

# Đẩy mạnh ứng dụng dữ liệu lớn trong lĩnh vực giáo dục và đào tạo

ThS Lò Thị Phương Nhung, ThS Nguyễn Mai Phương

Học viện Chính trị quốc gia Hồ Chí Minh

**Dữ liệu lớn là yếu tố góp phần tạo nên diện mạo mới, mang tính đặc thù của cuộc cách mạng công nghiệp 4.0. Trong giáo dục và đào tạo (GD&ĐT), dữ liệu lớn có tác động và đem lại lợi ích to lớn 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 phù hợp... Chính vì những lợi ích này, chúng ta cần có những giải pháp thúc đẩy mạnh mẽ ứng dụng dữ liệu lớn trong GD&ĐT.**

## Vai trò dữ liệu lớn trong GD&ĐT

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 phù hợp... Dưới đây là một số lý do mà ứng dụng của dữ liệu lớn ngày càng trở nên cần thiết trong lĩnh vực GD&ĐT [1].

*Thiết kế hệ thống giáo dục phù hợp:* giúp các tổ chức giáo dục xây dựng một cấu trúc học tập chung cho mọi học viên nhưng vẫn đáp ứng được năng lực và thể mạnh khác nhau của các học viên.

*Cải thiện hệ thống đánh giá:* các trường có thể theo dõi năng lực của học viên thông qua nhiều môn học khác nhau cả trên phương diện cá nhân và tập thể, từ đó phát triển các giải pháp phù hợp nhằm hỗ trợ học viên. Phân tích thống kê điểm số của học viên qua các môn học khác nhau cũng giúp định hướng tốt hơn và xây dựng lộ trình học tập phù hợp.

*Giúp định hướng nghề nghiệp cho học viên:* phần lớn các trường học hiện nay không có hướng dẫn

cụ thể và định hướng gì cho nghề nghiệp tương lai của học viên. Dữ liệu lớn sẽ giúp giải quyết những tình huống khó xử thông qua việc phân tích dữ liệu học viên và tạo báo cáo thành tích. Nếu báo cáo đó chỉ ra rằng bạn là nhà văn sáng tạo, thì bạn được khuyến khích nên theo đuổi những nghề nghiệp liên quan đến lĩnh vực đó. Những tư vấn này thường được thực hiện bởi cố vấn thông qua các buổi phỏng vấn ngắn và đánh giá cho điểm. Nhưng ngày nay với sự phát triển của dữ liệu lớn, quá trình đó đã được cải thiện rất nhiều. Qua việc phân tích dữ liệu lớn, các nhà phân tích sẽ chỉ ra rằng bên cạnh là một nhà văn sáng tạo, bạn cũng giỏi về khoa học, điều này sẽ giúp bạn đưa ra quyết định nghề nghiệp đúng đắn hơn.

*Đề xuất phương pháp học mới:* một số học viên thường học bằng phương pháp đọc và ghi nhớ, nhưng một số có thể nắm bắt mọi thứ qua việc viết ra, hay một số ít hiểu bài tốt hơn bằng việc xem video hay sử dụng các phương pháp khác. Tuy nhiên, học viên thường phải làm quen với các phương pháp cụ thể mà đôi khi

phương pháp này không phù hợp và cản trở việc phát triển của học viên. Điều đó cũng ảnh hưởng đến kết quả học của học viên. Vậy tại sao không thiết kế các phương pháp học tập khác nhau và để học viên tự lựa chọn phương pháp học phù hợp. Dữ liệu lớn sẽ giúp ích cho việc này. Thông qua phân tích dữ liệu lớn, các trường có thể xây dựng phương pháp học khác nhau, giúp việc học trở nên thú vị hơn. Qua đó cũng giúp học viên hiểu thêm về thế mạnh, điểm yếu của bản thân và cung cấp cho họ những tài liệu học tập hữu ích.

## Tiềm năng ứng dụng

Các hạ tầng của giáo dục số trong bối cảnh ứng dụng những thành tựu của cuộc cách mạng công nghiệp 4.0 với những đặc tính nổi trội như: internet vạn vật (IoT), dữ liệu lớn, điện toán đám mây đã mang lại nhiều cơ hội và khả năng to lớn giúp cho việc tái tạo, sản sinh tri thức, chia sẻ thông tin, “san bằng” các rào cản trong việc tiếp cận kiến thức. Đặc biệt, trong những thập kỷ gần đây, với sự xuất hiện và những tính năng nổi trội của dữ liệu lớn

## Diễn đàn Khoa học và Công nghệ

đã đưa đến những giới hạn mới trong cách chúng ta giám sát, hiểu và đánh giá các quy trình trong giáo dục và đưa ra những quyết định sáng suốt trong nỗ lực cải thiện hiệu quả giáo dục. Tuy không có định nghĩa thống nhất duy nhất tồn tại nhưng dữ liệu lớn thường được đặc trưng bởi khối lượng lớn, tốc độ và sự đa dạng trong kỷ nguyên kỹ thuật số [2]. So với các thế hệ dữ liệu trước đó được thu thập thông qua nỗ lực đáng kể của con người, việc sử dụng phổ biến các công cụ kỹ thuật số trong cuộc sống hằng ngày tạo ra một lượng dữ liệu (khối lượng) chưa từng có với tốc độ ngày càng tăng (tốc độ) và từ các phương thức và thước đo thời gian khác nhau (đa dạng) [3]. Những dữ liệu này đòi hỏi các tài nguyên máy tính đáng kể và các phương pháp phân tích thay thế để xử lý và diễn giải.

