

KINH NGHIỆM CỦA MỘT SỐ NƯỚC TRÊN THẾ GIỚI VỀ ỨNG DỤNG DỮ LIỆU LỚN TRONG GIÁO DỤC VÀ BÀI HỌC CHO VIỆT NAM

TS. TRẦN THỊ NGỌC MINH^(*)

Tóm tắt: *Dữ liệu lớn (Big data) đã và đang đem lại nhiều lợi ích cho các quốc gia trên thế giới. Dữ liệu lớn có tác động mạnh mẽ đến nhiều lĩnh vực, trong đó có lĩnh vực giáo dục. Bài viết nghiên cứu kinh nghiệm ứng dụng hiệu quả dữ liệu lớn trong giáo dục của một số nước trên thế giới, như Mỹ, Anh, Nhật Bản và Singapore, từ đó rút ra những bài học cho Việt Nam.*

Từ khóa: *dữ liệu lớn; giáo dục; kinh nghiệm; Việt Nam*

1. Kinh nghiệm của một số nước trên thế giới về ứng dụng dữ liệu lớn trong giáo dục

- *Kinh nghiệm của Mỹ*

Các chuyên gia về học tập trực tuyến trong giáo dục đại học Mỹ dự đoán, những năm tới, phân tích học tập (Learning Analysis - LA) sẽ được sử dụng rộng rãi trong giáo dục trực tuyến để xác định kiểu hành vi của sinh viên, cải thiện tỷ lệ học tập và duy trì của sinh viên. Phân tích học tập là quá trình thu thập và phân tích một cách có hệ thống các tập dữ liệu lớn từ các nguồn trực tuyến, nhằm mục đích cải thiện quá trình học tập.

Phân tích học tập sử dụng các mô hình dự đoán cung cấp thông tin hữu ích. Đây là cách tiếp cận đa lĩnh vực dựa trên xử lý dữ liệu, nâng cao công nghệ học tập, khai thác dữ liệu giáo dục và trực quan hóa. Mục đích của phân tích học tập là điều chỉnh các cơ hội giáo dục cho phù hợp với nhu cầu và khả năng của cá nhân người học thông qua các hành động, như can

thiệp với học sinh có nguy cơ hoặc cung cấp phản hồi vào nội dung giảng dạy. Ngược lại, khai thác dữ liệu giáo dục cố gắng tạo ra các phản hồi có hệ thống và tự động cho người học. Trong khi phân tích học tập hướng vào việc áp dụng các phương pháp và mô hình đã biết để giải quyết các vấn đề ảnh hưởng đến việc học của học sinh và hệ thống tổ chức học tập, khai thác dữ liệu giáo dục tập trung vào việc phát triển các phương pháp phân tích dữ liệu tính toán mới.

Những năm gần đây, việc ứng dụng dữ liệu lớn và phân tích học tập trở nên phổ biến trong giáo dục đại học ở Mỹ. Ví dụ, hệ thống lưu trữ hồ sơ sinh viên đại học đã duy trì thông tin kết quả về sinh viên (như điểm trong mỗi khóa học). Thông tin này có thể được các nhà nghiên cứu tổ chức sử dụng để nghiên cứu các mô hình thành tích của học sinh theo thời gian, thường là từ học kỳ này sang học kỳ khác hoặc từ năm này sang năm khác. Dữ liệu sẽ được thu thập cho mỗi giao dịch của sinh viên trong một khóa học, đặc biệt nếu khóa học được cung cấp trực tuyến dưới dạng điện tử. Mọi mục nhập của sinh viên về đánh giá khóa học, mục nhập bảng thảo luận, mục nhập

^(*) Học viện Báo chí và Tuyên truyền, Học viện Chính trị quốc gia Hồ Chí Minh

blog hoặc hoạt động wiki có thể được ghi lại, tạo ra hàng nghìn giao dịch cho mỗi sinh viên trong mỗi khóa học. Hơn nữa, dữ liệu này sẽ được thu thập trong thời gian thực hoặc gần thực khi nó được giao dịch và sau đó được phân tích để đề xuất các hướng hành động.

