pearson Correlation	-.092 <sup>*</sup>	-.358**	.560**	.011	.540**
Sig. (2-tailed)	.014	.000	.000	.776	.000
N	715	715	715	715	715

Đối với các yếu tố môi trường, 3 yếu tố không có ý nghĩa thống kê là Mức độ hỗ trợ của giảng viên, Điều kiện cơ sở vật chất và Ảnh hưởng của bạn bè với giá trị sig trên 10%. Nguyên nhân có thể là do dữ liệu có biến động quá lớn hay phân phối không đều giữa các nhóm hoặc các biến xảy ra ngẫu nhiên mà không có mối liên hệ thực sự giữa các biến.

Phương pháp học tập có mức tương quan lớn nhất, hơn 67%, sau đó là Mức độ cạnh tranh và Mức độ yêu thích ngành học với mức tương quan lần lượt là 56% và 54%. Bảng phân tích các yếu tố môi trường dựa trên ngành cho thấy sinh viên ngành Sư phạm Vật Lý có đánh giá tích cực hơn hẳn so với ngành Sư phạm Toán học. Dựa trên thông tin thống kê, phương pháp học tập của 3 khóa K14, K19 và K20 đạt hiệu quả tốt nhất với giá trị trung bình trên 3,5. Dữ liệu cho thấy giá trị trung bình Mức độ yêu thích ngành học chênh lệch không đáng chỉ từ khoảng 3 đến 3,4. Như vậy, Phương pháp học tập hiệu quả và Mức độ yêu thích đều đóng vai trò quan trọng bởi chúng ảnh hưởng đến cách sinh viên tiếp cận và tiếp nhận kiến thức, cũng như sự tập trung và động viên của họ trong quá trình học. Đặc biệt, Mức độ cạnh tranh có số liệu thống kê cao hơn hẳn so với các yếu tố còn lại với giá trị trung bình phần lớn trên 3,6. Sự cạnh tranh có ảnh hưởng lớn đến kết quả học tập, vì nó có thể tạo ra áp lực và thúc đẩy sinh viên nỗ lực hơn. Ngoài ra, Mức độ thích ứng với môi trường cũng thể hiện mối tương quan nhưng chỉ ở mức 32,2%.

Trong các yếu tố môi trường, 2 yếu tố bao gồm: Chất lượng giảng viên và Sự phù hợp của chương trình đào tạo có hệ số tương quan âm. Điều này có thể do sự không phù hợp giữa phương pháp giảng dạy của giảng viên và cách học của sinh viên, sự thiếu sự hài lòng về chất lượng của bài giảng, hoặc thiếu sự tương tác tích cực giữa giảng viên và sinh viên. Bên cạnh đó, Chương trình đào tạo không đáp ứng được nhu cầu thực tế của sinh viên hoặc không cung cấp đủ kiến thức và kỹ năng cần thiết cho họ trong thực tế công việc, điều này có thể phản ánh qua sự thiếu hụt kiến thức thực tiễn, thiếu sự linh hoạt trong phương pháp giảng dạy, hoặc không đáp ứng được các tiêu chuẩn ngành nghề hoặc yêu cầu của thị trường lao động. Theo số liệu thống kê cho thấy, 4 khóa đầu đánh giá Chương trình đào tạo rất phù hợp với giá trị trung bình trên 4,0. Tuy nhiên, 3 khóa sau đó cảm thấy chưa thực sự hài lòng về Chương trình đào tạo, đặc biệt giá trị trung bình K20 đánh giá chỉ còn 2,95. Như vậy, có thể thấy Chương trình đào tạo đã không còn phù hợp với ngành nghề đào tạo của sinh viên, yêu cầu sự thay đổi và cập nhật sao cho phù hợp với thị trường lao động trong hiện tại và tương lai.

