

KHAİ THÁC DỮ LIỆU MỞ KHI DẠY XÁC SUẤT THỐNG KÊ CHO SINH VIÊN, NGHIÊN CỨU TẠI TRƯỜNG ĐẠI HỌC HẢI DƯƠNG

Nguyễn Thị Thu Hà
Trường Đại học Hải Dương
Email: uhdhangyenthu76.edu@gmail.com

Tóm tắt: Trong bối cảnh cuộc Cách mạng Công nghiệp 4.0 và sự phát triển của kinh tế số, năng lực làm việc với dữ liệu đã trở thành một yêu cầu thiết yếu đối với nguồn nhân lực chất lượng cao. Đối với học phần Xác suất Thống kê, dữ liệu mở đóng vai trò là cầu nối thiết yếu giữa lý thuyết và thực hành. Thay vì sử dụng các bộ dữ liệu giả định, được đơn giản hóa trong giáo trình, dữ liệu mở cung cấp cho sinh viên các bộ dữ liệu xác thực từ thế giới thực. Tuy nhiên, việc giảng dạy học phần Xác suất Thống kê tại các trường đại học hiện nay vẫn còn nặng về lý thuyết, thiếu kết nối với các vấn đề thực tiễn, dẫn đến việc sinh viên gặp khó khăn trong việc ứng dụng kiến thức. Bài viết trình bày kết quả nghiên cứu việc ứng dụng dữ liệu mở và đề xuất một giải pháp tích hợp có hệ thống nguồn tài nguyên dữ liệu mở vào giảng dạy học phần Xác suất thống kê giúp sinh viên có cơ hội tiếp cận tình huống thực tế, rèn luyện kỹ năng phân tích và tư duy phản biện góp phần nâng cao hứng thú học tập, phát triển năng lực ứng dụng kiến thức thống kê vào thực tiễn.

Từ khóa: dữ liệu mở, Xác suất Thống kê, Trường Đại học Hải Dương

EXPLOITING OPEN DATA IN TEACHING PROBABILITY AND STATISTICS TO UNIVERSITY STUDENTS: A CASE STUDY AT HAI DUONG UNIVERSITY

Nguyen Thi Thu Ha
Hai Duong University
Email: uhdhangyenthu76.edu@gmail.com

Abstract: In the context of the Fourth Industrial Revolution and the growth of the digital economy, data competency has become an essential requirement for high-quality human resources. For the Probability and Statistics course, open data serves as a vital bridge between theory and practice. Instead of using hypothetical, simplified datasets from textbooks, open data provides students with authentic, real-world datasets. However, the current teaching of the Probability and Statistics course at universities remains heavily theoretical and disconnected from practical problems, leading to difficulties for students in applying their knowledge. This paper presents the results of research into the application of open data and proposes a solution for systematically integrating open data resources into the teaching of the Probability and Statistics course. This approach helps students encounter realistic situations, practice analytical skills and critical thinking, and thereby enhances their interest in learning and develops their ability to apply statistical knowledge in practice.

Keywords: open data, Probability and Statistics, Hai Duong University

Nhận bài: 12/06/2025

Phản biện: 12/07/2025

Duyệt đăng: 17/07/2025

I. ĐẶT VẤN ĐỀ

Trong bối cảnh chuyển đổi số và phát triển mạnh mẽ của khoa học dữ liệu (DL), giáo dục đại học hiện nay không chỉ đòi hỏi truyền đạt kiến thức mà còn cần trang bị cho sinh viên (SV) năng lực phân tích DL thực tiễn, kỹ năng phản biện DL, và khả năng ra quyết định dựa trên số liệu. Đặc biệt với học phần Xác suất Thống kê (XSTK), thường đánh giá là trừu tượng và khô khan, việc khai thác DL mở (open data) được xem như một hướng tiếp cận hiệu quả để kết nối lý thuyết với thực tiễn. Gould (2010) đã chỉ ra rằng việc sử dụng DL thực tế, phức tạp và có bối cảnh rõ ràng sẽ giúp SV tham gia vào quá trình học một cách tích cực hơn, thay vì chỉ tiếp nhận các công thức một cách thụ động.

