

ỨNG DỤNG DỮ LIỆU LỚN TRONG CÔNG TÁC ĐÀO TẠO VÀ HƯỚNG PHÁT TRIỂN TẠI HỌC VIỆN CHÍNH TRỊ QUỐC GIA HỒ CHÍ MINH

★ ThS LÒ THỊ PHƯƠNG NHUNG

★ ThS NGUYỄN MAI PHƯƠNG

Học viện Chính trị quốc gia Hồ Chí Minh

● **Tóm tắt:** *Dữ liệu lớn (Big data) là một trong những vấn đề cốt lõi trong cuộc Cách mạng công nghiệp 4.0. Hiện nay, dữ liệu lớn đã và đang được ứng dụng trong mọi mặt của đời sống xã hội. Trong công tác đào tạo, dữ liệu lớn làm thay đổi căn bản từ công tác quản lý, vận hành đến hoạt động giảng dạy, nghiên cứu. Bài viết tổng quan về ứng dụng dữ liệu lớn trong công tác đào tạo trên thế giới và Việt Nam, trên cơ sở đó đề xuất một số giải pháp nhằm ứng dụng hiệu quả dữ liệu lớn tại Học viện Chính trị quốc gia Hồ Chí Minh (Học viện).*

● **Từ khóa:** *dữ liệu lớn, ứng dụng dữ liệu lớn, giáo dục - đào tạo, Học viện Chính trị quốc gia Hồ Chí Minh.*

Dữ liệu lớn là thuật ngữ chỉ một tập hợp dữ liệu rất lớn và phức tạp mà những công cụ, ứng dụng xử lý dữ liệu truyền thống không thể đảm nhiệm được. Dữ liệu lớn là cốt lõi để sử dụng, phát triển internet vạn vật (IoT) và trí tuệ nhân tạo (AI). Dữ liệu lớn chứa nhiều thông tin phục vụ cho việc kinh doanh, nghiên cứu khoa học, giáo dục, dự đoán dịch bệnh sắp phát sinh và thậm chí là xác định tình trạng giao thông theo thời gian thực...

Ứng dụng dữ liệu lớn đem đến những lợi thế vượt bậc trong giáo dục như chương trình giảng dạy tùy chỉnh, cải thiện hệ thống đánh giá, định hướng nghề nghiệp cho sinh viên và đề xuất

phương pháp học tập phù hợp. Sự phát triển của dữ liệu lớn giúp các cơ sở đào tạo kiểm soát tốt hơn thông tin học viên; đổi mới phương thức học và giảng dạy theo hướng tích cực, hiệu quả hơn.

1. Tổng quan về ứng dụng dữ liệu lớn trong công tác đào tạo trên thế giới và Việt Nam

Trên thế giới, ứng dụng dữ liệu lớn trong hoạt động giáo dục, đào tạo thường tồn tại dưới dạng dữ liệu hành chính và dữ liệu quá trình học tập. Sự xuất hiện của dữ liệu lớn trong giáo dục bắt nguồn từ ít nhất hai xu hướng chính trong kỷ nguyên kỹ thuật số. *Thứ nhất*, việc ghi chép và lưu trữ dữ liệu trong môi trường truyền thống ngày càng được số hóa, dẫn đến lượng lớn thông



Học viên nghiên cứu, học tập tại Thư viện Học viện Chính trị quốc gia Hồ Chí Minh _ Ảnh: Viện TTKH

tin được chuẩn hóa. Cụ thể, Hệ thống thông tin học sinh (SIS) đã được áp dụng rộng rãi để lưu trữ và sắp xếp thông tin hồ sơ học sinh (thí dụ: nhân khẩu học, nền tảng học vấn) và hồ sơ học tập (thí dụ: ghi danh khóa học và kết quả học tập). SIS giúp quản lý và phân tích các dữ liệu của học viên tại một cơ sở giáo dục trong nhiều thập kỷ trên quy mô lớn. Thứ hai, các hoạt động học tập vốn khó ghi chú trong các lớp học truyền thống (face-to-face) giờ đây có thể được ghi lại một phần thông qua Hệ thống quản lý học tập (LMS). Trong hầu hết các trường hợp, LMS được giảng viên sử dụng để phân phối tài liệu giảng dạy, quản lý bài tập của học viên và giao tiếp với học viên. Từ việc truy cập vào các mô đun khóa học đến sửa chữa bài tiểu luận, những thông tin này có thể dễ dàng được lưu trữ tới hàng nghìn mục dữ liệu mỗi học viên. Ngoài SIS và LMS, những đổi mới trong môi trường học tập kỹ thuật số làm phong phú thêm các hình thức su

phạm, đồng thời thu thập “dữ liệu số” của học sinh. Sự đa dạng này dẫn đến dữ liệu được đa phương thức với khối lượng lớn.

