

TỔNG QUAN VỀ PHÂN TÍCH DỮ LIỆU LỚN TRONG GIÁO DỤC ĐẠI HỌC

AN OVERVIEW OF BIG DATA ANALYTICS IN HIGHER EDUCATION SECTOR

Mai Anh Thơ

Trường Đại học Nông Lâm TP.HCM, Việt Nam

Ngày toà soạn nhận bài 14/5/2018, ngày phản biện đánh giá 24/5/2018, ngày chấp nhận đăng 15/6/2018.

TÓM TẮT

Trong bối cảnh nhu cầu về đảm bảo chất lượng tại các cơ sở giáo dục đại học Việt Nam đang được quan tâm từ các bên liên quan, việc tìm ra các biện pháp cụ thể và hiệu quả để đảm bảo và nâng cao chất lượng là rất cấp thiết. Để đạt được điều này, phân tích dữ liệu giáo dục được xem là một giải pháp khá hữu hiệu. Bài viết được thực hiện nhằm đưa ra cái nhìn bao quát về dữ liệu lớn trong ngữ cảnh của giáo dục đại học, đặc biệt tại Việt Nam như là một mối quan tâm mới xuất hiện gần đây trong các nghiên cứu liên ngành. Bài viết cũng xác định rõ các loại dữ liệu cần được thu thập, lưu trữ và phân tích tại các cơ sở giáo dục đại học. Qua đó, bài viết nêu lên những vấn đề tồn tại và gợi ý những định hướng ban đầu cho việc ứng dụng lĩnh vực nghiên cứu này vào thực tiễn giáo dục đại học.

Từ khóa: *dữ liệu lớn; dữ liệu lớn trong giáo dục; phân tích dữ liệu lớn; giáo dục đại học; chất lượng giáo dục đại học.*

ABSTRACT

It has recently been an urgency to create specific solution for quality assurance in terms of education. Big data analytics came out as an efficient tool to enhance the quality of higher education. This article aims to present an overview of big data in the context of higher education, especially in Vietnam as an emerging interdisciplinary research field. The article then identifies the types of data that need to be collected, stored and analyzed at educational institutions, summarize data analytic methods that are currently being applied. Current situation of Vietnam higher education is also addressed as a background of some initial suggestions for applying big data analytics in this sector.

Keywords: *Big data; educational big data; big data analytics; higher education; quality of higher education.*

1. MỞ ĐẦU

Các cơ sở giáo dục đại học đang hoạt động trong môi trường cạnh tranh gay gắt dưới áp lực ngày càng tăng để đáp ứng với sự thay đổi của kinh tế, chính trị, xã hội và toàn cầu hóa. Việc đảm bảo tỉ lệ sinh viên đồng thời đảm bảo chất lượng đào tạo là mối quan tâm lớn hiện nay. Sự mong đợi ngày càng cao và khắt khe từ các bên liên quan cũng tạo áp lực cho nhà trường, trong điều kiện nguồn tài trợ từ chính phủ suy giảm, các yêu cầu về tính minh bạch và trách nhiệm giải trình ngày

càng cao. Trong bối cảnh đó, việc đưa ra các quyết định đúng đắn để ứng phó kịp thời với các vấn đề mà cơ sở giáo dục đang phải đối mặt là vô cùng quan trọng. Phân tích dữ liệu giáo dục, dựa trên tập dữ liệu khổng lồ đã có của tổ chức để phát hiện các thông tin tiềm ẩn, được xem là một công cụ hữu hiệu hỗ trợ cho quá trình ra quyết định.

Bài viết này trình bày các khái niệm cơ bản về dữ liệu lớn và phân tích dữ liệu lớn trong giáo dục đại học. Các cơ sở giáo dục cần xác định rõ các loại dữ liệu cần được thu

thập, lưu trữ và phân tích. Bài viết cũng tổng kết các phương pháp phân tích dữ liệu đang được áp dụng hiện nay, đồng thời trình bày chu trình 4 bước cơ bản để phân tích dữ liệu. Những lợi ích từ việc phân tích dữ liệu giáo dục cũng được tổng hợp từ các nghiên cứu để làm rõ sự cần thiết của lĩnh vực nghiên cứu này đối với các cơ sở giáo dục đại học. Sau cùng, bài viết tóm lược tình hình nghiên cứu và ứng dụng phân tích dữ liệu giáo dục tại Việt Nam, từ đó đưa ra những nhận định ban đầu nhằm chuẩn bị cho việc áp dụng phân tích dữ liệu lớn tại các cơ sở giáo dục đại học của Việt Nam.

2. DỮ LIỆU LỚN VÀ CÁC ĐẶC TRƯNG CỦA DỮ LIỆU LỚN

Dữ liệu được xác định là một trong những yếu tố chính dẫn tới sự thay đổi trong thế kỷ 21. Những mô tả như “cuộc cách mạng dữ liệu” hay “kỷ nguyên của dữ liệu lớn” hay đơn giản là “dữ liệu lớn” đã cho thấy sự gia tăng đáng kể số lượng dữ liệu mà chúng ta đang tạo ra hàng ngày trong tất cả các lĩnh vực của đời sống.