Để ứng dụng dữ liệu lớn và phân tích học tập trong giáo dục, thì các yêu cầu xử lý giao dịch chủ yếu là điện tử, thay vì thủ công. Hướng dẫn trực tiếp truyền thống có thể hỗ trợ các quy trình ra quyết định dựa trên dữ liệu truyền thống, nhưng để chuyển sang các ứng dụng phân tích học tập rộng rãi hơn và đặc biệt nhạy cảm với thời gian, điều quan trọng là các giao dịch hướng dẫn phải được thu thập khi chúng xảy ra. Điều này có thể thực hiện được trong trường hợp hệ thống quản lý khóa học/quản lý học tập (CMS/LMS).

Hầu hết các CMS đều cung cấp khả năng giám sát liên tục hoạt động của học sinh cho dù đó là câu trả lời, bài đăng trên bảng thảo luận, quyền truy cập vào tài liệu đọc, hoàn thành bài kiểm tra hoặc một số đánh giá khác. Sử dụng đầy đủ các tính năng của CMS cơ bản, một khóa học trực tuyến mạnh mẽ kéo dài 15 tuần có thể tạo ra hàng nghìn giao dịch cho mỗi học viên. Ghi chép và phân tích thời gian thực của các giao dịch này có thể được sử dụng để cung cấp cho ứng dụng phân tích học tập. Điều quan trọng đối với loại ứng dụng này là không chờ đợi ứng dụng phân mềm phân tích học tập. Các giao dịch giảng dạy cũng nên được tích hợp với các tài nguyên khác, như dữ liệu từ hệ thống thông tin của trường đại học (sinh viên, khóa học, giảng viên) và một chương trình phân mềm phân tích.

Một lưu ý quan trọng là độ chính xác của dữ liệu không bao giờ được xâm phạm, vì tính kịp thời của dữ liệu, cả về độ chính xác và khi kết thúc kỳ chấm điểm hoặc học kỳ để ghi lại các biện pháp hiệu suất. Điều này quan trọng là do việc giám sát các giao dịch của sinh viên trên cơ sở thời gian thực cho phép các cảnh báo theo thời gian thực. Giảng viên hướng dẫn có thể thực hiện hoặc can thiệp kịp thời để hỗ trợ sinh viên. Do đó, một CMS hoặc một cái gì đó tương tự trở nên quan trọng nhằm thu thập và

cung cấp dữ liệu này vào cơ sở dữ liệu “lớn” để xử lý kịp thời là điều quan trọng và phải có trong ứng dụng phân tích học tập.

Trên thực tế, có 08 loại ứng dụng dữ liệu lớn và phân tích học tập trong giáo dục đại học đã được mô tả, gồm: 1) Giám sát thành tích cá nhân của sinh viên; 2) Phân biệt thành tích của sinh viên theo các đặc điểm đã chọn, như chuyên ngành, năm học, dân tộc...; 3) Xác định các ngoại lệ để can thiệp sớm; 4) Dự đoán tiềm năng để tất cả sinh viên đạt được thành tích tối ưu; 5) Ngăn chặn sự giảm sút số lượng sinh viên của một khóa học hoặc chương trình; 6) Xác định và phát triển các công nghệ giảng dạy hiệu quả; 7) Phân tích các kỹ thuật và công cụ đánh giá tiêu chuẩn; 8) Kiểm tra và đánh giá chương trình giảng dạy.

Trong các khóa học trực tuyến, CMS thường xuyên cung cấp số liệu thống kê theo dõi khóa học và hệ thống cảnh báo sớm thô sơ cho phép người hướng dẫn theo dõi những sinh viên không trả lời trên blog hoặc bảng thảo luận, không truy cập tài liệu đọc, hoặc không kịp thời làm bài tập. Các số liệu thống kê về khóa học này được duy trì theo thời gian thực và người hướng dẫn có thể xem lại chúng thường xuyên khi họ muốn. Một lần nữa, những sinh viên không tham gia như mong muốn có thể được gửi một email bày tỏ lo ngại về hiệu suất của họ. Không có can thiệp nào trong số này yêu cầu phân tích học tập, nhưng cách tiếp cận này có thể được nâng cao đáng kể bằng cách mở rộng số lượng và tính chất của dữ liệu được thu thập. Ví dụ, một phản hồi của học sinh trên bảng thảo luận có thể được phân tích để tìm ra các mẫu để xác định mức độ sâu sắc và chất lượng tham gia của học sinh với tài liệu khóa học. Những mẫu này được phát hiện bằng cách kiểm tra hàng nghìn, thậm chí hàng chục nghìn câu trả lời của sinh viên khác.