Các kỹ thuật khai thác dữ liệu cho dữ liệu lớn trong giáo dục được Baker và Siemens⁽¹⁾ phân loại thành nhiều phương pháp, bao gồm: các phương pháp suy luận mô hình hóa kiến thức; thuật toán khám phá cấu trúc (chú trọng cấu trúc nội dung, kỹ năng giáo dục và cấu trúc mạng xã hội của người học); khai thác mối quan hệ (bao gồm khai thác mẫu tuần tự và khai thác tương quan); hình dung; và khám phá mô hình (bao gồm việc sử dụng mô hình trong các phân tích tiếp theo). Với khối lượng, tốc độ và sự đa dạng, tất cả “dữ liệu lớn” cho thấy thu thập thông tin về hành vi của người học có nhiều giá trị trong lĩnh vực nghiên cứu giáo dục. Các dữ liệu số về hành động của học sinh mang lại thông tin cụ thể và chi tiết hơn về quá trình học tập. Bằng cách kết hợp dữ liệu hành vi với khảo

sát hoặc thang đo tâm lý, các nhà nghiên cứu có thể lập bản đồ chuỗi hành động với các đặc điểm nhận thức và kiểm tra xem các dấu vết hành vi quan sát được có phù hợp với các giả định lý thuyết hay không và tinh chỉnh lý thuyết ở cấp độ chi tiết. Những thông tin phong phú này giúp các nhà nghiên cứu hiểu rõ cơ chế tác động của chính sách cụ thể và giải quyết các vấn đề liên quan đến chính sách.

Các dữ liệu lớn trong lĩnh vực giáo dục hiện nay bao gồm: dữ liệu lớn cấp vi mô (thí dụ: dữ liệu luồng nhấp chuột), dữ liệu lớn cấp trung bình (thí dụ: dữ liệu văn bản) và dữ liệu lớn cấp vĩ mô (thí dụ: dữ liệu thể chế).

Dữ liệu lớn cấp vi mô là dữ liệu tương tác “siêu nhỏ” với số giây giữa các hành động có thể thu thập dữ liệu riêng lẻ từ hàng triệu người học tiềm năng. Hầu hết dữ liệu cấp vi mô được thu thập tự động trong quá trình tương tác giữa người học và môi trường học tập tương ứng của họ, bao gồm hệ thống dạy học thông minh, các khóa học trực tuyến lớn (MOOC), các trò chơi mô phỏng.

Dữ liệu lớn cấp trung bình bao gồm các sản phẩm viết của học viên được số hóa và thu thập một cách có hệ thống trong nhiều môi trường học tập khác nhau, từ các bài tập trong khóa học đến tham gia diễn đàn thảo luận trực tuyến, hệ thống dạy học thông minh và các tương tác trên mạng xã hội. Đáng chú ý, dữ liệu cấp trung bình mang lại cơ hội nắm bắt thông tin sơ cấp về sự tiến bộ của người học trong khả năng nhận thức và xã hội, cũng như trạng thái cảm xúc.

Dữ liệu lớn cấp vĩ mô bao gồm dữ liệu được thu thập ở cấp thể chế. Thí dụ, dữ liệu nhân khẩu học và nhập học của học viên, dữ liệu dịch vụ trong khuôn viên trường, lịch học và dữ liệu đăng ký khóa học, cũng như dữ liệu về yêu cầu chính của trường đại học và dữ liệu về các yêu cầu đối với học sinh hoàn thành bằng cấp. Mặc

dù dữ liệu lớn cấp vĩ mô thường được thu thập trong khoảng thời gian nhiều năm nhưng không được cập nhật thường xuyên, thường chỉ một hoặc hai lần mỗi kỳ (thí dụ: thông tin về lịch trình khóa học, hồ sơ điểm).