### 2.4.3. Kết quả học tập

*Kết quả học tập trước đây*

**Bảng 2. Mối tương quan giữa kết quả học tập trước đây với GPA**

	Kết quả thi tốt nghiệp THCS	Môn Toán THPT	Môn Văn THPT	Môn Anh THPT	Môn Sử THPT	Môn Địa THPT
Pearson Correlation	.005	.582**	.539**	.472**	-.006	.149
Sig. (2-tailed)	.902	.000	.000	.000	.916	.009
N	715	715	715	715	306	306
	Môn GDCD THPT	Môn Lý THPT	Môn Hóa THPT	Môn Sinh THPT	Điểm thi tốt nghiệp THPT	Tiếng Anh đầu vào
Pearson Correlation	.051	.450**	.108	.046	.618**	-.026
Sig. (2-tailed)	.377	.000	.030	.350	.000	.482
N	306	409	409	409	715	715



	Hình học sơ cấp (2)	Khóa luận tốt nghiệp (8)	Đạy học phân hóa phần đại số (3)	Đạy học phân hóa phần hình học (3)	Một số chuyên đề nâng cao về dãy số và hàm số (2)
Pearson Correlation	.607**	.610**	.550**	.614**	.503**
Sig. (2-tailed)	.000	.000	.000	.000	.000

**Bảng 4. Mối tương quan giữa các học phần chuyên ngành của SP Vật Lý với GPA**

	Cơ đại cương (3)	Đại cương về khoa học trái đất (2)	Toán cao cấp (3)	Điện và từ (3)	Quang học (3)	Dao động và sóng (2)	Vật lý phân tử nhiệt học (3)
Pearson Correlation	.533**	.626**	.465**	.578**	.542**	.422**	.221**
Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.004

	Toán cho Vật Lý (3)	Vật lý nguyên tử và hạt nhân (2)	Kỹ thuật điện (3)	Lý luận dạy học Vật Lý (3)	Thực hành vật lý đại cương (2)	Tin học cho Vật Lý (2)
Pearson Correlation	.484**	.504**	.482**	.600**	.555**	.436**
Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

	Thí nghiệm Vật Lý phổ thông (3)	Thiết kế và tổ chức hoạt động trọng dạy học Vật Lý (4)	HDDH Vật Lý bằng TA (2)	HD tổ chức HĐTN (2)	Vật Lý chất rắn (2)
Pearson Correlation	.624**	.658**	.499**	.681**	.382**
Sig. (2-tailed)	.000	.000	.000	.000	.000

	Phương pháp giải bài tập Vật Lý phổ thông (3)	Cơ học lượng tử (2)	Kỹ thuật điện tử (3)	Kiểm tra đánh giá trong dạy học Vật Lý (2)	Lịch sử Vật Lý (2)	Phát triển chương trình Vật Lý phổ thông (2)
Pearson Correlation	.601**	.596**	.602**	.581**	.662**	.492**
Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

	Phương pháp nghiên cứu khoa học (2)	Khóa luận tốt nghiệp (8)	Các phương pháp dạy học hiện đại trong Vật Lý (3)	Hướng dẫn tổ chức dạy học tích hợp theo chủ đề môn KHTN (3)	Vật Lý lý thuyết (2)
Pearson Correlation	.529**	.596**	.608**	.514**	.556**
Sig. (2-tailed)	.000	.000	.000	.000	.000



**Bảng 7. Mối tương quan giữa các môn tự chọn ngành Sư phạm Vật Lý với GPA**

	Tự chọn 1 (Tin học, PTNL thông tin trong kĩ nguyên số)	Tự chọn 2 (3) (Hà nội học, Biển và hải đảo,...)	Tự chọn 3 (5) (Tiếng Anh, Hàn, Trung)	Tự chọn 4 (2) (Âm nhạc và cảm thụ âm nhạc, Mỹ thuật,...)	Tự chọn 5 (2) (Tiếng Anh/ Hàn/ Trung chuyên ngành)	Tự chọn 6 (2) (Điện động lực học, Vật lý mô phỏng,...)	Tự chọn 7 (2) (Khoa học vật liệu, Vật lý thống kê,...)	Tự chọn 8 (2) (Lý thuyết tương đối, lý thuyết chất rắn,...)
Pearson Correlation	.337**	.508**	.271**	.432**	.318**	.557**	.098	.668**
Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.211	.000

