

THỰC TRẠNG ỨNG DỤNG PHÂN TÍCH DỮ LIỆU VÀ DỮ LIỆU LỚN TRONG HOẠT ĐỘNG THƯƠNG MẠI ĐIỆN TỬ TẠI HÀ NỘI

● TRẦN THỊ HƯƠNG - VŨ NGỌC THÀNH - PHẠM MAI CHI

TÓM TẮT:

Cùng với xu hướng chuyển đổi số trong các mặt hoạt động khác nhau của nền kinh tế - xã hội, thương mại điện tử đang bùng nổ mạnh mẽ. Việc chuyển từ phương thức kinh doanh truyền thống sang thương mại điện tử đã mang lại lợi ích cho các đơn vị kinh doanh cũng như khách hàng. Việc này cũng kéo theo sự xuất hiện rất nhiều dữ liệu, được xem như nguồn nguyên, nhiên liệu mới. Nếu khai thác được nguồn tài nguyên mới này, thông qua hệ thống thu thập và phân tích dữ liệu lớn, sẽ mang lại những lợi ích cho cả chủ thể kinh doanh và khách hàng trong hoạt động thương mại điện tử. Bài viết đã phân tích thực trạng công tác thu thập, lưu trữ, làm sạch và khai thác dữ liệu trong hoạt động thương mại điện tử cũng như chỉ ra những rào cản đối với hoạt động này.

Từ khoá: dữ liệu lớn, phân tích dữ liệu lớn, thương mại điện tử.

1. Đặt vấn đề

Trong những năm gần đây, dữ liệu lớn là một từ khóa được nhắc đến rất nhiều trong cả học thuật, trong thực tế kinh doanh và công nghệ. Các tập đoàn toàn cầu đang có nhu cầu ngày càng lớn đối với công cụ phân tích kinh doanh dựa trên dữ liệu lớn. Nhiều doanh nghiệp công nghệ thông tin hàng đầu đã xem dữ liệu lớn như nền tảng cho sự phát triển.

Phân tích dữ liệu lớn liên quan đến việc thu thập, lưu trữ, xử lý và phân tích dữ liệu tổng hợp từ nhiều nguồn khác nhau. Khoa học phân tích dữ liệu lớn liên quan đến việc thu thập, lưu trữ, xử lý và phân tích dữ liệu quy mô lớn trên các hệ thống máy tính dựa trên đám mây. Phân tích dữ liệu lớn cho phép doanh nghiệp nhanh chóng kết hợp dữ liệu có cấu trúc (như dữ liệu từ hệ thống CRM, ERP) và dữ liệu phi cấu trúc (như nhật ký máy,

nhật ký máy chủ, và website). Bằng cách sử dụng các công cụ phân tích phù hợp, doanh nghiệp có thể (i) phân tích cú pháp, biến đổi và trực quan hóa dữ liệu; (ii) sử dụng dữ liệu này để phát hiện những quy luật, mô hình của số liệu kinh doanh cũng như những vấn đề bất thường, phân tích nguyên nhân gốc rễ; (iii) xây dựng hệ thống cảnh báo sớm hoặc gợi ý, hỗ trợ quá trình ra quyết định.

Thế giới hiện nay đang chứng kiến sự phát triển mạnh mẽ của thương mại điện tử. Đây là tiền đề và cơ hội cho xu hướng ứng dụng phân tích dữ liệu lớn trong lĩnh vực này. Đổi lại, hệ thống phân tích dữ liệu lớn cũng sẽ mang lại những lợi ích đáng kể cho cả chủ thể kinh doanh và khách hàng trong hoạt động thương mại điện tử.