Ở Việt Nam hiện nay, ngành giáo dục đã xác định ứng dụng công nghệ thông tin là 1 trong 9 nhóm nhiệm vụ trọng tâm nhằm triển khai có hiệu quả Nghị quyết Hội nghị Trung ương 8 khóa XI (ngày 4-11-2013) về đổi mới căn bản, toàn diện giáo dục và đào tạo. Tháng 1-2017, Thủ tướng Chính phủ có Quyết định số 117/QĐ-TTg phê duyệt đề án Tăng cường ứng dụng công nghệ thông tin trong quản lý và hỗ trợ các hoạt động dạy - học, nghiên cứu khoa học góp phần nâng cao chất lượng giáo dục và đào tạo giai đoạn 2016-2020, định hướng đến năm 2025. Tháng 6-2020, Thủ tướng Chính phủ ra Quyết định số 749/QĐ-TTg phê duyệt Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030, theo đó, giáo dục là 1 trong 8 lĩnh vực trọng điểm cần được ưu. Bên cạnh đó, hàng loạt chính sách thúc đẩy chuyển đổi số giáo dục được ban hành, như: các quy định ứng dụng công nghệ thông tin trong quản lý, tổ chức đào tạo qua mạng, quy chế đào tạo từ xa trình độ đại học, quy định quản lý, vận hành sử dụng hệ thống cơ sở dữ liệu toàn ngành, mô hình ứng dụng công nghệ thông tin trường phổ thông, chuẩn dữ liệu kết nối; hướng dẫn nhiệm vụ công nghệ thông tin cho khối đại học, phổ thông hàng năm và nhiều văn bản chỉ đạo điều hành khác. Dữ liệu lớn không chỉ được ứng dụng trong quản lý, điều hành giáo dục mà còn trở thành một ngành học mới ở Việt Nam.

2. Một số giải pháp ứng dụng hiệu quả dữ liệu lớn trong công tác đào tạo tại Học viện

Trong bối cảnh số hóa, với mục tiêu xây dựng hệ thống quản trị Học viện thông minh; “tăng

cường ứng dụng công nghệ thông tin vào các bài giảng, vào công tác quản lý lớp, quản lý học viên; xây dựng cơ sở vật chất hiện đại, đồng bộ để Học viện xứng đáng là trung tâm quốc gia đào tạo, bồi dưỡng cán bộ lãnh đạo, quản lý, cán bộ khoa học lý luận của hệ thống chính trị⁽²⁾ thì việc nghiên cứu và ứng dụng dữ liệu lớn trong giáo dục vào thực tế công tác giảng dạy, đào tạo và nghiên cứu khoa học của Học viện là nhiệm vụ cấp thiết, cần sớm triển khai thực hiện. Với điều kiện thực tế, Học viện cần có phương hướng chung trong xây dựng lộ trình ứng dụng dữ liệu lớn vào các hoạt động đào tạo, nghiên cứu và giảng dạy, cụ thể: *Thứ nhất*, hình thành hệ thống dữ liệu lớn của Học viện từ hai nguồn: (1) số hóa toàn bộ dữ liệu của hệ thống Học viện (dữ liệu nội sinh) và (2) liên kết, trao đổi, hợp tác trong nước và quốc tế để khai thác, xử lý dữ liệu nhằm xây dựng dữ liệu lớn của Học viện (dữ liệu ngoại sinh). *Thứ hai*, vận hành và ứng dụng dữ liệu lớn (bao gồm: dữ liệu lớn của Học viện và dữ liệu lớn của quốc gia và quốc tế) trong các hoạt động của Học viện. *Thứ ba*, nghiên cứu và xây dựng các đề án chi tiết bảo đảm việc hình thành, ứng dụng dữ liệu lớn trong công tác đào tạo, nghiên cứu và giảng dạy tại Học viện.