Khi nói đến dữ liệu lớn là đang nói đến khối lượng lớn dữ liệu, bao gồm cả dữ liệu có cấu trúc và không có cấu trúc mà chúng ta đang tạo ra và phải đối mặt hàng ngày. Dữ liệu lớn thường được xác định thông qua ba chữ V [1]:

Volume (Khối lượng dữ liệu): để nói về việc các tổ chức thu thập dữ liệu từ nhiều nguồn khác nhau.

Velocity (Tốc độ sản sinh dữ liệu): các luồng dữ liệu được sản sinh với tốc độ chưa từng thấy và cần được xử lý kịp thời.

Variety (Sự đa dạng của dữ liệu): dữ liệu có trong tất cả các loại định dạng - từ dữ liệu có cấu trúc, dữ liệu số trong cơ sở dữ liệu truyền thống cho đến các tài liệu văn bản không có cấu trúc, email, video, âm thanh, dữ liệu, mã cổ phiếu và các giao dịch tài chính.

Ngày nay, ngoài ba chữ V trên dữ liệu lớn được mở rộng thêm bốn chữ V [2]:

Veracity (Sự đáng tin cậy của dữ liệu): Dữ liệu có thể bị nhiễu và bất thường. Vì

vậy, dữ liệu cần được làm sạch và lọc nhiễu để có thể trở nên đáng tin cậy.

Validity (Tính hợp lệ của dữ liệu): chỉ những dữ liệu hợp lệ mới có thể dẫn đến các quyết định đúng. Do vậy, tính đúng đắn và chính xác của dữ liệu rất quan trọng.

Volatility (Sự biến động của dữ liệu): quá trình dữ liệu được tạo ra là quá trình liên tục. Một số dữ liệu có thể không còn có giá trị sau một thời điểm nhất định. Do vậy, dữ liệu cần được kiểm tra tính hợp lệ và xác định thời gian cần lưu trữ.

Value (Giá trị của dữ liệu): những tri thức có thể rút trích ra từ dữ liệu.

Chính sự ra đời và thống trị của Internet cùng với sự mở rộng các hệ thống số trên tất cả các lĩnh vực làm phát sinh khối lượng lớn các dữ liệu phi cấu trúc bên cạnh các dữ liệu có cấu trúc truyền thống. Việc thu thập, phân tích dữ liệu lớn để phát hiện ra các tri thức tiềm ẩn bên trong dữ liệu là rất quan trọng, giúp hỗ trợ quá trình ra quyết định. Ngày nay, với sự phát triển nhanh chóng của công nghệ và công cụ, khả năng phân tích các dữ liệu phức tạp đã và đang làm thay đổi cách thức dự đoán và ra quyết định, đem lại các tiềm năng to lớn trong cạnh tranh trí tuệ.

3. DỮ LIỆU LỚN TRONG GIÁO DỤC ĐẠI HỌC

3.1. Khái niệm

Trong bối cảnh giáo dục tạo đại học, dữ liệu lớn bao hàm quy trình xử lý dữ liệu quản lý và vận hành để đánh giá hoạt động và quá trình phát triển của một tổ chức giáo dục nhằm dự báo hoạt động trong tương lai và xác định những vấn đề tiềm tàng liên quan đến hoạt động xây dựng chương trình đào tạo, nghiên cứu, giảng dạy và học tập [3]. Các quan điểm khác cho rằng để đáp ứng nhu cầu cải tiến hiệu quả hoạt động, giáo dục đại học phải áp dụng công cụ phân tích dữ liệu vào hệ thống. Dù là một lĩnh vực mới xuất hiện gần đây trong giáo dục, một số học giả đã khẳng định rằng phân tích dữ liệu lớn là rất thích hợp để giải quyết các thách thức chính mà giáo dục đại học hiện nay đang phải đối mặt [4].

3.2. Các loại dữ liệu trong giáo dục đại học

Dữ liệu lớn trong giáo dục đại học bao gồm cả các hệ thống cơ sở dữ liệu lưu trữ khối lượng lớn dữ liệu có cấu trúc cũng như dữ liệu thu thập được từ những tương tác và hoạt động cụ thể trong quá trình nghiên cứu,

giảng dạy và học tập, được chia làm 2 loại [5]:

Static data (Dữ liệu tĩnh): dữ liệu có cấu trúc chứa các thông tin mà các tổ chức giáo dục đang thu thập và lưu trữ trong các hệ quản trị cơ sở dữ liệu, bao gồm:

Bảng 1. Dữ liệu giáo dục có cấu trúc

TT	Loại dữ liệu	Mô tả
1	Dữ liệu tuyển sinh và nhập học	Số lượng hồ sơ đăng ký, số lượng nhập học, điểm trúng tuyển...
2	Dữ liệu người học	Bao gồm: tên, tuổi, trường cấp 3, quá trình học tập, thông tin liên lạc...
3	Dữ liệu về đội ngũ	Số lượng giảng viên, nhân viên, hồ sơ đội ngũ...
4	Dữ liệu tài chính	Thu nhập, chi tiêu, ngân sách, thuế, lợi nhuận...
5	Dữ liệu cựu người học	Địa chỉ, thông tin liên hệ hiện tại, vị trí công việc, nơi làm việc...
6	Kế hoạch giảng dạy, học tập và nghiên cứu	Dữ liệu về danh sách các môn học mở trong từng học kỳ, phân công giảng viên, số sinh viên đăng ký, danh sách phân bổ đề tài...
7	Cơ sở vật chất	Số phòng học, phòng thí nghiệm, trang thiết bị, phần mềm, thư viện, ký túc xá...