- *Kinh nghiệm của Vương quốc Anh*

Hiện nay, việc cung cấp và sử dụng dữ liệu của ngành giáo dục đại học tại Vương quốc Anh nằm trong danh sách những nước tốt nhất trên thế giới (thậm chí là dẫn đầu toàn cầu).

Vương quốc Anh là quốc gia duy nhất có cơ quan chuyên trách thu thập dữ liệu (HESA). Trong năm 2013, cơ quan này đã tiến hành khảo sát sâu rộng về bối cảnh thông tin hiện tại, xác định được 525 bộ sưu tập dữ liệu giáo dục đại học riêng biệt và 93 tổ chức riêng biệt thu thập dữ liệu này. Không có tổ chức nào sẽ phải áp dụng theo 525 bộ sưu tập hằng năm, vì nhiều bộ sưu tập trong số này dành riêng cho các khóa học cụ thể hoặc diễn ra định kỳ 6 tháng hoặc ít thường xuyên hơn⁽¹⁾. Trong số 93 tổ chức thu thập dữ liệu học sinh, có một số tổ chức lớn trong toàn ngành thực hiện nhiều hoạt động thu thập, như HESA, SLC và UCAS. Nhiều tổ chức khác như PSRB (các cơ quan chuyên môn, luật định và quản lý) thường chỉ tiến hành một số lượng nhỏ các bộ dữ liệu nhỏ để được công nhận. Tuy nhiên, hiện có sự trùng lặp trong bộ dữ liệu được thu thập bởi các tổ chức khác nhau và không có sự chia sẻ dữ liệu thường xuyên giữa những người thu thập dữ liệu, cũng như không có yêu cầu đối với những người thu thập phải cộng tác với nhau.

Hầu hết các hệ thống được cung cấp trong một cơ sở giáo dục sẽ thu thập dữ liệu về những gì sinh viên đang làm, do đó, cung cấp thông tin hữu ích, có thể được sử dụng trong phân tích học tập. Các chỉ số phổ biến nhất hiện đang được sử dụng trong phân tích hệ thống: Sử dụng VLE (phần mềm học trực tuyến); Điểm danh; Sử dụng thư viện (cả thông qua hệ thống trực tuyến và mượn sách vật lý); Nộp bài tập và điểm số... Các trường đại học cũng kết hợp dữ liệu từ hệ thống thông tin sinh viên (SIS) vào hệ thống phân tích của họ, bao gồm cả việc mở rộng dữ liệu về sự tham gia như quốc tịch và tình trạng kinh tế. Dữ liệu có thể bao gồm thông tin đăng nhập vào SIS để biết thời gian biểu và kết quả kỳ thi.

Một số trường bắt đầu khám phá khả năng kết hợp nhiều loại dữ liệu hơn vào hệ thống phân tích của họ. Đại học Lancaster đang xem xét việc thu thập và sử dụng dữ liệu mà sinh viên đang truy cập vào máy tính thư viện và trong bao lâu. Thực tế, không phải mọi tổ chức đều

sử dụng cùng một số liệu. Một số loại dữ liệu sẽ hữu ích hơn trong một số tổ chức nhất định so với những cơ sở khác. Cụ thể, Edinburgh nhận thấy rằng, thông tin về việc sử dụng VLE không quá hữu ích cho họ, bởi Edinburgh là một trường đại học chuyên đào tạo trực tuyến.

- Kinh nghiệm của Nhật Bản

Năm 2013, hai chương trình nghiên cứu đã được Cơ quan Khoa học và Công nghệ Nhật Bản (JST) tài trợ: 1) “Công nghệ cốt lõi tiên tiến để tích hợp dữ liệu lớn” hỗ trợ việc tạo ra, cải tiến và hệ thống hóa công nghệ thông tin sáng tạo và các phương pháp toán học cơ bản để thu được kiến thức và cái nhìn sâu sắc mới từ việc sử dụng Dữ liệu lớn trên các lĩnh vực khác nhau. Chương trình này được giám sát bởi Giáo sư Masaru Kitsuregawa, Tổng Giám đốc Viện Tin học quốc gia. 2) Chương trình “Công nghệ ứng dụng tiên tiến để tăng cường sử dụng dữ liệu lớn để khám phá khoa học nhiều lĩnh vực và giải quyết vấn đề xã hội”, tập trung vào việc ứng dụng công nghệ dữ liệu lớn; được giám sát bởi Giáo sư Yuzuru Tanaka tại Trường Cao học Khoa học Thông tin và Công nghệ tại Đại học Hokkaido.