Các nhà nghiên cứu trên thế giới đã thực hiện rất nhiều các nghiên cứu liên quan đến xây dựng cơ sở dữ liệu lớn, lưu trữ, xử lý và phân tích/

khai thác dữ liệu lớn nói chung, ứng dụng trong thương mại điện tử nói riêng. Khi tìm kiếm trên Google scholar cho từ khóa “Big Data Analytics”, chỉ cần 0,12 giây đã cho ra hơn 100 nghìn kết quả về các công trình nghiên cứu về phân tích dữ liệu lớn. Trong lĩnh vực thương mại điện tử, nền tảng Spark được nghiên cứu khá nhiều và được chứng minh mang lại hiệu quả trong xử lý dữ liệu lớn nói chung (Shaikh et al. 2019), dự đoán tỷ lệ đánh giá của khách hàng (Mishra, Kang, and Woo 2019), đánh giá rủi ro về chất lượng sản phẩm trong hoạt động thương mại điện tử (Liu et al. 2019), và duy trì lượng tồn kho hiệu quả trong những dịp cao điểm như Black Friday cho các nhà kinh doanh thương mại điện tử (Kulkarni, Rajan, and Wang 2020).

Fu (2019) đã nghiên cứu về tiềm năng của các công nghệ phân tích dữ liệu lớn trong phương pháp và chiến lược ra quyết định chính xác trong hoạt động thương mại điện tử; Suguna và cộng sự (2016) nghiên cứu về ứng dụng Hadoop trong xây dựng hệ gợi ý cho hoạt động thương mại điện tử. Thêm vào đó, Le và Liaw (2017) đã phân tích khả năng áp dụng phân tích dữ liệu lớn để phân tích phản hồi của khách hàng. Ngoài ra, các nhà nghiên cứu còn đi sâu vào khả năng khám phá những tri thức tiềm ẩn trong Big Data Clickstreams (Xylogiannopoulos, Karampelas, and Alhajj 2020), khả năng phát hiện những gian lận thương mại trong quá trình giao dịch điện tử (Zhou et al. 2019).

Các nhà nghiên cứu ở Việt Nam cũng đang đi sâu vào nghiên cứu lĩnh vực này. Tuy nhiên, các công trình chủ yếu tập trung về mảng kỹ thuật với rất nhiều đề tài nghiên cứu các cấp được thực hiện và các viện nghiên cứu dữ liệu lớn được thành lập. Đối với chủ đề về hệ thống phân tích dữ liệu lớn trong hoạt động thương mại điện tử, có khá ít công trình được công bố, nhất là những nghiên cứu về mặt quản trị trong quá trình xây dựng và triển khai hệ thống phân tích dữ liệu lớn. Vũ Đức Thi (2017), nghiên cứu tổng quan về các phương pháp xử lý dữ liệu lớn trong thương mại điện tử. Nghiên cứu của Trịnh Thu Trang (2019) mới chỉ dừng lại ở phân tích tổng quan những lợi ích mà dữ liệu lớn có thể mang lại cho hoạt động thương mại điện tử.

Qua tổng quan các nghiên cứu trước, nhóm nghiên cứu phát hiện thấy khoảng trống nghiên

cứu về phân tích thực tế công tác thu thập và phân tích dữ liệu lớn trong hoạt động thương mại điện tử ở Việt Nam nói chung và ở Hà Nội, một thành phố lớn có tốc độ phát triển thương mại điện tử mạnh mẽ của cả nước, nói riêng. Do vậy, nhóm tác giả đã thực hiện nghiên cứu về Thực trạng ứng dụng phân tích dữ liệu và dữ liệu lớn trong hoạt động thương mại điện tử ở Việt Nam nói chung và Hà Nội nói riêng.

2. Phương pháp nghiên cứu

2.1. Phương pháp thu thập dữ liệu

Nghiên cứu được thực hiện thông qua khảo sát bằng bảng hỏi đối với các đơn vị kinh doanh thương mại điện tử. Bảng hỏi được thu thập trong tháng 3 năm 2021. Các câu hỏi chính trong bảng hỏi bao gồm:

- i. Thông tin chung về ngành nghề, lĩnh vực kinh doanh, số năm hoạt động, lượng đơn hàng thông qua hình thức kinh doanh trực tuyến của đơn vị.
- ii. Mức độ thường xuyên thu thập và phân tích dữ liệu nhằm phục vụ hoạt động kinh doanh của đơn vị.
- iii. Thực trạng việc tổ chức, các kênh và công cụ thực hiện thu thập và phân tích dữ liệu lớn trong hoạt động kinh doanh thương mại điện tử.
- iv. Các nhân tố có thể ảnh hưởng đến thực trạng và hiệu quả công tác thu thập và phân tích dữ liệu và dữ liệu lớn trong hoạt động kinh doanh thương mại điện tử.