Dữ liệu lớn được thu thập trong lĩnh vực giáo dục hiện nay bao gồm: dữ liệu lớn cấp vĩ mô (ví dụ: dữ liệu luồng nhấp chuột), dữ liệu lớn cấp trung bình (ví dụ: dữ liệu văn bản) và dữ liệu lớn cấp vĩ mô (ví dụ: dữ liệu thể chế). Dữ liệu lớn cấp vĩ 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 vĩ 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 mở trực tuyến lớn (MOOC), mô phỏng và trò chơi. Dữ liệu lớn cấp trung bình bao gồm các sản phẩm viết của học viên được máy tính hóa và thu thập một cách có hệ thống trong các hoạt động viết tại 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 tự nhiên dữ liệu thô 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. Ví dụ về dữ liệu lớn cấp vĩ mô bao gồm: 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 chúng không được cập nhật và thường chỉ một hoặc hai lần mỗi kỳ.

### **Ứng dụng dữ liệu lớn cấp vĩ mô**

Việc ứng dụng dữ liệu vĩ mô trong giáo dục hiện nay chủ yếu để tìm hiểu: 1) Các thành phần của nhận thức; 2) Siêu nhận thức và tự điều chỉnh; 3) Trạng thái tình cảm, đánh giá, phân loại; 4) Kiến thức của học sinh, trên cơ sở đó, nhiều cơ sở giáo dục cũng như các công ty dữ liệu về giáo dục hiện nay khai thác dữ liệu ở cấp độ vĩ mô để xác định các kiến thức có thể hiện thực hóa nhằm tăng cường việc giảng dạy và học tập và cách cá nhân hóa các tài nguyên giáo dục kỹ thuật số.

### **Ứng dụng dữ liệu lớn cấp trung bình**

Dữ liệu lớn cấp trung bình chủ yếu liên quan đến kho văn bản. Các bài viết trên máy tính của học

sinh được thu thập một cách có hệ thống trên quy mô lớn đang tăng lên khi văn bản học thuật chuyển từ văn bản giấy sang văn bản kỹ thuật số. Trong khi các bài thi mỗi năm một lần cấp quốc gia như kỳ thi ACT (bài thi đánh giá thí sinh về năng lực và những kiến thức cần thiết sẵn sàng cho việc học tập ở bậc đại học), SAT (bài thi chuẩn hóa cho việc đăng ký vào một số trường đại học của Mỹ) trước đây chính là cơ hội hiếm có để thu thập kho tài liệu viết lớn thì việc nộp bài tập của học sinh cho hệ thống quản lý học tập (LMS) chính là kho tài liệu viết lớn có thể truy cập được. Bên cạnh các bài tập của khóa học, dữ liệu dạng văn bản có thể bắt nguồn từ các diễn đàn thảo luận trực tuyến, hệ thống dạy kèm thông minh, cơ sở dữ liệu trang web, mã lập trình và nhiều nguồn khác. Mỗi điểm dữ liệu trung bình thường được thu thập trong các khoảng thời gian từ phút đến giờ. Tuy nhiên, một cá nhân có thể tham gia vào các hoạt động viết với tần suất và mức độ thường xuyên khác nhau. Ví dụ: một sinh viên có thể nộp bài tập viết mỗi tuần cho LMS trong một học kỳ để hoàn thành một lớp học nhưng có thể tham gia vào các tương tác trên mạng xã hội với cường độ khác nhau trong suốt nhiều năm của chương trình học.

Hiện nay, dữ liệu lớn cấp trung bình được ứng dụng để cung cấp thông tin chi tiết về: 1) Các quá trình nhận thức (ví dụ: hoạt động nhận thức, kiến thức và kỹ năng), 2) Các quá trình xã hội (ví dụ: cấu trúc diễn ngôn và cộng tác), 3) Các quá trình hành vi (ví dụ: sự tham gia và rời bỏ của người học), và 4) Các quá trình cảm xúc (ví dụ, tình cảm, động cơ).

**Ứng dụng dữ liệu lớn cấp vĩ mô**

Dữ liệu lớn cấp vĩ mô được thu thập trong khoảng thời gian nhiều năm, với tỷ lệ thu thập thấp so với các cấp độ khác. Ví dụ: dữ liệu tổ chức toàn trường đại học bao gồm dữ liệu nhân khẩu học và nhập học của sinh viên; ghi danh khóa học và điểm; lịch trình khóa học và mô tả khóa học... Những dữ liệu này không được cập nhật, nhiều nhất là vài tuần một lần và thường chỉ một hoặc hai lần mỗi kỳ. Ví dụ, thông tin nhân khẩu học của học sinh thường chỉ được thu thập một lần và chỉ được cập nhật theo yêu cầu của học sinh. Tuy nhiên, những dữ liệu như vậy có thể tạo cơ hội cho quản trị viên tham gia vào quá trình ra quyết định dựa trên dữ liệu để cải thiện việc ra quyết định hành chính, nâng cao trải nghiệm của học sinh và cải thiện năng lực quản trị của trường học.