Fluid data (tạm dịch là Dữ liệu thể “lỏng”): Trong ngữ cảnh của giáo dục đại học, fluid data có nội hàm là các loại dữ liệu phi cấu trúc, được sinh ra khi giảng viên, người học tương tác với các hệ thống học trực tuyến và các ứng dụng công nghệ trong nhà trường. Đây được gọi là “dấu chân kỹ thuật số” (digital footprint). Khi giảng viên, sinh viên tương tác với các công nghệ học tập, họ để lại những dấu vết dữ liệu mà có thể tiết lộ tình cảm, kết nối xã hội, ý định và mục tiêu của họ. Các nhà nghiên cứu có thể sử dụng dữ liệu này để phát hiện các vấn đề tiềm tàng của giảng viên và người học theo thời gian, từ học kỳ này sang học kỳ khác hoặc từ năm này sang năm khác. Trong lúc chưa tìm được thuật ngữ chính xác diễn tả đầy đủ ý nghĩa của loại dữ liệu này thì người nghiên cứu xin dùng khái niệm fluid data với nội hàm như trên. Loại dữ liệu đa dạng và sản sinh liên tục này, sẽ là nguồn dữ liệu lớn quan trọng mà các cơ sở giáo dục đại học cần khai thác thông qua việc phân tích dữ liệu lớn.

4. PHÂN TÍCH DỮ LIỆU TRONG GIÁO DỤC ĐẠI HỌC

4.1. Khái niệm

Theo Educause, phân tích dữ liệu được định nghĩa một cách tổng quát là dựa vào dữ liệu để ra quyết định [6]. Dữ liệu được tập hợp lại, sau đó được chuyển thành các thông tin có ích. Thông qua việc phân tích dữ liệu trong giáo dục đại học, cả cơ sở giáo dục và người học sẽ có thể thực hiện một số quyết định đúng đắn dựa trên dữ liệu [7]. Phân tích dữ liệu đem đến sự đổi mới dưới 2 hình thức: (a) Sustainable (Bền vững): giúp cải tiến hệ thống đang có và (b) Disruptive (Hỗn loạn): tạo ra hệ thống mới hoàn toàn khác với hệ thống đang có.

Có 5 cấp độ phân tích dữ liệu: (1) *Cá nhân*: phân tích hiệu suất cá nhân dưới góc độ mục tiêu và thành tựu; (2) *Môn học*: liên quan đến phát triển khái niệm; (3) *Khoa/Bộ môn*: liên quan đến các mô hình dự đoán, các

mẫu thành công hay thất bại; (4) *Cơ sở giáo dục*: liên quan đến nguồn lực phân bổ và hiệu quả hoạt động; (5) *Cấp vùng*: liên quan đến đối sánh giữa các hệ thống và tiêu chuẩn [8].

Phân tích dữ liệu trong giáo dục đại học có thể được phân thành 2 loại chính như sau:

Learning Analytics (Phân tích dữ liệu học tập): giúp người học đưa ra các quyết định cần thiết để đạt được những kết quả mong đợi cụ thể, giúp gia tăng sự thành công của người học. Thông qua việc phân tích này, các mẫu hành vi của người học có thể được nhận diện. Giảng viên có thể dựa vào các mẫu hành vi để xác định các sinh viên có nguy cơ, từ đó hỗ trợ cơ sở giáo dục các biện pháp ngăn ngừa thích hợp. Hiện nay, với số lượng sinh viên bỏ học, hoặc bị buộc thôi học ngày càng gia tăng, thì phân tích dữ liệu người học thực sự là công cụ hữu ích để cải thiện vấn đề này.

Academic Analytics (Phân tích dữ liệu quản lý): giúp cho việc ra quyết định liên quan đến hoạt động quản lý và các hoạt động khác của cơ sở giáo dục. Phân tích này được định nghĩa là sự kết hợp của thông tin, công nghệ, văn hóa và quản trị cơ sở giáo dục [9]. Phân tích các dữ liệu liên quan đến các quyết định bên trong của nhà trường như: tuyển sinh và nhập học, tài chính, quản lý hành chính, hội đồng trường... giúp cải thiện hiệu quả hoạt động của cơ sở giáo dục.