- Chương trình Công nghệ cốt lõi tiên tiến để tích hợp dữ liệu lớn, nhằm hỗ trợ: 1) Tạo ra, cải tiến và hệ thống hóa công nghệ cốt lõi thế hệ tiếp theo để giải quyết các thách thức chung đối với một số lĩnh vực dữ liệu; 2) Phân tích tích hợp dữ liệu lớn trong nhiều lĩnh vực. Các mục tiêu phát triển cụ thể bao gồm: Công nghệ để vận hành ổn định các hệ thống quản lý dữ liệu quy mô lớn (nén, truyền và lưu trữ dữ liệu lớn); công nghệ thu thập hiệu quả kiến thức cần thiết bằng cách tìm kiếm, so sánh, trực quan hóa thông tin đa dạng và các phương pháp toán học, thuật toán cho các công nghệ trên.

Chương trình này bao gồm 04 đề xuất cụ thể như sau: 1) Thành lập NLP có cấu trúc chuyên sâu về kiến thức (xử lý ngôn ngữ tự nhiên) và xây dựng hạ tầng tri thức. 2) Thu thập và phân tích dữ liệu bảo mật - bảo mật với bảo đảm kiểm soát thông tin và ứng dụng của nó vào y học cá nhân hóa và dịch tễ học di truyền. 3) Dữ

liệu cực lớn - Sự hội tụ của dữ liệu lớn và HPC để xử lý Yottabyte. 4) Khám phá kiến thức sâu sắc từ dữ liệu phức hợp và giá trị tạo ra của nó.

Chương trình Công nghệ ứng dụng tiên tiến để tăng cường sử dụng dữ liệu lớn để khám phá khoa học nhiều lĩnh vực và giải quyết vấn đề xã hội nhằm hỗ trợ các dự án và nghiên cứu hợp tác. Trong đó, việc sử dụng dữ liệu lớn có thể mang lại tác động xã hội to lớn bằng cách giải quyết các vấn đề kinh tế - xã hội đầy thách thức và đạt được sự sáng tạo giá trị. Các lĩnh vực cụ thể được quan tâm là khoa học đời sống, khoa học vật liệu, chăm sóc sức khỏe và y tế, xã hội và kinh tế, hệ thống hạ tầng đô thị, phòng chống và giảm nhẹ thiên tai, nông nghiệp, lâm nghiệp và thủy sản, không gian bên ngoài và môi trường của trái đất. Các mục tiêu dài hạn là tạo ra thực nghiệm mới và nâng cao độ tinh vi của các công nghệ ứng dụng thế hệ tiếp theo cần thiết để đạt được các mục tiêu và thiết lập công nghệ hệ thống phân tích dữ liệu lớn toàn diện và tích hợp để sử dụng trong nhiều lĩnh vực.

Chương trình đã đưa ra **02 đề xuất cụ thể** sau:

Một là, phát triển nền tảng tạo tri thức được thúc đẩy bởi dữ liệu lớn trong nghiên cứu thuốc thông qua quy trình sản xuất. Mục tiêu của dự án này là thiết lập một nền tảng cho phép thống nhất dữ liệu và thúc đẩy nghiên cứu sẽ tối ưu hóa các hệ thống để tiếp cận phát triển được phẩm từ quan điểm toàn diện, tương quan và cấp cao. Nghiên cứu sẽ được thúc đẩy bởi dữ liệu lớn, bắt đầu bằng việc xác định các mẫu cho các hướng phát triển phân tử chỉ dựa trên khối lượng lớn dữ liệu hợp chất và protein. Các mẫu này sẽ được kết hợp với một thư viện hợp chất được tạo ảo và xác định các mục tiêu cho các hợp chất, với các hợp chất ứng cử viên được đánh giá thêm về tính khả thi tổng hợp và sản xuất của chúng. Số lượng lớn dữ liệu liên quan đến sản xuất cũng sẽ thúc đẩy việc phát triển các phương pháp đánh giá hoạt động an toàn của các nhà máy sản xuất để đảm bảo rằng chúng được trang bị đầy đủ. Mục tiêu cuối cùng là thiết lập các mô hình nâng cao để đánh giá rủi ro, quản lý rủi ro và kiểm soát chất lượng.