2.2. Phương pháp phân tích dữ liệu

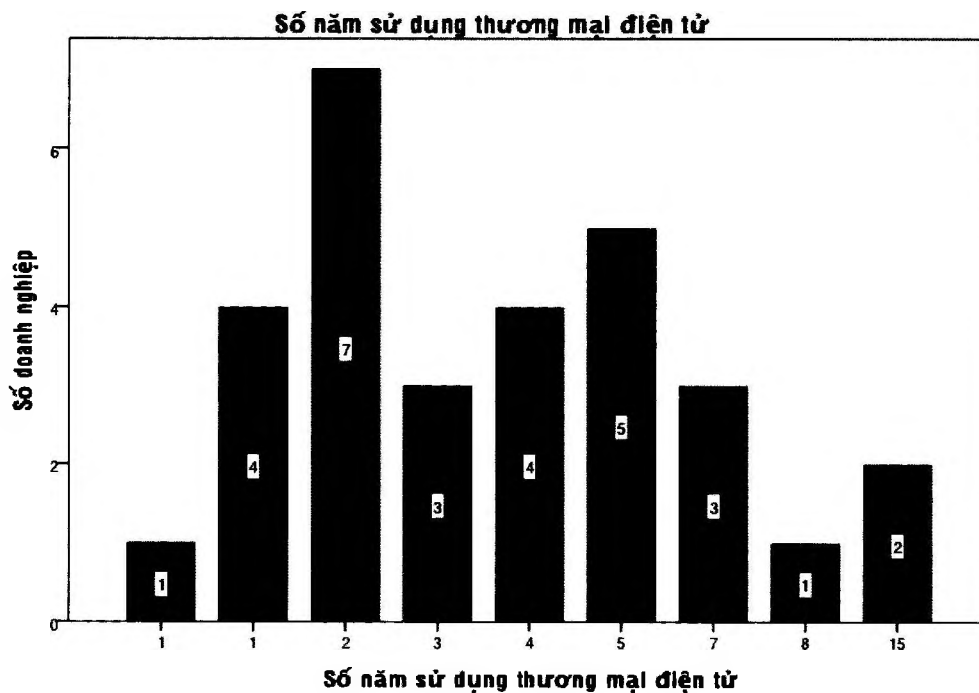
Dữ liệu thu thập được thông qua bảng hỏi được phân tích thống kê qua Excel, Google Forms analytics, và SPSS nhằm đánh giá thực trạng việc ứng dụng phân tích dữ liệu lớn trong hoạt động thương mại điện tử ở Việt Nam nói chung và Hà Nội nói riêng, cũng như các nhân tố ảnh hưởng đến thực trạng này.

3. Kết quả nghiên cứu và thảo luận

3.1. Phân tích mô tả về mẫu điều tra

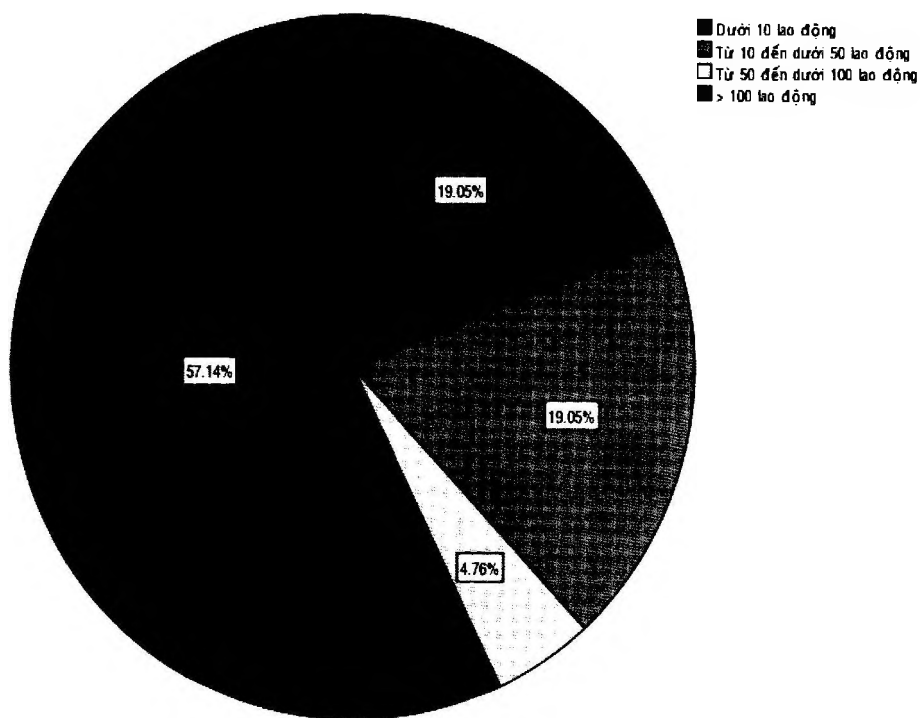
Trong tháng 3 năm 2021, nhóm nghiên cứu đã thực hiện khảo sát với 30 đơn vị có hoạt động kinh doanh thương mại điện tử. Các đơn vị này có số năm hoạt động biến thiên từ 1 đến 75 năm, trong đó, số năm triển khai kinh doanh thương mại điện tử biến thiên từ 1 đến 15 năm (Hình 1).