Phân tích dữ liệu học tập đem lại lợi ích cho người học, đội ngũ giảng viên, hoạt động giảng dạy, học tập và nghiên cứu, trong khi phân tích dữ liệu quản lý đem lại lợi ích cho các nhà quản trị, nhà tài trợ, nhà tiếp thị, các doanh nghiệp, chính phủ và các tổ chức lớn khác. Để cải thiện giáo dục đại học, cần phân tích cả dữ liệu học tập và dữ liệu quản lý.

4.2. Các phương pháp phân tích dữ liệu

Có rất nhiều phương pháp phân tích dữ liệu giáo dục khác nhau nhằm khám phá ra các dữ liệu tiềm tàng thú vị ẩn bên trong các tập dữ liệu giáo dục. Các phương pháp này được phân thành 2 loại chính: Phân tích định lượng và Phân tích định tính.

Phân tích định lượng: liên quan đến các mô hình tính toán và thống kê, bao gồm các phương pháp sau:

- **Phân tích thống kê**: Bằng việc áp dụng các số liệu thống kê và các phương pháp toán học, kiến thức có thể được phát hiện ra từ dữ liệu. Phân tích thống kê có liên quan đến các dấu vết mà tính toán số có thể thực hiện được như cho phép đếm số lượt truy cập, phân tích số lần nhấp chuột và tính toán thời gian dành cho công việc. Một số công cụ phân tích thống kê có thể sử dụng như IBM SPSS và MATLAB.
- **Trực quan hóa (Visualization)**: Thông tin thống kê có thể diễn giải thành biểu đồ, sơ đồ, lập bản đồ mindmap, heat map, 3D plots, scatterplots, sự đánh giá các mô hình và sơ đồ. Các kỹ thuật hiển thị thông tin trực quan này, thể hiện được rõ ràng những gì có thể đo lường một cách chính xác để có cái nhìn sâu sắc hơn về quá trình học tập [10].
- **Phân tích mạng xã hội định lượng**: Social Network Analysis (SNA) có thể sử dụng dữ liệu môi trường học tập để trợ giúp giảng viên hiểu được bầu không khí của lớp học của mình và trợ giúp cho sinh viên khi cần thiết.

Phân tích định tính: Bằng phương pháp định tính, các khuyến nghị có thể được cung cấp cho những người học dựa trên hoạt động học tập. Khảo sát của Khalil và Ebner [11] thu được hai phương pháp phân tích định tính chính:

- **Trí tuệ cảm xúc**: dựa trên cảm xúc, liên quan đến tâm lý học và xã hội học.
- **Phân tích mạng xã hội định tính**: bằng cách quan sát, phỏng vấn và khảo sát, các dữ liệu định tính sau đó được phân tích nhờ vào việc thiết lập một mạng xã hội để đưa vào một ngữ cảnh rộng hơn.

Dù đi theo phương pháp phân tích dữ liệu nào, việc phân tích dữ liệu cũng được thực hiện dựa trên chu trình mà Khalil và Ebner đưa ra gồm 4 bước như sau:

- **Tạo ra dữ liệu:** dữ liệu được tạo ra từ các bên liên quan trong các môi trường học tập trực tuyến, các ứng dụng công nghệ.
- **Lưu trữ dữ liệu: người học thông qua các hệ thống học tập, đã để lại nhiều dữ liệu có thể** theo dõi được. Người học không chỉ là người sử dụng dữ liệu mà còn là người tạo ra dữ liệu.
- **Phân tích dữ liệu:** các phương pháp phân tích định tính và định lượng khác nhau được áp dụng để khám phá các mẫu ẩn bên trong bộ dữ liệu giáo dục.
- **Hành động:** kết quả phân tích phải được diễn giải thành hành động cụ thể như: dự báo, can thiệp, khuyến nghị, cá nhân hóa và phản hồi.

Chu trình này sẽ được lặp lại một cách tối ưu, tương tự như nhận định của Campbell và Oblinger đó là có những điểm tương đồng trong các giai đoạn giữa các phương pháp phân tích dữ liệu [12]. Phân tích dữ liệu là một vòng lặp mở của các giai đoạn nên được kết thúc vào cuối chu trình bằng cách đem lại sự tối ưu hóa cho môi trường học tập và các bên liên quan.

4.3. Những lợi ích của phân tích dữ liệu giáo dục

Theo [13], nhóm tác giả đã tổng kết 112 bài báo nghiên cứu về phân tích dữ liệu từ các cơ sở dữ liệu tin cậy như Google Scholar, Educational Resources Information Center (ERIC), ProQuest, và EBSCO HOST, và đưa ra các lợi ích của phân tích dữ liệu giáo dục như sau:

- **Xác định các học phần mục tiêu:** Bằng việc kiểm tra xu hướng chọn ngành xét tuyển của sinh viên và lợi ích trong các ngành khác nhau, cơ sở giáo dục có thể xác định các học phần mục tiêu phù hợp hơn với nhu cầu và sở thích của người học, tập trung nguồn lực giáo dục và dạy học vào các chương trình có đông sinh viên đăng ký nhất. Nhà trường cũng có thể dự đoán số lượng sinh viên tốt nghiệp để xây dựng kế hoạch tuyển sinh dài hạn hơn [14].