Sự phát triển của các nguồn DL mở đã tạo ra một cơ hội lớn để đổi mới phương pháp giảng dạy. Theo Hiệp hội Thống kê Hoa Kỳ (American Statistical Association, 2014), việc để sinh viên

làm việc trực tiếp với DL là một phần cốt lõi của việc đào tạo thống kê trong thế kỷ 21, giúp họ phát triển tư duy phản biện và kỹ năng giải quyết vấn đề. Roberts et al. (2019) đã trình bày mô hình sử dụng DL mở trong các khóa học nhập môn thống kê, nhấn mạnh rằng DL mở giúp SV thu hút hơn và dễ tiếp cận kiến thức định lượng. Bên cạnh đó, Hardin (2018) cũng đã chỉ ra rằng việc cho SV tự tải và xử lý DL động từ Internet, thay vì chỉ học qua các bộ DL cố định giúp họ hình thành tâm thế ứng dụng DL thực, chuẩn bị tốt hơn cho công việc phân tích trong thực tế. Xu hướng tiếp tục được củng cố qua dự án “Open Case Studies” (Wright et al., 2023), giới thiệu các case study mở, mang tính phân tích chuyên sâu, giúp SV thực hành tư duy thống kê với DL thật, nhiều chiều và thách thức. Tại Việt Nam, việc tích hợp học liệu mở (OER) cũng đã được các cơ sở giáo dục chú trọng. Một nghiên cứu của Học viện

Chính trị quốc gia Hồ Chí Minh (2024) cho thấy khai thác học liệu mở giúp giảng viên (GV) chủ động hơn trong thiết kế bài giảng, cải thiện chất lượng đào tạo mà vẫn giảm chi phí biên soạn tài liệu. Đồng thời, Bộ Giáo dục cũng đang triển khai kho DL mở giáo trình, hướng đến hơn 300 tài nguyên trong giai đoạn 2024 - 2026, tạo điều kiện thuận lợi cho việc tra cứu, áp dụng giảng dạy.

Tại Trường Đại học Hải Dương hiện nay, giảng dạy XSTK ở trường thiên về lý thuyết và ví dụ mẫu; các tình huống thực tế với số liệu mở còn rất hạn chế. Do đó, thực hiện nghiên cứu khai thác DL mở trong giảng dạy XSTK tại Trường Đại học Hải Dương là cần thiết để tăng tính kết nối giữa kiến thức và thực tế, tăng hứng thú và động lực học tập của SV, rèn luyện kỹ năng xử lý, phân tích DL thực cho SV.

II. NỘI DUNG NGHIÊN CỨU

2.1. Khái niệm dữ liệu mở

Dữ liệu mở được định nghĩa theo nhiều quan điểm khác nhau, theo Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 đã đưa ra định nghĩa DL mở của cơ quan nhà nước: là dữ liệu được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ”. Theo Điều 3 Luật Dữ liệu 2024 quy định: DL mở là dữ liệu mà mọi cơ quan, tổ chức, cá nhân có nhu cầu đều được tiếp cận, chia sẻ, khai thác, sử dụng. Như vậy DL mở là dữ liệu mà bất kỳ ai cũng có thể truy cập, sử dụng và chia sẻ. Dù định nghĩa theo cách nào thì DL mở luôn đảm bảo được 2 đặc trưng: khả năng sẵn sàng và khả năng tiếp cận dễ dàng. Các đặc điểm nổi bật bao gồm tính sẵn có, khả năng truy cập dễ dàng (thường miễn phí và qua mạng), định dạng máy đọc được và cho phép tái sử dụng.

2.2. Vai trò của khai thác dữ liệu mở khi giảng dạy xác suất thống kê

Việc khai thác dữ liệu mở trong giảng dạy môn XSTK đóng vai trò quan trọng trong đổi mới giáo dục đại học thời kỳ số hóa, giúp môn học trở nên thực tế, hấp dẫn và gắn với nhu cầu nghề nghiệp.

Thứ nhất, DL mở giúp xây dựng cầu nối giữa lý thuyết và thực tiễn. Thay vì dựa vào dữ liệu giả định, SV được tiếp cận với dữ liệu thật từ các cơ quan như Tổng cục Thống kê hay WHO, giúp hiểu sâu hơn các khái niệm như phân phối chuẩn thông qua các ví dụ thực tế như chiều cao, cân nặng của dân số.

