

Một số công cụ xử lý số liệu trong nghiên cứu khoa học, áp dụng cho sinh viên ngành kinh tế - xã hội

Lâm Thái Bảo Ngọc *

*Trường Đại học Tiền Giang

Received: 27/04/2023; Accepted: 06/05/2023; Published: 28/5/2023

Abstract: Scientific research is one of the tasks that play a very important role in training creative thinking ability, research capacity and is one of the “important - mandatory - necessary” measures to aim at improving the quality of life. high quality training, better meet the increasingly demanding needs of society. The article introduces the characteristics and compares the advantages and disadvantages of a number of data processing tools in scientific research such as SPSS, STATA, SAS, Eviews,... and big data applications major in scientific research at universities.

Keywords: Tool, data processing, scientific research, big data

1. Đặt vấn đề

Hiện nay, khi mà thế giới đang trở nên “phẳng” hơn thì vai trò của các trường đại học và chất lượng giáo dục tại các trường đại học càng trở nên quan trọng hơn bao giờ hết. Bất cứ một trường đại học nào cũng đều có hai nhiệm vụ chính và quan trọng nhất đó chính là: đào tạo và nghiên cứu khoa học. Đây là hai hoạt động có mối quan hệ hữu cơ, hai nhiệm vụ cơ bản chiến lược của nhà trường, trong đó việc giảng viên nhà trường tích cực tham gia các hoạt động nghiên cứu khoa học là một trong những biện pháp *quan trọng - bắt buộc - cần thiết* để hướng đến nâng cao chất lượng đào tạo, đáp ứng tốt hơn nhu cầu ngày càng khắt khe của xã hội. Với chức năng giảng dạy, đại học đào tạo những chuyên gia có kỹ năng cao và có văn hóa; với chức năng nghiên cứu, trường đại học là trung tâm sáng tạo ra tri thức mới và chuyển giao công nghệ cho nền kinh tế, do đó, cũng có thể xem các trường đại học là một phương tiện nối kết công dân trong một xã hội.

Trong nghiên cứu kinh tế - xã hội, việc xử lý số liệu bằng các phần mềm là bắt buộc với các đề tài nghiên cứu sử dụng phương pháp định lượng nhằm đơn giản hóa quá trình xử lý dữ liệu và chạy mô hình. Tuy nhiên, việc lựa chọn phần mềm nào cần phụ thuộc vào đặc điểm nguồn dữ liệu và khả năng sử dụng phần mềm của người nghiên cứu. Trong bài viết này, tác giả giới thiệu và so sánh một số công cụ xử lý số liệu nhằm hỗ trợ nghiên cứu viên, sinh viên tối đa trong việc tìm kiếm, lựa chọn và sử dụng các phần mềm phù hợp với đề tài nghiên cứu khoa học của mình.

Sử dụng thành thạo các công cụ xử lý số liệu là một phần quan trọng trong quá trình ứng dụng công

nghệ thông tin và chuyển đổi số trong quá trình giáo dục đại học. Đặc biệt, trong thời đại phát triển của Big Data (dữ liệu lớn) thì vai trò của người dạy (giảng viên) sẽ phải nâng cấp hơn khi bổ sung thêm nội dung khai thác Big Data trong giảng dạy, nghiên cứu khoa học và ứng dụng vào thực tiễn cho sinh viên.

2. Nội dung nghiên cứu

2.1. Big Data

Big Data là các tập dữ liệu có khối lượng lớn và phức tạp. Độ lớn đến mức các phần mềm xử lý dữ liệu truyền thống không có khả năng thu thập, quản lý và xử lý dữ liệu trong một khoảng thời gian hợp lý.

Những tập dữ liệu lớn này có thể bao gồm các dữ liệu có cấu trúc, không có cấu trúc và bán cấu trúc, mỗi tập có thể được khai thác để tìm hiểu insights.

Các giải pháp Big Data cung cấp các công cụ, phương pháp và công nghệ được sử dụng để nắm bắt, lưu trữ, tìm kiếm và phân tích dữ liệu trong vài giây để tìm mối quan hệ và hiểu biết về cải tiến và lợi ích cạnh tranh mà trước đây không có. [2] [4]

Thực tế là các kho dữ liệu hiện tại đã là rất lớn và chúng còn đang tăng trưởng một cách nhanh chóng theo cấp số mũ. Các nguồn dữ liệu này không chỉ xuất phát từ con người (như nguồn thu thập từ các mạng xã hội) mà còn từ máy tính (các loại thông tin nhật ký được tạo bởi máy tính) và các vệ tinh (nguồn dữ liệu thu thập từ vệ tinh) được tạo ra liên tục dẫn đến vấn đề là làm thế nào chúng ta có thể giải quyết nguồn dữ liệu lớn và tăng trưởng liên tục như vậy. Hai yếu tố cần được quan tâm để giải quyết vấn đề này là:

Xử lý dữ liệu nhanh

Nguồn lưu trữ dữ liệu đáng tin cậy

Nhờ sự phát triển của Big Data mà ngày nay,

chúng ta dễ dàng kiểm tra mức độ đạo văn của các sản phẩm khoa học.