- **Cải tiến chương trình giảng dạy:** Thông qua việc phân tích dữ liệu lớn có thể giúp cho giảng viên thay đổi và điều chỉnh để cải tiến chương trình, tối ưu hóa tiềm năng học tập của sinh viên [15].
- **Quan tâm đến kết quả học tập của người học, hành vi và quá trình:** Phân tích dữ liệu có thể giúp các nhà giáo dục hiểu được trải nghiệm học tập của người học thông qua các tương tác của người học với các công cụ công nghệ như e-learning và mobile-learning. Sử dụng dữ liệu lớn cũng cho thấy hành vi học tập, tác động đến việc học tập thích ứng và mức độ kiên trì trong quá trình học tập. Bằng việc hiểu các yếu tố tác động đến kết quả học tập của người học, giảng viên có thể đánh giá các loại kiến thức và điều chỉnh nội dung giáo dục cho phù hợp.
- **Cá nhân hóa học tập:** Phân tích dữ liệu lớn có thể giúp phát triển các khóa học mang tính cá nhân hóa, phù hợp với khả năng học tập của từng sinh viên. Cá nhân hóa có thể mang lại một cuộc cách mạng quan trọng trong các hệ thống e-learning và ứng biên kết quả [16]. Điều này đảm bảo tỉ lệ thành công của lớp học và sự tham gia tích cực hơn của người học [17].
- **Cải thiện hiệu suất làm việc của giảng viên:** Thông qua việc phân tích dữ liệu được tạo ra khi sử dụng công cụ hướng dẫn của công nghệ và các công cụ nghiên cứu từ thư viện trực tuyến [18], các nhà phân tích có thể xác định hành vi trực tuyến của các giảng viên. Qua đó, có thể xác định hiệu quả làm việc cũng như xác định các điểm cần cải thiện của giảng viên, giúp giảng viên có cơ hội chuẩn bị tốt hơn khi tương tác với sinh viên trong môi trường học tập công nghệ nhằm nâng cao hiệu quả trong hoạt động dạy học.
- **Cơ hội việc làm, học tập nâng cao sau tốt nghiệp:** Sử dụng dữ liệu lớn cho phép các cơ sở giáo dục xác định các cơ hội việc làm được đăng tải cho sinh viên tốt nghiệp và giúp mục tiêu giáo dục phù hợp hơn với nhu cầu thị trường việc làm. Nhà

trường cũng có thể dự đoán khả năng có việc làm hay thất nghiệp, hoặc chưa xác định tình huống về cơ hội việc làm [19]. Sử dụng dữ liệu lớn có thể giúp các bên liên quan trong hệ thống giáo dục hiểu rõ hơn về triển vọng nghề nghiệp và đánh giá tốt hơn khả năng thích ứng với nghề nghiệp của chương trình học. Trong môi trường học tập toàn cầu, thông tin này không chỉ tạo thuận lợi cho việc lập kế hoạch giáo dục đại học và sau đại học tốt hơn, mà còn hữu ích cho các tổ chức khi họ ra quyết định tuyển dụng và đầu tư ngân sách cho sinh viên tốt nghiệp các ngành nghề khác nhau.

- *Cộng đồng nghiên cứu và thực hiện phân tích dữ liệu:* Cộng đồng nghiên cứu cũng được hưởng lợi từ việc sử dụng dữ liệu lớn trong giáo dục. Các nhà nghiên cứu có thể dễ dàng chia sẻ thông tin và hợp tác. Họ có thể xác định những khoảng trống giữa thị trường lao động và cơ sở giáo dục từ đó nghiên cứu để giải quyết vấn đề.

Từ các lợi ích trên cho thấy lĩnh vực phân tích dữ liệu giáo dục là thật sự cần thiết và đem lại hiệu quả thiết thực hỗ trợ nâng cao chất lượng cho các cơ sở giáo dục.

5. PHÂN TÍCH DỮ LIỆU TRONG GIÁO DỤC ĐẠI HỌC TẠI VIỆT NAM

Phân tích dữ liệu giáo dục tại Việt Nam đang được quan tâm trong thời gian gần đây, cùng với sự ảnh hưởng của cuộc cách mạng công nghiệp 4.0 kéo theo những tác động đến các cơ sở giáo dục, đặc biệt là giáo dục đại học. Hội thảo InSITE 2017 do đại học RMIT Việt Nam tổ chức vào giữa tháng 08/2017 thu hút hơn 150 nhà giáo dục trong nước và quốc tế đã khẳng định vai trò quan trọng của phân tích dữ liệu người học trong giáo dục. Đây là phương pháp hữu dụng đối với giảng viên, người làm công tác giáo dục, những người soạn thảo chương trình học và các nhà nghiên cứu giáo dục, vì giúp hé lộ năng lực tiềm ẩn cũng như biểu đồ phức tạp thể hiện hành vi học tập của sinh viên [20]. Đại học FPT cũng đã đưa ra những bài toán cơ bản của Big Data trong lĩnh vực giáo dục như: cá