Thứ hai, DL mở giúp làm phong phú trải nghiệm học tập theo định hướng năng lực. Qua

việc làm việc với DL thực tế, SV được rèn luyện kỹ năng xử lý dữ liệu lộn xộn, phân tích và trình bày dữ liệu – những năng lực cốt lõi trong chương trình giáo dục đại học hiện nay. Ví dụ, SV có thể phân tích mối liên hệ giữa tỷ lệ tốt nghiệp THPT và thất nghiệp tại các tỉnh, buộc họ phải vận dụng nhiều kỹ năng phân tích và kiểm định giả thuyết.

Thứ ba, DL mở tạo điều kiện triển khai các phương pháp học tập tích cực như học theo dự án, học giải quyết vấn đề. SV có thể chủ động lựa chọn đề tài, khai thác dữ liệu và trình bày kết quả phân tích, giúp phát triển tư duy nghiên cứu và năng lực lập luận. Ví dụ, một nhóm SV đã phân tích mối quan hệ giữa ô nhiễm không khí và tỷ lệ bệnh hô hấp tại Hà Nội giai đoạn 2015–2020, dựa trên DL từ Bộ Y tế và Bộ TN&MT.

Thứ tư, DL mở hỗ trợ giảng viên đổi mới phương pháp giảng dạy bằng việc xây dựng các tình huống gắn với thực tiễn như biến động giá xăng dầu hay chỉ số lạm phát, thay vì các ví dụ giả định khô khan.

Cuối cùng, DL mở giúp gắn kết dạy học XSTK với yêu cầu nghề nghiệp, chuẩn bị cho SV năng lực làm việc với DL số trong thời đại công nghiệp 4.0. Nhờ đó, SV hình thành tư duy logic, định lượng, và kỹ năng giải thích dữ liệu – những phẩm chất thiết yếu trong thị trường lao động hiện đại.

2.3. Thực trạng khai thác dữ liệu mở khi dạy xác suất thống kê tại trường Đại học Hải Dương

2.3.1. Khái quát chung về khảo sát

Mẫu khảo sát: Tiến hành khảo sát 08 GV tham gia giảng dạy học phần XSTK và 120 SV tại Trường Đại học Hải Dương về khai thác DL mở khi giảng dạy và học tập học phần XSTK.

Phương pháp khảo sát: tiến hành khảo sát bằng bảng hỏi với 20 câu hỏi mô tả mức độ đạt được theo thang Likert 5 mức độ (mức độ từ 1-5 tương đương với rất yếu - yếu – trung bình – khá – tốt) của 08 GV tham gia giảng dạy học phần XSTK và 120 SV vào đầu năm học 2024-2025.

2.3.2. Kết quả khảo sát

Kết quả khảo sát cho thấy GV có nhận thức khá tốt về vai trò của DL mở trong dạy học, với điểm trung bình đạt 4,0/5, đa số GV cho rằng cần tích hợp DL mở vào bài giảng (4,5/5). Điều này phản ánh tinh thần đổi mới và sẵn sàng tiếp cận các phương pháp dạy học hiện đại, theo hướng ứng dụng DL thực tiễn. Tuy nhiên, sự nhận thức này mới dừng lại ở mức tư duy và mong muốn, chưa thực sự được chuyển hóa thành hành động cụ thể và có hệ thống. Mặc dù nhận thức

tốt, nhưng năng lực triển khai của GV lại chưa hiệu quả, điểm trung bình chỉ đạt (3,2/5). Khoảng 60% cho biết có thể tìm kiếm DL mở (3,0/5), nhưng tỷ lệ tự tin sử dụng phần mềm xử lý (như Excel nâng cao, Python, R) vào giảng dạy chưa thực sự cao (3,1/5). Điều này chỉ ra rằng một bộ phận GV chưa được đào tạo bài bản về kỹ năng xử lý DL thực hoặc còn hạn chế về công nghệ. Việc từng sử dụng DL mở trong giảng dạy cũng chỉ đạt

(3,1), chủ yếu ở dạng ví dụ minh họa chứ chưa đi sâu vào bài tập, dự án hay kiểm tra đánh giá có tính phân tích DL. Một thực trạng rõ ràng là thiếu hệ thống hỗ trợ giảng dạy với dữ liệu mở. Dù đa số GV đồng ý rằng trường cần xây dựng kho DL mở nội sinh, điểm trung bình ở tiêu chí có tài liệu, hướng dẫn về dạy học với dữ liệu mở cho SV chỉ 2,8/5, phản ánh sự thiếu hụt cả về công cụ, tài liệu và chính sách hỗ trợ triển khai.