Nghiên cứu và xây dựng lộ trình ứng dụng dữ liệu lớn trong hoạt động giáo dục tại Học viện cần thực hiện một số giải pháp sau:

Một là, nâng cao nhận thức về việc ứng dụng công nghệ thông tin, số hóa và ứng dụng dữ liệu lớn trong hoạt động quản lý đào tạo, nghiên cứu, giảng dạy của Học viện

Công nghệ mang đến những lợi thế vượt bậc trong giáo dục - đào tạo như chương trình giảng dạy tùy chỉnh, cải thiện hệ thống đánh giá, định hướng nghề nghiệp cho học viên và đề xuất phương pháp học phù hợp, hỗ trợ quản lý, giám sát hiệu quả, nâng cao chất lượng giảng dạy...

Theo xu hướng hiện nay, việc ứng dụng dữ liệu lớn trong giáo dục - đào tạo chủ yếu trong các khâu lưu trữ, quản lý, phân tích bộ dữ liệu lớn từ hồ sơ học viên với tiêu chí lấy người học làm trung tâm. Dữ liệu lớn có khả năng thay đổi cách thức tổ chức hoạt động, tương tác với học viên cũng như quản lý, vận hành hoạt động của cơ quan/đơn vị. Tại Học viện, việc ứng dụng dữ liệu lớn có thể giúp thay đổi cách thức tổ chức, quản lý đào tạo, chương trình và phương pháp giảng dạy, giải quyết các vấn đề tồn đọng trong hệ thống, nâng cao hiệu quả chương trình học tập. Đảng ủy, Ban Giám đốc Học viện, Ban Giám đốc các Học viện trực thuộc cũng như lãnh đạo Vụ Quản lý đào tạo, các trung tâm, các Ban Quản lý đào tạo cần nâng cao nhận thức về tính cấp thiết của việc ứng dụng công nghệ thông tin, số hóa nói chung và ứng dụng dữ liệu lớn nói riêng trong hoạt động quản lý đào tạo, nghiên cứu, giảng dạy của Học viện. Nâng cao nhận thức về vai trò, lợi ích của dữ liệu lớn và việc ứng dụng trong công việc.

Hai là, thực hiện một số bước trong quy trình chuẩn hóa dữ liệu nhằm hình thành dữ liệu lớn từ nguồn nội sinh của Học viện

Dữ liệu lớn vừa là cách thức tổ chức vừa là một hình thức lưu trữ dữ liệu trong ngành giáo dục và đào tạo. Dữ liệu lớn trong các trường học được hình thành từ nhóm các dữ liệu trong các mảng: trường học, lớp học, học viên, đội ngũ cán bộ, cơ sở vật chất, học liệu, bài thi, điểm thi... Thực chất, đây là quá trình chuyển đổi số các dữ liệu, tức hoạt động số hóa và công cụ xử lý tương ứng. Do đó, Học viện cần số hóa, chuẩn hóa lại dữ liệu để có thể thu thập, tập hợp dữ liệu vào cơ sở dữ liệu chủ - dữ liệu lớn dùng chung để khai thác. Dữ liệu của Học viện có thể thu thập từ tất cả các hoạt động quản lý, phục vụ của các vụ chức năng và hoạt động giảng dạy, nghiên cứu

của các viện chuyên ngành cùng cơ sở dữ liệu của thư viện. Dữ liệu lớn của Học viện bao gồm: Dữ liệu lớn cấp vĩ mô bao gồm các dữ liệu về thông tin học viên, chương trình giảng dạy, thời khóa biểu, thông tin về cán bộ, giảng viên, nghiên cứu viên được tổng hợp từ Vụ Quản lý đào tạo, Vụ Tổ chức - Cán bộ, Vụ Quản lý khoa học. Dữ liệu lớn cấp trung bình bao gồm: dữ liệu về các bài luận, tiểu luận, luận văn, luận án, các bài kiểm tra của học viên, đề tài các cấp. Dữ liệu lớn cấp vi mô được thu thập thông qua hoạt động của học viên trên các ứng dụng học trực tuyến, qua việc sử dụng các tài nguyên trong thư viện số và thậm chí qua việc truy cập của học viên vào mạng internet của Học viện. Như vậy, cần chuẩn hóa các quy trình, nghiệp vụ của các đơn vị chức năng cũng như các đơn vị trực thuộc để các dữ liệu thu thập từ các đơn vị được xử lý phù hợp và tích hợp vào cơ sở dữ liệu dùng chung của Học viện.