Hai là, đổi mới công nghệ đồng hóa dữ liệu lớn để cách mạng hóa dự báo thời tiết khắc nghiệt trong phạm vi rất ngắn. Nghiên cứu này nhằm đổi mới công nghệ đồng hóa để tận dụng triệt để dữ liệu lớn từ các công nghệ thế hệ tiếp theo của Nhật Bản như radar thời tiết mảng theo giai đoạn, vệ tinh thời tiết địa tĩnh và máy tính 10-Petaflops K hàng đầu thế giới. Hệ thống dự báo thời tiết số cập nhật siêu nhanh trong 30 giây sáng tạo để dự báo thời tiết khắc nghiệt trong 30 phút sẽ được phát triển, hỗ trợ phòng ngừa và giảm nhẹ thiên tai đồng thời tạo ra một bước đột phá khoa học trong khí tượng học.

Về Chương trình giáo dục: Chương trình cấp bằng thạc sĩ về phát triển nguồn nhân lực cho đổi mới dữ liệu lớn được phát triển vào năm 2013 tại các trường sau đại học của Đại học Keio. Sinh viên được yêu cầu tham gia một trong ba lớp học về các công cụ dữ liệu lớn trong năm đầu tiên, một khóa học về dữ liệu lớn thực tế trong năm thứ hai và ba trong số tám lớp học liên quan đến công nghệ cho dữ liệu lớn để hoàn thành chương trình học của mình. Thông qua thực hành, sinh viên phát triển các kỹ năng và kiến thức về quản lý và phân tích dữ liệu lớn. Sinh viên và giáo sư làm việc với các nhà cung cấp dữ liệu lớn (Hadoop, MongoDB và Mahout) để giải quyết các vấn đề dữ liệu lớn thực sự trong năm lĩnh vực ứng dụng sau: thương mại điện tử, khoa học đời sống, quản trị, lưu lượng truy cập và kiểm soát đám đông.

- Kinh nghiệm của Singapore

Hệ thống giáo dục Singapore là một trong những điểm mạnh nổi bật giúp cho đất nước này trở thành điểm đến của rất nhiều du học sinh trên thế giới.

Công nghệ đã và sẽ luôn là một phần quan trọng trong giáo dục Singapore. Tại Singapore, Big data Analytics tạo ra các cơ hội mới để cải thiện quá trình giáo dục bằng cách giúp giáo viên và người học đưa ra quyết định thông minh hơn sớm hơn trong quá trình học tập. Các phát triển, trong việc sử dụng khoa học dữ liệu để thúc đẩy đổi mới quy trình, đang phát triển

nhANH chóng. Có thể thấy, việc ứng dụng dữ liệu lớn trong giáo dục tạo ra các công cụ và ứng dụng mới để giúp sinh viên và giáo viên sử dụng thời gian của họ một cách hiệu quả.

Sử dụng tốt nhất dữ liệu lớn làm công cụ hỗ trợ việc ra quyết định của người dạy và người học, điều quan trọng là người dạy và người học phải hiểu cách phân tích dữ liệu có thể hỗ trợ quá trình học tập. Ở Singapore, ba cách phổ biến trong việc sử dụng dữ liệu lớn là tận dụng khả năng đo lường sự hiểu biết, cá nhân hóa trải nghiệm học tập và giúp thiết kế các khóa học thú vị hơn. Cụ thể:

Một là, ứng dụng dữ liệu lớn cho phép giảng viên đo lường, giám sát và phản hồi trong thời gian thực về sự hiểu biết của sinh viên về tài liệu. Phân tích, hiển thị phương pháp học tập của sinh viên học có thể giúp các nhà giáo dục điều chỉnh phương pháp giảng dạy của họ và đáp ứng yêu cầu của sinh viên. Đây là bước phát triển quan trọng đối với giảng viên, bởi nó sẽ củng cố khả năng của người học để giải quyết bất kỳ thành kiến vô thức nào có thể có đối với sự tham gia hoặc hiệu suất của sinh viên.