Bảng. Kết quả khảo sát về khai thác DL mở của SV khi học XSTK

STT	Nội dung	Điểm TB
1	SV nhận thức được DL mở giúp hiểu bài sâu hơn	4,2
2	Mong muốn được học với tình huống dữ liệu mở	4,5
3	Có thể tìm và tải dữ liệu từ các cổng mở	2,9
4	Từng làm bài tập với DL thực trong môn học	3,0
5	Được giảng viên hướng dẫn khai thác DL mở	3.1
6	Có thiết bị, phần mềm, internet để học và phân tích DL mở	3.2

Kết quả khảo sát cho thấy, SV có nhận thức ban đầu tương đối tốt về vai trò của khai thác DL mở giúp hiểu bài sâu hơn (điểm trung bình 4,2/5). Tuy nhiên, điều tích cực là SV rất quan tâm đến ứng dụng thực tiễn: 80% mong muốn được học thông qua tình huống sử dụng dữ liệu thực (4,5/5). Điều này cho thấy nếu được tổ chức học tập hợp lý, SV sẵn sàng tiếp nhận cách học mới, thực tiễn hơn. Một điểm yếu lớn nhất ở SV với điểm trung bình chỉ 2,9/5. SV thiếu cả kỹ năng tìm kiếm, xử lý DL, lẫn sử dụng phần mềm phân tích. Điều này đồng nghĩa với việc họ thiếu nền tảng kỹ thuật cần thiết để tiếp cận học tập dựa trên DL mở. Hệ quả là các bài tập, dự án có gắn với DL thực chưa được SV thực hiện nhiều, hoặc nếu có thì mang tính hình thức, mô phỏng đơn giản. SV đánh giá mức hỗ trợ ở mức 3,2/5, phản ánh rằng họ thiếu môi trường học tích cực, thiết bị, phần mềm chuyên dụng, cũng như hướng dẫn cụ thể từ GV. SV cho rằng có đủ điều kiện cơ bản về thiết bị và internet, nhưng vẫn thiếu điều kiện thực hành nhóm, chia sẻ dữ liệu hoặc truy cập nguồn học liệu mở.

Nguyên nhân:

Đối với GV: Thiếu kỹ năng về khai thác và xử lý DL mở, nhiều GV chưa được đào tạo bài bản về các công cụ và phần mềm phân tích DL, dẫn đến tâm lý e ngại khi triển khai các bài học gắn với dữ liệu thực. Nguyên nhân nữa là, chưa có kinh nghiệm tích hợp DL mở vào thiết kế bài giảng:

Phần lớn GV vẫn tiếp cận việc dạy học XSTK theo hướng truyền thống (lý thuyết – bài tập mẫu), ít được tiếp cận với các phương pháp dạy học theo dự án, tình huống thực tiễn có sử dụng DL mở. Khối lượng công việc lớn, thiếu thời gian chuẩn bị DL thực tế: Việc tìm kiếm, làm sạch và xây dựng tình huống từ DL mở là quá trình tốn nhiều thời gian, trong khi GV còn phải đảm nhận nhiều nhiệm vụ giảng dạy khác. Thiếu tài liệu hướng dẫn và cộng đồng chuyên môn chia sẻ: GV ít có cơ hội tiếp cận các ví dụ mẫu, ngân hàng DL mở dùng cho giáo dục, hay các tài liệu thiết kế học liệu tích hợp DL thực.