nhân hóa giáo dục trực tuyến, đào tạo các nhà khoa học dữ liệu hay ứng dụng giúp phát hiện gian lận báo cáo khoa học [21]. Bài viết của giáo sư Hồ Tú Bảo tại báo Tia sáng tháng 07/2017 cũng khẳng định: các nước phát triển như Mỹ, Đức, Pháp, Hàn Quốc, Trung Quốc, Nhật Bản đều đã xây dựng chương trình chiến lược quốc gia với cốt lõi là câu chuyện của số hóa, kết nối và phân tích dữ liệu lớn. Quốc gia nào làm chủ nguồn dữ liệu lớn và phức tạp này, đưa chúng vào mọi lĩnh vực của sản xuất và cuộc sống, làm cho sản xuất và cuộc sống thông minh và hiệu quả hơn, sẽ là quốc gia thắng cuộc trong cách mạng công nghiệp 4.0 này [22].

Một số đại học trong nước đã quan tâm đến khoa học dữ liệu, khai phá dữ liệu áp dụng vào nhiều lĩnh vực khác nhau trong đời sống như: khai phá dữ liệu từ ngân hàng dữ liệu virus cúm; ứng dụng khai phá dữ liệu xây dựng hệ thống tổng hợp thông tin thị trường chứng khoán Việt Nam; xây dựng mô hình và hệ thống dự báo dịch tả khu vực Hà Nội dựa trên kỹ thuật khai phá dữ liệu...

Riêng đối với lĩnh vực giáo dục, khai phá dữ liệu giáo dục, lĩnh vực nghiên cứu gần với phân tích dữ liệu giáo dục đã bước đầu có các nghiên cứu rải rác tại các trường: Đại học Cần Thơ với ứng dụng các kỹ thuật trong khai phá dữ liệu để hỗ trợ sinh viên lập kế hoạch học tập [23]; phát hiện môn học quan trọng ảnh hưởng đến kết quả học tập của sinh viên ngành CNTT [24]; hệ thống dự báo năng lực học tập và hỗ trợ sinh viên lựa chọn môn học [25][26] [27]. Đại học Đà Nẵng với ứng dụng khai phá dữ liệu xây dựng công cụ dự đoán kết quả học tập của sinh viên [28]; Đại học Bách Khoa TP.HCM với áp dụng kỹ thuật khai phá dữ liệu để phân loại dữ liệu giáo dục [29].

Những thông tin trên cho thấy các cơ sở giáo dục đại học Việt Nam đã bắt đầu nhận thức được tầm quan trọng của việc khai phá dữ liệu giáo dục, tuy nhiên các nghiên cứu trong lĩnh vực này cũng còn giới hạn về số lượng cũng như phạm vi ứng dụng. Các nghiên cứu hiện nay dừng lại ở việc sử dụng các kỹ thuật khai phá dữ liệu để phát hiện các

vấn đề từ dữ liệu của người học, và đưa ra các dự đoán, hỗ trợ cho người học trong từng bài toán nhỏ, rời rạc, tại một đơn vị cụ thể. Các nghiên cứu chủ yếu tập trung vào các giải thuật khai phá dữ liệu dưới góc nhìn của khoa học máy tính, bài toán giáo dục chỉ là minh họa cho việc áp dụng giải thuật. Qua quá trình tìm hiểu, tác giả chưa tìm thấy nghiên cứu nào đề cập đến phân tích dữ liệu lớn dưới góc nhìn tổng quát về những lợi ích và tác động mà phân tích dữ liệu lớn giáo dục mang lại cho công tác đảm bảo và nâng cao chất lượng của các cơ sở giáo dục đại học tại Việt Nam.

Hiện nay, giải pháp học tập trực tuyến lớn nhất được 72% các trường xếp hạng cao của thế giới sử dụng là Blackboard đã tích hợp các công cụ phân tích dữ liệu như: Blackboard Analytics for Learn, Blackboard Predict, X-Ray Learning Analytics [30]. Tại Việt Nam đã có một số trường đại học bắt đầu áp dụng giải pháp của Blackboard như: Đại học Quốc tế thuộc Đại học Quốc gia TP.HCM, Đại học Sư Phạm Kỹ Thuật TP.HCM, Đại học Giao thông Vận tải... Tuy vậy, các trường nếu không hiểu rõ ý nghĩa của việc phân tích dữ liệu giáo dục, cũng như không đề ra được chính sách thích hợp để có đủ nguồn dữ liệu chất lượng, sẽ không thể khai thác hiệu quả các công cụ mà Blackboard hỗ trợ. Việc tích hợp các công cụ phân tích dữ liệu vào các giải pháp giáo dục của cổng thông tin trực tuyến lớn hỗ trợ công nghệ dạy học như BlackBoard càng khẳng định tầm quan trọng và sự cần thiết của phân tích dữ liệu giáo dục.