Hình 1: Số năm triển khai kinh doanh thương mại điện tử của các đơn vị được khảo sát



Nguồn: Tác giả phân tích từ kết quả khảo sát bằng phần mềm SPSS và Excel.

Hình 2: Quy mô lao động của các đơn vị được khảo sát



Nguồn: Tác giả phân tích từ kết quả khảo sát bằng phần mềm SPSS và Excel.

Quy mô lao động của các đơn vị được khảo sát trình bày trong Hình 2. Có 57,14% các đơn vị có số lượng lao động trên 100 người, trong đó có 5 doanh nghiệp được khảo sát có trên 1.000 lao động. Lĩnh vực kinh doanh của các đơn vị được khảo sát cũng khá đa dạng, bao gồm các đơn vị kinh doanh nông sản, hàng dệt may thời trang, điện lạnh, đồ gia dụng, mỹ phẩm, điện tử viễn thông, công nghệ thông tin, và dịch vụ truyền thông.

Hình 3 biểu diễn phân bố của các đơn vị được khảo sát theo số lượng đơn hàng trung bình mỗi tháng (20% số đơn vị được khảo sát có lượng đơn hàng tương đối nhỏ, khoảng dưới 100 đơn/tháng; 33,33% số đơn vị có từ trên 100 đến 1.000 đơn hàng/tháng; 23,33% là tỷ lệ 2 nhóm có số đơn hàng từ 1.001 đến 10.000 đơn/tháng và trên 10.000 đơn/tháng. Như vậy, lượng dữ liệu sinh ra bởi hoạt động kinh doanh thương mại điện tử cũng như hoạt động quản trị tác nghiệp như quản trị nhân sự của các đơn vị kinh doanh này rất lớn.