Đối với SV: SV chưa được hướng dẫn đầy đủ cách tìm kiếm, lựa chọn, làm sạch và trực quan hóa DL từ các nguồn mở. Họ chủ yếu học thụ động theo nội dung GV cung cấp sẵn. SV chưa được tổ chức các hoạt động học gắn với dữ liệu thực tế: Các bài tập hoặc đánh giá vẫn mang tính lý thuyết hoặc mô phỏng đơn giản, không yêu cầu SV làm việc trực tiếp với bộ DL mở. Nhiều SV còn bỡ ngỡ với các công cụ phân tích DL như Excel nâng cao, SPSS, Python ..., dẫn đến khó khăn khi làm bài tập có yếu tố thực tiễn. Trường chưa có kho dữ liệu nội bộ (ví dụ từ khảo sát SV, DL học tập, nghiên cứu khoa học SV...) để GV và SV có thể sử dụng phục vụ giảng dạy và học tập. Nhà trường chưa có định hướng rõ ràng trong việc khuyến khích sử dụng DL mở như một xu thế đổi

mới dạy học, dẫn đến việc khai thác chỉ mang tính tự phát ở một số GV cá nhân.

2.4. Một số giải pháp nâng cao hiệu quả khai thác dữ liệu mở khi giảng dạy học phần Xác suất thống kê tại trường trường Đại học Hải Dương

2.4.1. Tích hợp dữ liệu mở vào bài giảng học phần Xác suất Thống kê

a. Mục đích, ý nghĩa: Việc tích hợp dữ liệu mở vào bài giảng Xác suất – Thống kê nhằm tăng tính thực tiễn và hấp dẫn cho môn học, giúp sinh viên hiểu được sự liên hệ giữa lý thuyết và thực tiễn đời sống, kinh tế – xã hội. Ngoài ra, việc này còn góp phần rèn luyện năng lực xử lý, phân tích dữ liệu – kỹ năng thiết yếu trong thời đại số và giúp SV làm quen với dữ liệu lớn, dữ liệu phi cấu trúc, phục vụ cho học tập bậc cao và công việc sau này.

b. Cách thức thực hiện giải pháp

Bước 1: Chọn nội dung học phần phù hợp để tích hợp DL mở, như thống kê mô tả, ước lượng, kiểm định giả thuyết, hồi quy – tương quan.

Bước 2: Tìm kiếm, xử lý DL mở từ các nguồn uy tín như Cổng DL mở quốc gia, Tổng cục Thống kê, World Bank... Ví dụ, khi giảng bài “ước lượng khoảng tin cậy”, GV có thể dùng DL tỷ lệ sinh viên có việc làm sau tốt nghiệp theo ngành.

Bước 3: Thiết kế nội dung bài giảng bằng cách đưa các bài tập, tình huống gắn liền với DL mở – như phân tích phân phối cân nặng từ DL thực tế.

Bước 4: Tổ chức lớp học theo hướng nhóm nhỏ, tương tác với DL, sử dụng phần mềm Excel, SPSS để xử lý và trình bày kết quả.

c. Chú ý khi thực hiện giải pháp: Chọn DL phù hợp trình độ SV, tránh quá lớn hay quá phức tạp; không sử dụng DL vi phạm riêng tư. Hướng dẫn SV dùng phần mềm hỗ trợ và luôn đảm bảo mục tiêu học thuật thống kê không bị che khuất bởi kỹ thuật xử lý dữ liệu.

2.4.2. Biên soạn tài liệu và hướng dẫn sử dụng dữ liệu mở trong học tập

a. Mục đích, ý nghĩa

Hỗ trợ SV chủ động tiếp cận, sử dụng và khai thác DL mở trong quá trình học, nghiên cứu, làm bài tập, khóa luận. Đồng thời giúp hình thành kỹ năng thu thập và xử lý DL từ nhiều nguồn khác nhau.

b. Cách thức thực hiện

GV xây dựng danh mục các nguồn DL mở đáng tin cậy.

Biên soạn cẩm nang ngắn hướng dẫn cách truy cập, tải, xử lý và sử dụng phần mềm khai thác DL.

Tổ chức hoạt động thực hành: cho SV tự chọn chủ đề xã hội (giá xăng, ô nhiễm, tai nạn...) và thực hiện phân tích thống kê cơ bản bằng bảng, biểu đồ.

c. Chú ý khi thực hiện giải pháp

Ngôn ngữ hướng dẫn nên đơn giản, minh họa trực quan bằng hình ảnh, video ngắn. Cần hướng dẫn SV kiểm tra độ tin cậy và nguồn gốc dữ liệu. Khuyến khích SV ghi lại nhật ký học tập hoặc báo cáo ngắn sau mỗi lần thực hành khai thác DL mở.