Do vậy, các cơ sở giáo dục Việt Nam cần đặc biệt đầu tư và đẩy mạnh nghiên cứu, ứng dụng phân tích dữ liệu giáo dục trong thời gian tới bởi lẽ nếu không ý thức rõ và bắt kịp xu thế, lợi thế của việc nắm giữ và khai thác dữ liệu lớn giáo dục, các cơ sở giáo dục đại học sẽ có nguy cơ tụt hậu ngay trên sân nhà.

6. KẾT LUẬN

Phân tích dữ liệu lớn trong giáo dục là một xu thế mà hầu hết các trường đại học chất lượng trên thế giới đã và đang tiếp tục

triển khai vì các lợi ích thiết thực mà nó đem lại như đã tổng kết trong mục 2.5.

Tại Việt Nam, để có thể nghiên cứu và triển khai phân tích dữ liệu lớn trong giáo dục đại học, các cơ sở giáo dục trước hết cần có cơ chế quản lý một cách hệ thống và đồng bộ các loại dữ liệu đã trình bày ở mục 2.2. Bên cạnh các dữ liệu tĩnh truyền thống, được lưu trữ trong các hệ quản trị cơ sở dữ liệu, nhà trường cần triển khai và lưu trữ dữ liệu từ các hệ thống hỗ trợ học tập áp dụng các công nghệ dạy học và các ứng dụng trực tuyến, các diễn đàn, các trang mạng xã hội, website, các thiết bị thu thập thông tin... nhằm có được nguồn dữ liệu lớn quý giá để tiến hành phân tích. Đây cũng là thách thức đối với nhiều cơ sở giáo dục đại học Việt Nam, bởi lẽ, việc áp dụng các hệ thống học tập trực tuyến hiện nay ở các trường còn khá hạn chế, và việc triển khai tại một số trường cũng chưa mang tính hiệu quả cao. Đó là lý do tại sao các nghiên cứu và công bố về việc phân tích và khai thác các loại dữ liệu này nhằm hỗ trợ cho công tác quản trị, ra quyết định và cải thiện chất lượng trong nhà trường tại Việt Nam còn rất khiêm tốn.

Bên cạnh đó, phân tích dữ liệu giáo dục là lĩnh vực nghiên cứu liên ngành gồm Công nghệ giáo dục, Thống kê, Toán học, Khoa học máy tính và Khoa học thông tin. Do vậy, để nghiên cứu và triển khai phân tích dữ liệu thành công, các cơ sở giáo dục cần xây dựng được đội ngũ các chuyên gia có kinh nghiệm trong các lĩnh vực trên cùng phối hợp thực hiện.

Việc sẵn sàng về nguồn dữ liệu và đội ngũ là bước đầu để các trường đại học Việt Nam có thể tiếp cận và triển khai phân tích dữ liệu lớn giáo dục. Các nghiên cứu của thế giới cũng chỉ ra rất nhiều các vấn đề khác cần quan tâm trong phân tích dữ liệu giáo dục. Không phải là quốc gia tiên phong trong lĩnh vực nghiên cứu này, các cơ sở giáo dục của Việt Nam cần nhanh chóng tìm hiểu, học hỏi kinh nghiệm, nắm bắt các cơ hội và thách thức để có thể triển khai và thu nhận các lợi ích thiết thực nhằm nâng cao chất lượng giáo dục đại học.

TÀI LIỆU THAM KHẢO

- [1] SAS. (Jan 07). *Big Data: What it is and why it matters*. Available at: www.sas.com/en_us/insights/big-data/what-is-big-data.html
- [2] R. J. Vatsala*, Sathyaraj R***, "A Review of Big Data Analytics in Sector of Higher Education," *Journal of Engineering Research and Application*, vol. 7, no. 6, p. 7, 2017.
- [3] A. G. Picciano, "The Evolution of Big Data and Learning Analytics in American Higher Education," *Journal of Asynchronous Learning Networks*, vol. 16, no. 3, p. 12, Jun 2012.
- [4] G. Siemens. (2011, Feb 07). *How Data and Analytics Can Improve Education*. Available at: <http://www.oreilly.com/ideas/education-data-analytics-learning>
- [5] X. ShackLock, "From Brick to Clicks: The Potential of Data and Analytics in Higher Education," Higher Education Commission, January 2016.
- [6] K. E. A. Angela Van Barneveld, John P. Campbell, "Analytics in Higher Education: Establishing a Common Language," January 2012. Educause Learning Initiative
- [7] J. M. K. J Meenakurami, "Learning Analytics and Its Challenges in Education Sector a Survey," *International Journal of Computer Applications*, 0975-8887, p. 5, 2015.
- [8] T. Bomatpalli, "Significance of Big Data and Analytics in Higher Education," *International Journal of Computer Applications*, 0975-8887, vol. 68, no. 14, p. 3, April 2013.
- [9] P. J. Goldstein, "Academic Analytics in The Education Sector: Needs, Opportunities and Challenges," *International Journal of Research in Computer and Communication Technology*, 2014.
- [10] E. Duval, "Attention please! Learning analytics for visualization and recommendation," presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK 11), Banff, AB, Canada, February 27 - March 01, 2011.
- [11] M. E. Mohammad Khalil, "Learning Analytics: Principles and Constraints," presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, June 2015.
- [12] D. G. O. John P. Campbell, "Academic Analytics," EDUCAUSE Publications October 15, 2007. Available at: <https://library.educause.edu/resources/2007/10/academic-analytics>
- [13] J. T. A. Sandra Nunn, Therese Kanai, Mansureh Kebritchi, " Learning Analytics Methods, Benefits, and Chanllenges in Higher Education: A Systematic Literature Review," *Online Learning*, vol. 20, no. 2, p. 17, 2016.
- [14] M. A. Alaa Althubaiti, "Medical Colleges in Saudi Arabia: Can We Predict Graduate Numbers?," *Higher Education Studies*, vol. 4, no. 3, p. 8, March 28, 2014. Canadian Center of Science and Education
- [15] P. Graciela M. Armayor, Sean T. Leonard, "Graphic Strategies for Analyzing and Interpreting Curricular Mapping Data," *American Journal of Pharmaceutical Education*, vol. 75, no. 5, p. 10, June 15, 2010.
- [16] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges," *British Journal of Educational Technology*, 2014. British Educational Research Association
- [17] N. M. Ana Borray. (2017, April 15, 2018). *Can Making College Completion Paths More Personalized Increase Student Success?* Available at: <https://er.educause.edu/blogs/2017/3/can-making-college-completion-paths-more-personalized-increase-student-success>
- [18] M. M. R. Beijie Xu, "Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data," *Educational Technology & Society*, vol. 15, no. 3, 2012. International Forum of Educational Technology & Society
- [19] C.-F. T. Bangsuk Jantawan, "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment," *International Journal of Computer Science and Information Security*, vol. 11, no. 10, pp. 1-7, October 2013.