Sự khác biệt trong thống kê thấy rõ ràng ở bảng 2.7 và 2.8 với các học phần tự chọn. Nguyên nhân xuất phát từ nội dung từ một số học phần này có liên quan đến môn chuyên ngành, trong khi đó, một số khác lại là những môn liên quan đến các lĩnh vực khác như: âm nhạc, nghệ thuật, kinh tế, phân tích dữ liệu, lịch sử,...

## 2.5. Một số khuyến nghị về tổ chức đào tạo nhằm cải thiện chất lượng sinh viên

### 2.5.1. Đối với nhà trường

Do tính chất môi trường đại học có sự biệt so với môi trường phổ thông, trong khi các cấp học trước đây tập trung vào việc cung cấp kiến thức cơ bản và toàn diện cho học sinh, đại học lại tập trung vào sự chuyên sâu và chuyên nghiệp trong đào tạo. Phương pháp học tập ở đại học cũng yêu cầu tính chủ động, tự chủ của sinh viên. Chính vì vậy, việc thích ứng với tiến độ và trình độ đại học có thể là thách thức đối với sinh viên. Và để sinh viên dễ dàng có thể thích ứng và tận hưởng môi trường học tập, nhà trường nên tạo ra một không gian mở và hỗ trợ giúp sinh viên cảm thấy thoải mái học tập và nghiên cứu tại trường. Đảm bảo rằng sinh viên có thể truy cập vào các nguồn lực học thuật như thư viện, phòng lab và các dịch vụ tư vấn học vụ cũng như sẵn sàng hỗ trợ sinh viên trong việc phát triển kỹ năng học tập và nghiệp vụ thông qua các khóa học và hội thảo.

Không chỉ vậy, niềm đam mê và yêu thích ngành học cũng là một trong những yếu tố quan trọng giúp sinh viên cố gắng và nỗ lực học tập. Do đó, nhà trường cũng cần tạo ra một môi trường học tập đầy sôi động và hấp dẫn. Điều này bao gồm việc xây dựng một chương trình học phong phú và linh hoạt, cho phép sinh viên chủ động lựa chọn những môn học phù hợp với sở thích và mục tiêu cá nhân của mình. Sự kết hợp giữa lý thuyết và thực hành trong quá trình giảng dạy cũng rất quan trọng, giúp sinh viên áp dụng những kiến thức học được vào thực tế và hiểu sâu hơn về ngành học của mình.

Bên cạnh đó, tạo ra các cơ hội cho sinh viên tham gia vào các dự án nghiên cứu và thực hành, cũng như hỗ trợ họ trong việc phát triển kỹ năng mềm và nghiệp vụ. Đồng thời, tạo điều kiện cho sinh viên giao lưu và hỗ trợ lẫn nhau trong quá trình học tập cũng là một yếu tố quan trọng. Một cộng đồng học thuật sôi động và hỗ trợ giúp sinh viên cảm thấy được đánh giá cao và động viên, tạo ra một môi trường học tập tích cực và thú vị. Việc tạo điều kiện cho sinh viên tìm hiểu chuyên sâu và khám phá sự đam mê trong ngành học của mình cũng được coi là chìa khóa để giúp sinh viên yêu thích và hứng thú hơn với việc học.