Hiện nay, dữ liệu lớn cấp vĩ mô trong lĩnh vực giáo dục chủ yếu được ứng dụng phổ biến trong: 1) Hệ thống cảnh báo sớm, 2) Hướng dẫn khóa học và hệ thống thông tin, 3) Phân tích quản trị giáo dục.

**Đẩy mạnh ứng dụng dữ liệu lớn trong GD&ĐT**

Đẩy mạnh ứng dụng dữ liệu lớn trong GD&ĐT sẽ góp phần không bỏ lỡ cơ hội mà cuộc cách mạng công nghiệp lần thứ 4 mang lại. Vì vậy, các cấp, các ngành nói chung và ngành GD&ĐT ở nước ta nói riêng cần tập trung triển khai một số nội dung cụ thể sau:

*Một là*, tiếp tục đẩy mạnh triển khai Chính phủ điện tử, hướng đến Chính phủ số trong toàn ngành, trong đó chú trọng triển khai

hoàn thiện hệ thống cơ sở dữ liệu (CSDL) toàn ngành GD&ĐT kết nối, liên thông, chia sẻ dữ liệu từ Trung ương đến địa phương, nhà trường và đồng bộ với các CSDL quốc gia, CSDL chuyên ngành khác góp phần hình thành CSDL mở quốc gia; đẩy mạnh các dịch vụ công trực tuyến phục vụ người dân; thực hiện số hóa triệt để, sử dụng văn bản điện tử, sổ sách học bạ, sổ điểm điện tử thay thế văn bản, tài liệu giấy; hoạt động chỉ đạo, điều hành, giao dịch, họp, tập huấn được thực hiện chủ yếu trên môi trường mạng.

*Hai là*, tăng cường chất lượng công tác dự báo, hoàn thiện cơ chế, chính sách theo hướng đi trước một bước. Trong đó, chú trọng chính sách hoàn thiện CSDL quản lý giáo dục, theo các quy định về chia sẻ, khai thác dữ liệu; hoàn thiện hành lang pháp lý thúc đẩy phát triển hình thức dạy - học trực tuyến qua mạng; chính sách quản lý các khóa học trực tuyến đảm bảo chất lượng thông qua các quy định về điều kiện mở trường mở lớp, thẩm định cấp phép nội dung...

*Ba là*, hoàn thiện cơ sở hạ tầng mạng đồng bộ, thiết bị công nghệ thông tin thiết thực phục vụ dạy - học, tạo cơ hội học tập bình đẳng giữa các vùng, miền có điều kiện kinh tế - xã hội khác nhau, ưu tiên hình thức thuê dịch vụ và huy động nguồn lực xã hội hóa cùng tham gia thực hiện.

*Bốn là*, thúc đẩy phát triển học liệu số ở tất cả các cấp học, ngành học, môn học gắn với việc thẩm định nội dung, kết nối, chia sẻ học liệu giữa các địa phương, nhà trường; hình thành kho học liệu số, học liệu mở dùng chung toàn ngành, liên kết với quốc tế,

đáp ứng nhu cầu tự học, học tập suốt đời, thu hẹp khoảng cách giữa các vùng miền; tiếp tục đổi mới cách dạy và học trên cơ sở áp dụng công nghệ số, khuyến khích và hỗ trợ áp dụng các mô hình GD&ĐT mới dựa trên các nền tảng số.

*Năm là*, triển khai mạng xã hội giáo dục có sự kiểm soát và định hướng thống nhất, tạo môi trường số kết nối, chia sẻ giữa cơ quan quản lý giáo dục, nhà trường, gia đình, giáo viên, giảng viên, học sinh, sinh viên; phát triển các khóa học trực tuyến mở, hình thành mạng học tập mở của người Việt Nam; triển khai hệ thống học tập trực tuyến dùng chung toàn ngành phục vụ công tác bồi dưỡng giáo viên, hỗ trợ dạy học cho các vùng khó khăn.

*Sáu là*, đào tạo, bồi dưỡng đội ngũ cán bộ quản lý, nhà giáo về kiến thức, kỹ năng công nghệ thông tin, an toàn thông tin cần thiết để tác nghiệp trên môi trường số, đáp ứng yêu cầu chuyển đổi số.

**TÀI LIỆU THAM KHẢO**

[1] <https://insight.isb.edu.vn/ung-dung-cua-big-data-trong-giao-duc/>  
 [2] D. Laney (2001), "3D data management: Controlling data volume, velocity and variety", *META Group Research Note*, **6(70)**, p.1.  
 [3] I. Roll and P.H. Winne (2015), "Understanding, evaluating, and supporting self-regulated learning using learning analytics", *Journal of Learning Analytics*, **2(1)**, pp.7-12.