2.2. Phân tích dữ liệu

Phân tích dữ liệu (Data Analysis) là quá trình kiểm tra các tập dữ liệu để tìm ra xu hướng và đưa ra kết luận về thông tin mà chúng đang lưu giữ.

Tùy theo các ngành nghề và mục đích nghiên cứu mà quá trình phân tích dữ liệu được gọi theo nhiều tên gọi khác nhau: phân tích chẩn đoán, phân tích dự đoán, phân tích theo mô tả, phân tích thống kê, phân tích đề xuất, phân tích khám phá,... Tuy nhiên, về bản chất tất cả đều thuộc một trong hai nhóm: phân tích định tính hoặc phân tích định lượng.

- Phân tích định tính
- Phân tích định lượng

2.3. Ứng dụng của Big Data trong nghiên cứu khoa học tại trường đại học ^[2]

Big Data tạo ra các phương pháp tiếp cận dựa trên dữ liệu để dạy sinh viên.

Có thể lưu trữ, quản lý, phân tích các bộ dữ liệu lớn bao gồm hồ sơ của sinh viên.

Duy trì bảo mật bằng cách sử dụng hệ thống quản lý big data có khả năng trích xuất phân cấp.

Big Data giúp bảo mật các bộ đề kiểm tra, ngân hàng đề thi.

Big Data cung cấp dữ liệu về các hoạt động trong lớp và giúp đưa ra quyết định cho giảng viên hay người điều hành tổ chức.

Big Data có thể đánh giá biểu cảm khuôn mặt và di chuyển của học sinh trong lớp từ đó giảng viên có thể đánh giá được chất lượng giảng dạy bằng máy ảnh độ phân giải cao, cảnh quay video và xử lý hình ảnh.

Cung cấp nguồn dữ liệu thứ cấp khổng lồ cho công tác nghiên cứu khoa học.

Trong công tác nghiên cứu khoa học, Big Data mang lại các nhà nghiên cứu cơ hội tiếp cận và giải quyết rất nhiều vấn đề hóc búa mà trước đây không thể hoặc rất khó thực hiện.

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...

Bảng 2.1. So sánh một số phần mềm phân tích dữ liệu

TT	Phần mềm	Đặc điểm	Ưu điểm	Nhược điểm
1	SAS (Statistical Analysis System)	Sử dụng ngôn ngữ lập trình SAS để phân tích dữ liệu, cung cấp các biểu diễn trực quan cho người nghiên cứu để đánh giá.	- Sử dụng cho các phân tích nâng cao, phân tích kinh doanh và các tác vụ quản lý dữ liệu khác nhau. - Có thể làm việc nhiều dữ liệu, nhiều file một lúc. - Quản lý file dữ liệu lớn. - Công cụ vẽ đồ thị mạnh.	- Mất nhiều thời gian để hiểu cách quản lý dữ liệu trong SAS. - Không mạnh về hồi quy logistic và ước lượng.
2	SPSS (Statistical Package for the Social Sciences)	Giao diện giữa người và máy cho phép sử dụng các menu thả xuống để chọn các lệnh thực hiện.	- Có bộ soạn thảo dữ liệu tương tự như Excel và dễ sử dụng. - Được sử dụng phổ biến trong phân tích thống kê, phân tích tương tác và phân tích lô. - Mạnh trong phân tích phương sai, phân tích nhiều chiều.	- Không hỗ trợ công cụ phân tích theo lược đồ mẫu. - Quản lý dữ liệu file chỉ đến 4.096 biến. - Ngôn ngữ cú pháp phức tạp hơn STATA.
3	STATA	Sử dụng các lệnh trực tiếp, hoặc sử dụng nhiều lệnh cho một nghiên cứu thực hiện cùng lúc.	- Cho phép thực hiện thao tác về dữ liệu một cách dễ dàng. - Có thể mạnh trong hồi quy, hồi quy logistic. - Cú pháp các lệnh đồ thị dễ sử dụng. - Đồ thị có chất lượng cao.	- Tại mỗi thời điểm, chỉ làm việc được một file. - Khả năng quản lý dữ liệu không bằng SAS. - Khả năng phân tích phương sai, phân tích nhiều chiều không cao.
4	AMOS	Công cụ hữu ích trong việc thực hiện phân tích mô hình cấu trúc tuyến tính (SEM), phân tích CFA, kiểm định SEM, phân tích Bootstrap, phân tích đa nhóm, phân tích tác động của biến điều tiết, biến trung gian.	- Thể hiện các phân tích và chỉ số dưới dạng hình ảnh đồ họa. - Phần mềm này rất phù hợp cho các bài nghiên cứu có sử dụng dữ liệu khảo sát thuộc ngành quản trị kinh doanh, kế toán, tài chính, ngân hàng, kinh tế phát triển,...	Không cần thiết nếu chỉ thực hiện một mô hình đơn giản, chỉ có một số biến phụ thuộc, không có biến trung gian.