2.4.3. Sử dụng dữ liệu mở trong kiểm tra, đánh giá kết quả học tập

a. Mục đích, ý nghĩa

Kiểm tra năng lực vận dụng kiến thức XSTK vào giải quyết vấn đề thực tiễn, thay vì chỉ đánh giá lý thuyết thuần túy. Góp phần phát triển tư duy phân tích DL và kỹ năng xử lý thông tin.

b. Cách thức thực hiện

Thiết kế đề kiểm tra, bài tập đánh giá dựa trên DL thực: Thay vì ra đề dạng truyền thống (bài toán đóng), GV thiết kế đề mở yêu cầu xử lý tình huống với DL thực. Cho phép SV sử dụng máy tính, phần mềm Excel hoặc các công cụ thống kê để xử lý DL.

Chấm điểm dựa trên tiêu chí: Khả năng xử lý và trực quan hóa DL, tính đúng, logic của phép tính và mức độ hiểu và vận dụng kiến thức thống kê, khả năng giải thích, trình bày kết quả.

c. Chú ý khi thực hiện giải pháp

Phải hướng dẫn SV kỹ cách tiếp cận và phân tích DL trước khi kiểm tra. Tránh đưa vào đề kiểm tra những bộ DL quá lớn, chưa xử lý hoặc không có ý nghĩa rõ ràng. Có thể tổ chức theo hình thức cá nhân hoặc nhóm, ưu tiên hình thức mở thay vì thi trắc nghiệm đóng kín.

III. KẾT LUẬN

Trong bối cảnh giáo dục đại học đang chuyển dịch mạnh mẽ theo hướng phát triển năng lực người học, việc khai thác DL mở trong giảng dạy học phần XSTK không chỉ là một xu thế đổi mới, mà còn là yêu cầu tất yếu để nâng cao tính thực tiễn, tính ứng dụng và sự hứng thú trong học tập của SV. Thông qua việc tích hợp DL mở vào

bài giảng, xây dựng tài liệu hướng dẫn SV tiếp cận và sử dụng DL, cũng như vận dụng DL mở trong kiểm tra, đánh giá, GV có thể tạo ra môi trường học tập gắn liền với thực tiễn, phát triển tư duy phân tích và kỹ năng xử lý thông tin cho người học.

Thực tiễn triển khai tại Trường Đại học Hải Dương cho thấy tiềm năng lớn của DL mở trong

việc hỗ trợ giảng dạy XSTK, đồng thời cũng bộc lộ một số khó khăn của GV, sự sẵn sàng của SV, cũng như việc lựa chọn dữ liệu phù hợp với nội dung học tập. Bài báo này kỳ vọng góp phần mở rộng nhận thức về vai trò của DL mở trong dạy học XSTK, đồng thời cung cấp một số định hướng cụ thể để triển khai hiệu quả tại các cơ sở giáo dục đại học, trong đó có Trường Đại học Hải Dương.

TÀI LIỆU THAM KHẢO

1. American Statistical Association. (2014). 2014 ASA Presidential Address: The Joy of Statistics - A Passion for the Art and Science of Helping People Make Decisions With Data.
2. Gould, R. (2010). Statistics and the Modern Student. *International Statistical Review*, 78(2), 297–315.
3. Hardin, J. (2018). Dynamic Data in the Statistics Classroom. *Technological Innovations in Statistics Education*, 11(1)
4. Roberts et al. (2019), Incorporating Open Data Into Introductory Courses in Statistics, *Journal of Statistics Education*, 27(3)
5. Schutt, R., & O'Neil, C. (2013). *Doing Data Science: Straight Talk from the Frontline*. O'Reilly Media.
6. Phạm Lê Văn. (2024). Kinh nghiệm xây dựng, khai thác học liệu mở phục vụ công tác đào tạo ở Việt Nam hiện nay. *Tạp chí Quản lý nhà nước*.
7. Wright, C., Meng, Q., Breshock, M. R., Atta, L., Taub, M. A., Jager, L. R., Muschelli, J., & Hicks, S. C. (2023). Open Case Studies: Statistics and Data Science Education through Real-World Applications. *arXiv*.