Ngoài ra, một môi trường học cạnh tranh tích cực là một yếu tố quan trọng kích thích hứng thú học tập và nghiên cứu cho sinh viên. Để có thể tạo ra và duy trì một môi trường học cạnh tranh tích cực ấy, nhà trường có thể xây dựng một cộng đồng học thuật sôi nổi, nơi mà sinh viên được khích lệ và động viên để tham gia vào các hoạt động ngoại khóa và học thuật. Không chỉ thế, sự cạnh tranh lành mạnh cũng được thể hiện thông qua các cuộc thi học thuật và các dự án nghiên cứu, nơi sinh viên có cơ hội thể hiện và phát triển tài năng của mình. Đồng thời, trường cũng cần đảm bảo rằng sinh viên có truy cập

vào các nguồn lực và dịch vụ hỗ trợ, từ thư viện đến tư vấn học vụ, để họ có thể vượt qua mọi thách thức trong hành trình học tập của mình. Tạo điều kiện cho phản hồi và cải thiện là một phần không thể thiếu trong quá trình nâng cao chất lượng môi trường giáo dục. Việc lắng nghe ý kiến của sinh viên và thực hiện điều chỉnh phù hợp không chỉ giúp nâng cao chất lượng giảng dạy mà còn tạo ra một môi trường học tập linh hoạt và đáng tin cậy. Không chỉ vậy, một môi trường học tập linh hoạt và đa dạng cũng là nơi mà sinh viên được tiếp cận những kiến thức và kỹ năng mới nhất và khám phá phát triển bản thân một cách hiệu quả.

Để đáp ứng nhu cầu không ngừng của xã hội hiện đại, nhà trường cũng nên áp dụng các biện pháp cải tiến mạnh mẽ trong chương trình đào tạo. Chúng tôi đề xuất cập nhật chương trình đào tạo tích hợp các kỹ năng mới và xu hướng công nghệ. Việc này có thể bao gồm việc thiết kế các học phần mới về công nghệ giáo dục, đổi mới phương pháp giảng dạy và đánh giá, cũng như đảm bảo rằng giáo viên được đào tạo để sử dụng công nghệ mới trong quá trình giảng dạy. Sự thay đổi này sẽ giúp sinh viên phát triển những kỹ năng nghiệp vụ như tư duy sáng tạo, giao tiếp hiệu quả và làm việc nhóm, tất cả đều là những yếu tố quan trọng trong việc thành công trong môi trường giáo dục ngày nay.

### **2.5.2. Đối với đội ngũ giảng viên**

Để nâng cao chất lượng học tập của sinh viên, giảng viên đóng vai trò quan trọng như là người hướng dẫn và truyền cảm hứng. Với môi trường học tập năng động, sáng tạo như hiện nay, giảng viên có thể thúc đẩy sự phát triển toàn diện của họ. Bằng cách sử dụng các phương pháp giảng dạy đa dạng và thú vị, giảng viên có thể kích thích sự tò mò và ham học của sinh viên, giúp họ tiếp cận vào những kiến thức và kỹ năng mới nhất. Việc cung cấp phản hồi xây dựng và khuyến khích sự tự học cũng là một phần quan trọng trong quá trình giảng dạy. Bằng cách thúc đẩy sự phát triển cá nhân và tạo ra các cơ hội cho sinh viên tham gia vào các hoạt động ngoại khóa và dự án nghiên cứu, giảng viên không chỉ giúp sinh viên thành công trong học tập mà còn chuẩn bị cho họ một nền tảng vững chắc cho sự nghiệp và cuộc sống sau này.