Công tác nghiên cứu khoa học của Học viện đã hình thành nên một hệ dữ liệu số tương đối lớn từ các công trình nghiên cứu khoa học các cấp do Học viện là cơ quan chủ trì nghiên cứu, các công trình do cán bộ của hệ thống Học viện tham gia hoặc chủ trì nghiên cứu, cũng như lý lịch khoa học của các nhà khoa học trong hệ thống Học viện. Nhóm dữ liệu lớn này sẽ phục vụ thiết thực cho công tác xã hội hóa các công trình nghiên cứu, tăng tính hiệu quả của kết quả nghiên cứu khi trở thành dữ liệu dùng chung, tránh lãng phí tài nguyên thông tin khoa học và góp phần giảm thiểu tình trạng “đạo văn”, trùng lặp vấn đề nghiên cứu.

Việc hình thành dữ liệu nội sinh thông qua chuyển đổi số toàn bộ quy trình hoạt động, dữ liệu các mặt công tác của Học viện vừa mang tính xây dựng, vừa mang tính ứng dụng. Bởi thực chất, trong quá trình số hóa, các dữ liệu đó được

bảo mật ở dạng thức số và quay lại phục vụ cho các hoạt động/lĩnh vực công tác khác của chính Học viện.

Ba là, nghiên cứu lựa chọn mô hình dữ liệu lớn và truy xuất dữ liệu lớn phù hợp với đặc điểm về dữ liệu cũng như chức năng, nhiệm vụ của Học viện

Cần phân tích cụ thể đặc điểm về cơ cấu tổ chức, chức năng, nhiệm vụ của Học viện nói chung và các đơn vị trực thuộc để lựa chọn mô hình phù hợp cho lưu trữ dữ liệu giáo dục của Học viện, giúp dễ dàng phân tích, truy xuất các thông tin nhằm nâng cao chất lượng giảng dạy, học tập, nghiên cứu, đồng thời giúp nâng cao hiệu quả hoạt động của các đơn vị chức năng.

Lựa chọn được mô hình dữ liệu lớn phù hợp sẽ giúp tích hợp đầy đủ các dữ liệu lớn cả cấp vi mô, vĩ mô và trung bình trong quá trình hoạt động của Học viện, tránh lãng phí nguồn lực. Thực tế trên thế giới hiện nay, việc ứng dụng dữ liệu lớn đã mang lại nhiều lợi ích cho cả cơ sở đào tạo, người học cũng như các đơn vị liên quan và rộng hơn là quốc gia, quốc tế. Với đặc điểm liên kết mở, môi trường số năng động và tính năng thu thập, lưu trữ, trích xuất ứng dụng, việc lựa chọn mô hình ứng dụng dữ liệu lớn trong Học viện cần tính tới những yêu cầu về kỹ thuật, công nghệ phù hợp với điều kiện và đặc thù của Học viện.

Đối tượng học viên của Học viện là các cán bộ dự nguồn cấp chiến lược, cán bộ đang giữ chức vụ quản lý và cán bộ thuộc diện quy hoạch. Do đó, lựa chọn mô hình dữ liệu lớn cần chú trọng vấn đề an toàn, bảo mật thông tin, tránh rò rỉ thông tin mật, thông tin cá nhân; cũng như giải quyết tốt sự xung đột giữa yêu cầu về quyền riêng tư và những yêu cầu về sự cởi mở và minh bạch hơn của môi trường mạng internet. Chìa khóa bảo đảm xây dựng thành công và khai thác hiệu quả cơ sở dữ liệu cho Học viện là phải xây dựng

được cơ chế để vừa đáp ứng an toàn và bảo mật thông tin cá nhân, vừa cho phép phân tích, khai thác dữ liệu để nâng cao chất lượng dạy và học.