Hai là, ứng dụng phân tích dữ liệu lớn giúp cho các khóa học trở nên thú vị với mức độ kiến thức khác nhau. Sử dụng phân tích dữ liệu để hiểu nơi mỗi sinh viên gặp khó khăn hoặc xuất sắc có thể cho phép bạn cung cấp các tài liệu ban đầu khác nhau cho mỗi sinh viên trong cùng một khóa học. Điều này sẽ cải thiện sự quan tâm của sinh viên đối với môn học, chỉ ra ai và khi nào nội dung học cụ thể nên được chuyển giao.

Ba là, ứng dụng dữ liệu lớn giúp thiết kế khóa học mới. Một thách thức quan trọng đối với các trường kinh doanh (và các trường đại học) là nhanh chóng hiểu được nhu cầu của ngành và đưa ra chương trình giảng dạy để đáp ứng những nhu cầu đó. Big data Analytics có thể được sử dụng để hiểu xu hướng thị trường và việc làm cũng như: Tổ chức các khóa học giới thiệu và các nguyên tắc học tập cơ bản xung quanh các ý tưởng mới nổi trong thế giới kinh doanh.

2. Một số bài học cho Việt Nam trong việc ứng dụng dữ liệu lớn vào lĩnh vực giáo dục

Ứng dụng dữ liệu lớn trong lĩnh vực giáo dục là vô cùng quan trọng, cung cấp vô số lợi ích cho cả giáo viên, học viên lẫn trường học. Phân tích dữ liệu lớn cung cấp cho các bên liên quan đến giáo dục một cái nhìn tổng thể, toàn diện về hiệu quả hoạt động của tổ chức, chương trình giảng dạy, giáo viên, học viên và triển vọng việc làm sau khi tốt nghiệp đối với sinh viên. Nó cũng cung cấp cho các học giả và nhà nghiên cứu thông tin cần thiết để xác định khoảng cách giữa giáo dục và các ngành công nghiệp để nhà giáo dục và các cơ sở giáo dục có thể khắc phục những thiếu sót này trong khóa học. Quan trọng hơn, khả năng cung cấp những khám phá này của dữ liệu lớn có thể giúp lĩnh vực giáo dục đạt được tiến bộ đáng kể để cải thiện quy trình học tập. Hơn nữa, dữ liệu lớn thúc đẩy giáo dục vì nó cố gắng theo kịp tốc độ phát triển của hệ thống thông tin và công nghệ mới. Do đó, từ thực tế ứng dụng dữ liệu lớn trong lĩnh vực giáo dục, ***cần lưu ý một số điểm sau:***

Thứ nhất, phát huy tối đa những lợi ích, hiệu quả của dữ liệu lớn trong giáo dục đối với cả người dạy, người học và cơ sở đào tạo. Đối với giáo viên: dữ liệu lớn góp phần cải tiến và phát triển chương trình, phương pháp giảng dạy; cải thiện hiệu suất của giảng viên... Ngày nay, việc áp dụng công nghệ chuyển đổi số và dữ liệu lớn trong giáo dục có thể hỗ trợ toàn diện cho giáo viên trong công việc. Điều này góp phần giải phóng năng lượng cho giáo viên - để giáo viên không còn làm một khối lượng công việc khổng lồ ở nhà, mà chỉ tập trung vào giảng dạy trên lớp; tạo ra những “giáo viên hạnh phúc” - không phải làm những việc “vụn vặt” hoặc lặp đi lặp lại, như soạn đề thi, soạn bài giảng, giao bài tập... Ứng dụng dữ liệu lớn để xử lý dữ liệu lớn với vài trăm nghìn đề thi, bài giảng, gợi ý nội dung liên quan bằng học máy (machine learning) và trí tuệ nhân tạo (AI), cho phép chấm thi trên giấy tự động, tiết kiệm 90% công chấm bài của giáo viên.