### **2.5.3. Đối với cá nhân sinh viên**

Bên cạnh các yếu tố môi trường và người dạy, nhân tố ở bản thân người học là chìa khóa quyết định trực tiếp đến kết quả học tập và một trong những yếu tố cần thiết là phương pháp học tập của sinh viên. Để xây dựng phương pháp học tập hiệu quả, việc tạo ra một kế hoạch học tập tổ chức và có mục tiêu là quan trọng. Bắt đầu bằng việc đặt ra các mục tiêu học tập cụ thể và đo lường được. Điều này sẽ giúp sinh viên biết bản thân đang hướng đến điều gì và tập trung vào những việc quan trọng nhất. Sau đó, lập một kế hoạch học tập chi tiết, bao gồm việc xác định thời gian học và các nhiệm vụ cụ thể mà bạn muốn hoàn thành. Một phần không kém phần quan trọng trong phương pháp học là kỹ thuật học. Thay vì chỉ sử dụng các kỹ thuật truyền thống, sinh viên nên tìm hiểu, thử nghiệm và tìm ra phương pháp học tập cá nhân thực sự phù hợp, như ghi chú, tóm tắt, hay sử dụng các biểu đồ và sơ đồ tư duy. Sử dụng các công cụ kỹ thuật số như ứng dụng ghi chú hoặc trình quản lý công việc để tổ chức thông tin một cách dễ dàng và tiện lợi. Môi trường học tập cũng đóng vai trò quan trọng trong hoạt động học tập của sinh viên. Một không gian yên tĩnh và thoải mái để tập trung vào việc học, tránh xa các yếu tố gây xao lạc như điện thoại hoặc tiếng ồn từ bên ngoài. Đồng thời, hãy khuyến khích bản thân bằng cách tạo ra các phương tiện giúp nâng cao tinh thần và động lực học tập. Bên cạnh đó, sinh viên nên tận dụng nguồn lực và hỗ trợ từ giáo viên, bạn bè và nguồn tài liệu học tập. Hỏi thăm và thảo luận với người khác khi gặp khó khăn, và đừng ngần ngại hỏi giáo viên về bất kỳ điều gì khi cần. Sự hỗ trợ này sẽ giúp bạn tiến xa hơn trên con đường học tập và phát triển cá nhân.

Ngoài một phương pháp học tập, sinh viên cũng cần khả năng quản lý thời gian hiệu quả và hợp lý. Để xây dựng một thời gian biểu phù hợp, sinh viên cần thiết lập một kế hoạch rõ ràng và nhất quán.

Bằng cách đặt ra các mục tiêu cụ thể và ưu tiên công việc theo thứ tự quan trọng, họ có thể tập trung vào những việc quan trọng nhất đầu tiên. OneNote là công cụ hữu ích để ghi chú và tổ chức các hoạt động hàng ngày, giúp họ có cái nhìn tổng quan về thời gian và không bị lạc hướng. Không chỉ vậy, sinh viên cũng cần phân chia thời gian một cách cân đối, dành thời gian cho học tập, công việc, thể dục và nghỉ ngơi. Việc duy trì sự cân bằng này không chỉ giúp sinh viên tối đa hóa hiệu suất làm việc mà còn giữ cho tinh thần luôn tỉnh táo và minh mẫn. Không chỉ là việc lập kế hoạch, việc thực hiện kế hoạch một cách nhất quán cũng rất quan trọng. Bằng cách tuân thủ kế hoạch đã đề ra, sinh viên có thể tránh được tình trạng rơi vào thói quen trì hoãn và lãng phí thời gian không cần thiết. Cuối cùng, việc đánh giá và điều chỉnh kế hoạch là một phần không thể thiếu của quản lý thời gian. Thường xuyên đánh giá tiến trình và điều chỉnh kế hoạch dựa trên những thay đổi và mục tiêu mới sẽ giúp sinh viên duy trì sự linh hoạt và hiệu quả trong công việc.

Nhìn chung, việc xây dựng một phương pháp học tập hiệu quả, phù hợp kết hợp với một thời gian biểu hợp lý và chi tiết, ưu tiên công việc, phân chia thời gian cân đối, thực hiện kế hoạch nhất quán và đánh giá định kỳ giúp sinh viên có thể cải thiện và nâng cao kết quả của hoạt động học tập đáng kể.