- [20] L. M. Thúy. (2017, Jan 15). *RMIT News: Đẩy mạnh phân tích dữ liệu về quá trình học*. Available at: <https://www.rmit.edu.vn/vi/tin-hoat-dong/day-manh-phan-tich-du-lieu-ve-qua-trinh-hoc>
- [21] V. Q. C. Nguyễn Việt Cường. (2017, Jan 15). *Những bài toán cơ bản của Big Data*. Available at: <https://tech.fpt.com.vn/nhung-bai-toan-co-ban-cua-big-data/>
- [22] H. T. Bảo. (2017, Jan 15). *Câu chuyện về Khoa học dữ liệu: Chỉ vài kẻ thắng cuộc*. Available at: <http://tiasang.com.vn/khoa-hoc-cong-nghe/Cau-chuyen-ve-khoa-hoc-du-lieu-chi-vai-ke-thang-cuoc-10789>
- [23] Nguyễn Chí Ngôn và Nguyễn Thái Nghe. *Hệ chuyên gia hỗ trợ sinh viên lập kế hoạch học tập (dựa trên phương pháp khai phá dữ liệu - data mining)*. Đề tài NCKH cấp trường. Đại học Cần Thơ, 2010.
- [24] Đỗ Thanh Nghị, Phạm Nguyên Khang, Nguyễn Minh Trung, Trịnh Trung Hưng, *Phát hiện môn học quan trọng ảnh hưởng đến kết quả học tập sinh viên ngành CNTT*, Tạp chí Khoa học Đại học Cần Thơ, 2014.
- [25] Huỳnh Lý Thanh Nhân, Nguyễn Thái Nghe. *Hệ thống dự đoán kết quả học tập và gợi ý lựa chọn môn học*. Kỷ yếu hội thảo quốc gia lần thứ XVI: Một số vấn đề chọn lọc của CNTT&TT, 2013, trang 110-118. NXB Khoa học kỹ thuật. ISBN: 987-604-67-0251-1
- [26] Nguyễn Thái Nghe. *Hệ thống dự báo năng lực học tập và hỗ trợ sinh viên lựa chọn môn học*. Đề tài NCKH cấp trường. Đại học Cần Thơ, 2013.
- [27] Nguyen Thai-Nghe. *An Introduction to Factorization Technique for Building Recommendation Systems*. Vol. 6/2013, pp. 44-53, Journal of Science - University of Da Lat, ISSN 0866-787X, 2013.
- [28] Nguyễn Thị Thanh Thủy, Nguyễn Trần Quốc Vinh. *Ứng dụng khai phá dữ liệu xây dựng công cụ dự đoán kết quả học tập của sinh viên*. Hội nghị sinh viên nghiên cứu khoa học lần thứ 8, Đại học Đà Nẵng, 2012.
- [29] Võ Thị Ngọc Châu, Nguyễn Hứa Phùng, *On Semi-supervised Learning with Sparse Data Handling for Educational Data Classification*, 4th International Conference FDSE, 2017.
- [30] B. Inc. (2018, May 15). Blackboard Analytics. Available at: <http://www.blackboard.com/education-analytics/>

Tác giả chịu trách nhiệm bài viết:

ThS. Mai Anh Thơ

Trường Đại học Nông Lâm Tp. HCM

Email: tho@hcmuaf.edu.vn