Đối với học viên/sinh viên: dữ liệu lớn cho phép cá nhân hóa việc học của từng học sinh để đạt được chuẩn đầu ra cần thiết; tạo điều kiện cho việc lựa chọn nghề nghiệp và nâng cao cơ hội việc làm sau khi tốt nghiệp của sinh viên... Với dữ liệu lớn, có thể xác định kỹ thuật nào là hiệu quả nhất cho từng người học. Điều đó thật tuyệt vời khi cá nhân hóa trong giáo dục giúp nâng cấp kỹ năng viết, khả năng giao tiếp hoặc đối phó với bài kiểm tra ở các môn học khác nhau của người học nhanh hơn. Dữ liệu lớn giúp thiết kế hệ thống dạy kèm thông minh và tương tác, phù hợp với nhu cầu và điểm yếu cá nhân của người học. Việc tạo ra những trải nghiệm học tập thú vị thông qua các giải pháp kỹ thuật số được thiết kế riêng hiện đã trở thành hiện thực và đáng được các nhà giáo dục quan tâm.

Đối với cơ sở giáo dục - đào tạo: dữ liệu lớn giúp sửa đổi cơ cấu tổ chức của cơ sở; tạo điều kiện thuận lợi cho quá trình ra quyết định, giúp cải tiến các khóa học, phân loại học sinh, tính được mức độ thành công của học viên/sinh viên...

Thứ hai, cần xây dựng hệ thống đánh giá quản lý giáo dục dữ liệu lớn. Tài nguyên dữ liệu giáo dục là nền tảng của giáo dục thông minh. Chỉ khi thiết lập một cơ chế đánh giá khoa học, chúng ta mới có thể thúc đẩy “nhất thể hóa” từ thu thập dữ liệu đến sử dụng dữ liệu, tạo nên sự phát triển tuần hoàn của giáo dục thông minh.

Thứ ba, đẩy mạnh việc bồi dưỡng đội ngũ quản lý giáo dục dữ liệu lớn. Cần tăng cường đào tạo nhân lực chuyên nghiệp, thiết lập và cải tiến hệ thống nhân sự dữ liệu lớn đa cấp và đa loại hình hỗ trợ trực tiếp cho chiến lược dữ liệu lớn quốc gia trong tương lai. Đội ngũ này là trung tâm quản trị dữ liệu lớn, chủ thể phát triển và đối tượng nghiên cứu của dữ liệu lớn.

Thứ tư, khắc phục những khó khăn, thách thức nảy sinh trong quá trình ứng dụng dữ liệu lớn vào lĩnh vực giáo dục. Về theo dõi dữ liệu, thu thập dữ liệu, phân tích dữ liệu, kết nối với

khoa học học tập, tối ưu hóa môi trường học tập, công nghệ mới nổi và các mối quan tâm đạo đức liên quan đến các vấn đề pháp lý và quyền riêng tư. Đây là các vấn đề lớn đang sau dữ liệu lớn trong giáo dục hiện nay.

Các vấn đề về lưu trữ và khả năng mở rộng: Khối lượng dữ liệu được tạo ra vượt quá khả năng xử lý của các công cụ dữ liệu lớn hiện có thể truy cập. Điều này có thể gây ra các sự cố nghiêm trọng khiến hệ thống gặp sự cố hoặc hoạt động chậm lại, dẫn đến trải nghiệm tiêu cực và giảm chất lượng phân tích.

Thứ năm, thường xuyên đầu tư, bảo trì, nâng cấp hạ tầng kỹ thuật công nghệ hiện đại, đáp ứng được các tiêu chuẩn kỹ thuật về lưu trữ, xử lý, bảo đảm an toàn, bảo mật dữ liệu. □

⁽¹⁾ Higher Education Commission, *Form bricks to clicks: The potential of data and analytics in higher education*, United Kingdom, January 2016, <https://www.policyconnect.org.uk/media/1128/download>, p.18

TÀI LIỆU THAM KHẢO

1. Bernard Marr, *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=7d13837360ba>
2. *How Big Data Is Influencing the Education*, <https://bigdataanalyticsnews.com/big-data-influencing-education/>
3. *Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review*, *Online Learning - Volume 20 Issue 2 - June 2016*, P.13-29
4. *Implications of the Data Revolution for Statistics Education*, *International Statistical Review* (2016), 84, 3, P.528-549
5. *The Evolution of Big Data and Learning Analytics in American Higher Education*, *Journal of Asynchronous Learning Networks*, Volume 16: Issue 3, P.9-20
6. Katrina Sin, Loganathan Muthu: *Application of Big Data in Education Data Mining and Learning Analytics - A Literature Review*, *Ictact Journal on soft computing: Special Issue on Soft Computing Models for Big Data*, July 2015, Volume: 05, Issue: 04