### 3. KẾT LUẬN

Xây dựng bộ dữ liệu chung về hồ sơ học tập của học sinh, trong đó có lưu trữ các dữ liệu liên quan đến các yếu tố ảnh hưởng tới kết quả học tập của học sinh và điểm các học phần tại đại học. Bằng cách áp dụng phương pháp phân tích thống kê sơ bộ, chúng tôi nhận thấy rằng thành tích học tập của sinh viên chịu ảnh hưởng từ nhiều yếu tố, trong đó phương pháp học tập có ảnh hưởng mạnh nhất, mức độ hỗ trợ của trường và mức độ phù hợp của chương trình đào tạo được đánh giá có mối tương quan nghịch biến. Dựa vào dữ liệu và kết quả thống kê, bài báo đưa ra một số khuyến nghị nhằm hỗ trợ quản lý và nâng cao chất lượng sinh viên. Bộ dữ liệu trên không chỉ dừng lại ở việc thu thập và xử lý dữ liệu mà còn tiến xa hơn bằng cách tích hợp các kỹ thuật tiên tiến như học máy và học sâu. Mục tiêu chính là phát triển các mô hình có khả năng phân tích, dự đoán và cung cấp cái nhìn sâu sắc về mối quan hệ giữa các yếu tố ảnh hưởng và kết quả học tập của học sinh. Từ đó, có định hướng và can thiệp kịp thời giúp cải thiện kết quả học tập của sinh viên.

### TÀI LIỆU THAM KHẢO

1. Chính phủ (2020), *Phê duyệt Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030*, Quyết định số 749/QĐ-TTg ngày 03/06/2020 của Thủ tướng Chính phủ <https://datafiles.chinhphu.vn/cpp/files/vbpq/2020/06/749.signed.pdf>.
2. N. T. T. An, N. T. N. Thứ, Đ. T. K. Oanh và N. V. Thành (2016). Những nhân tố ảnh hưởng kết quả học tập của sinh viên năm I-II Trường Đại học Kỹ thuật – Công nghệ Cần Thơ. *Tạp chí Khoa học - Trường Đại học Cần Thơ*, Phần C: Khoa học Xã hội, Nhân văn và Giáo dục (46), 82–89.
3. V. V. Kiệt, Đ. T. T. Phương (2017). Nghiên cứu các nhân tố cơ bản ảnh hưởng đến kết quả học tập của sinh viên. *Tạp chí Khoa học – ĐHQG Hà Nội: Nghiên cứu giáo dục*, Tập 3, số 3, tr7.
4. M. S. Farooq, A. H. Chaudhry, M. Shafiq, and G. Berhanu (2011). Factors affecting students' quality of academic performance: A case of secondary schoollevel. *Journal of Quality and Technology Management*, 7, 1–14.
5. N. O. Omarova, A. A. Echilova (2024). Big Data Technologies in the Education System. *Computational and Strategic Business Modelling* (pp.567-577), DOI:10.1007/978-3-031-41371-1\_47.

6. T. Hartati, N. Fitria, M. A. A. Harahap, D. Dasari (2023). Data-Driven Education: Data Processing as a Key to Improving the Quality of Mathematics Education. *ALSYSTECH Journal of Education Technology*, 2(1), DOI:10.58578/alsystech.v2i1.2361.

### **A DATASET OF HANOI METROPOLITAN UNIVERSITY STUDENTS INCLUDES FACTORS AFFECTING ACADEMIC PERFORMANCE AND LEARNING PROCESS**

**Abstract:** *This article presents the results of a study on the construction of a data base in the field of education sciences concerning the learning activities of natural sciences students at the Faculty of Pedagogy at Hanoi Metropolitan University. The data set not only focuses on information collection, but also high lights factors that affect students' learning processes to guide educational activities, analyze support for decision-making and improve learning effectiveness. Data collection uses two main methods: direct collect of persons through questionnaires and indirect collect through educational management units. After data cleaning and preprocessing, the dataset contains 992 samples and 89 data fields from 10 cohorts (K14 to K23). These data are divided into three main groups: demographic information, factors affecting learning results, and learning results (including general education performance and university module results). This data set, in particular, was processed and cleaned with preliminary statistical analysis tools. We then make some recommendations for universities, faculty and students to improve the quality of education. This is a solid foundation for the study of educational sciences, applying machine learning and deep learning methods to predict learners' academic outcomes.*

**Keywords:** *Machine Learning, Academic Outcomes, Data Science, Education Science, Statistical Analysis, Dataset, Influencing Factors.*