5	Eviews (Econometric Views)	Nghiên cứu dữ liệu chuỗi thời gian, dữ liệu chéo, dữ liệu mảng,... Trong đó dữ liệu lưu trữ theo định dạng độc quyền.	- Chuyên về kinh tế lượng. - Dự báo, mô phỏng hình và tạo các biểu đồ và bảng chất lượng cao. - Giao diện trực quan, dễ sử dụng.	Các mô hình định lượng được thiết lập sẵn khiến khả năng mở rộng mô hình bị giới hạn.
6	Statgraphics	Thực hiện và giải thích bằng ngôn ngữ đơn giản, cả hai chức năng thống kê cơ bản và nâng cao	- Bao quát toàn cảnh, trực quan hoá dữ liệu. - Khả năng phân tích đồ hoạ cao. - Tích hợp với giao diện R, giúp chia sẻ và đẩy nhanh nguồn thông tin. - Dễ dàng chuyển dữ liệu qua lại và thực thi các lệnh.	Khó khăn trong bước xử lý số liệu thô như đối biến số, tính các biến trung gian, mã hoá biến số.
7	MATLAB (Matrix Laboratory)	Cung cấp môi trường lập trình và tính toán kỹ thuật số. Cho phép sử dụng ma trận để tính toán các con số, vẽ thông tin cho các hàm và đồ thị, chạy các thuật toán, tạo giao diện người dùng và liên kết với các chương trình máy tính được viết bằng nhiều ngôn ngữ lập trình khác.	- Dễ dàng tạo ra các ứng dụng có giao diện đồ hoạ tuỳ chỉnh. - Hỗ trợ lượng lớn các hàm dùng trong đại số tuyến tính, thống kê, giải tích tối ưu hoá, phương trình vi phân,... - Linh hoạt trong việc sử dụng, ngôn ngữ lập trình trực quan, dễ học và khả năng ứng dụng cao.	- Chi phí cao. - Các bộ công cụ đi kèm viết riêng cho các nhu cầu nghiên cứu tính phí riêng.
8	Microsoft Excel	Chuyên về xử lý dữ liệu bảng tính	- Đồng bộ hoá, làm việc trên nhiều thiết bị cùng một lúc. - Dự đoán dữ liệu thống kê, công cụ dịch nhanh, giao diện đẹp	- Cần thời gian sử dụng thành thục các tính năng, các hàm. - Khả năng bảo mật kém.

3. Kết luận

Phân tích dữ liệu lớn trong nghiên cứu khoa học tại trường đại học và sử dụng các công cụ trong phân tích số liệu 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. Tại Việt Nam, để có thể nghiên cứu và triển khai phân tích dữ liệu lớn trong nghiên cứu khoa 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. Bên cạnh các dữ liệu tĩnh truyền thống, được lưu trữ trong các hệ thống 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 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. 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]. Hoàng Tuấn Long, Phạm Thu Thủy, “Quy trình quản lý dữ liệu cho các nghiên cứu xã hội học”, *Centre for International Forestry Research* (01/1/2020), <https://www.jstor.org/stable/resrep29156.9?seq=1>.
- [2]. Mai Anh Thơ, “Tổng quan về phân tích dữ liệu lớn trong giáo dục đại học”, *Tạp chí khoa học giáo dục kỹ thuật*, số 50, 96-104 (11/2018).
- [3] Một số công cụ xử lý số liệu trong nghiên cứu khoa học, <https://qlkh.uel.edu.vn/goc-nckh/mot-so-cong-cu-xu-li-so-lieu-trong-nckh>.