Bốn là, tăng cường quản trị dữ liệu và sử dụng các công cụ để phân tích dữ liệu lớn nhằm mang lại thông tin hữu ích góp phần đổi mới phương pháp đào tạo, bồi dưỡng và nghiên cứu

Quản trị dữ liệu là chìa khóa quan trọng để nâng cao chất lượng đào tạo, bồi dưỡng, nghiên cứu, giảng dạy; nâng cao hiệu suất, hiệu quả học tập của học viên. Việc phân tích dữ liệu lớn cũng đang biến đổi cách thức tương tác giữa cơ sở đào tạo với học viên và giảng viên. Với sự phát triển của khoa học dữ liệu, sự đổi mới trong cách thức thu thập và xử lý thông tin, nhiều đơn vị trong nước và quốc tế đã tận dụng được sức mạnh của Big data trong việc ra các quyết định quan trọng. Đây là con đường tất yếu mà Học viện cần hướng đến để biến các dữ liệu trở thành thông tin hữu dụng, phục vụ cho chính hoạt động của đơn vị. Học viện cần làm chủ hệ thống dữ liệu lớn và sử dụng các công cụ phân tích dữ liệu như là yếu tố hỗ trợ cho việc ra quyết định, nhờ đó tạo ra những tác động tích cực như nâng cao hiệu quả giảng dạy, nghiên cứu của cán bộ giảng viên và hiệu quả học tập của học viên.

Năm là, tăng cường đầu tư hạ tầng để phục vụ việc chuyển đổi số cũng như ứng dụng dữ liệu lớn trong quản lý, giảng dạy

Trong bối cảnh chuyển đổi số trong ngành giáo dục đang diễn ra mạnh mẽ, cần tăng cường đầu tư về hạ tầng công nghệ thông tin hiện đại, liên thông thống nhất trong toàn hệ thống Học viện, từ Học viện Trung tâm đến các Học viện trực thuộc và các trường chính trị tỉnh/thành phố, trường đào tạo cán bộ của các bộ/ngành Trung ương; từng bước xây dựng trung tâm tích hợp dữ liệu là đầu mối thu thập, lưu trữ và xử lý dữ liệu lớn. Đặc biệt, cần chú ý đầu tư hạ tầng

máy chủ để lưu trữ dữ liệu nhằm bảo đảm tính bảo mật thông tin. Ngoài ra, cần tập trung thực hiện Đề án Xây dựng thư viện điện tử, đầu tư xây dựng các phòng học đa phương tiện. Đầu tư các trang thiết bị công nghệ thông tin phục vụ giáo dục - đào tạo tại Học viện cần có trọng tâm, trọng điểm, tránh lãng phí. Ưu tiên các trang thiết bị có nhu cầu sử dụng cao, phục vụ trực tiếp cho nhiệm vụ giáo dục - đào tạo như máy tính, máy chiếu, bảng thông minh, thiết bị mô phỏng....

Sáu là, nâng cao chất lượng nguồn nhân lực để có thể xây dựng dữ liệu lớn của Học viện và tăng cường ứng dụng dữ liệu lớn nhằm nâng cao chất lượng giảng dạy và nghiên cứu

Trong thời gian tới, cần chú trọng xây dựng đội ngũ chuyên gia công nghệ thông tin, đặc biệt là các chuyên gia về khoa học dữ liệu - đóng vai trò trụ cột trong việc thiết kế và xây dựng dữ liệu lớn của Học viện. Bên cạnh đó, cần tăng cường bồi dưỡng kỹ năng sử dụng công nghệ thông tin cho cán bộ, giảng viên, nghiên cứu viên của Học viện nhằm thích ứng với việc chuyển đổi số. Đa dạng hóa các hình thức đào tạo, bồi dưỡng bằng các giải pháp cụ thể như: đào tạo, bồi dưỡng ngắn hạn, dài hạn; đào tạo, bồi dưỡng tập trung, không tập trung; mở rộng liên kết, hợp tác đào tạo, bồi dưỡng với các cơ quan khác ở địa phương và Trung ương... □

(1) Baker, R. S., & Siemens, G: *Educational data mining and learning analytics*, In R. K. Sawyer (Ed.): *Cambridge handbook of the learning sciences* (2nd ed.), Cambridge University Press, 2014, pp.253-274.

(2) *Học viện Chính trị quốc gia Hồ Chí Minh tổ chức Hội nghị trực tuyến tổng kết công tác năm 2020 và triển khai nhiệm vụ năm 2021*, <https://hcma.vn>, ngày 08-01-2021.