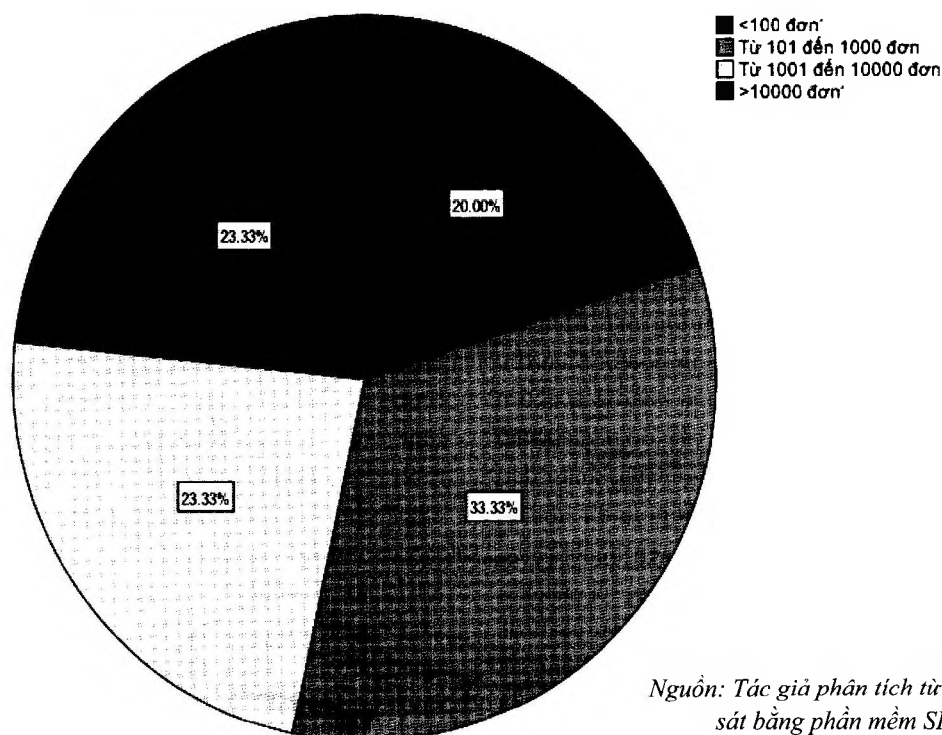
Các đơn vị được khảo sát sử dụng khá đa dạng các kênh khác nhau để triển khai hoạt động kinh doanh, trong đó có (i) các sàn giao dịch thương mại điện tử trong nước như Lazada, Tiki, Shopee (60%); (ii) các trang mạng xã hội (Facebook, Instagram,

Zalo,...) (76,7%); (iii) website của công ty (53,3%); (iv) sàn thương mại điện tử quốc tế (Amazon); và (v) các apps do công ty tự phát triển.

3.2. Thực trạng công tác thu thập dữ liệu phục vụ cho hoạt động kinh doanh thương mại điện tử

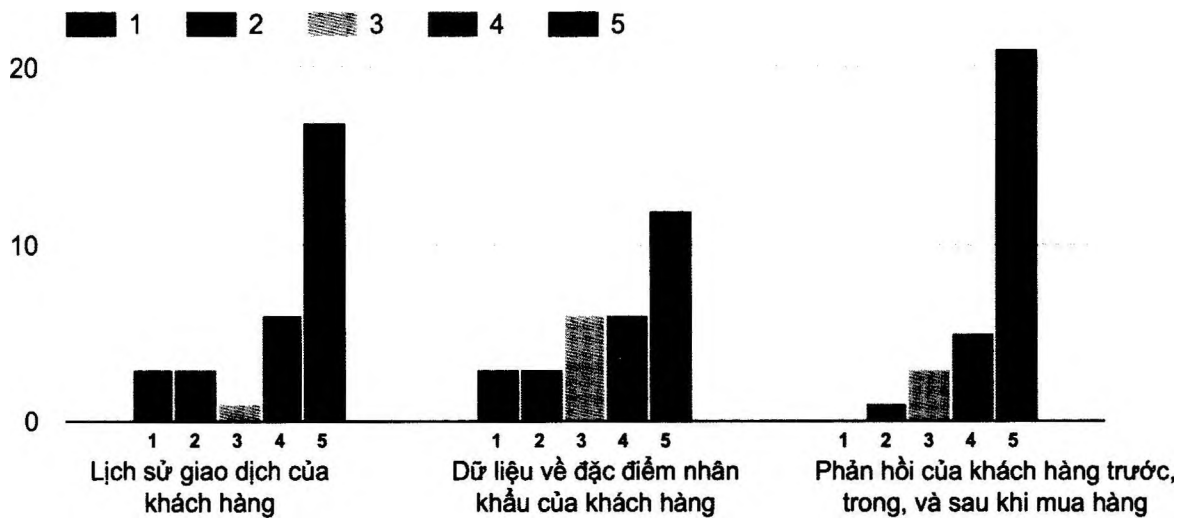
Để có thể khai thác được lợi ích từ dữ liệu trong hoạt động kinh doanh thương mại điện tử, điều quan trọng là các đơn vị kinh doanh phải thu thập và lưu trữ dữ liệu. Thông qua khảo sát, nhóm nghiên cứu nhận thấy, hầu hết các doanh nghiệp đều thực hiện việc thu thập dữ liệu liên quan đến hoạt động thương mại điện tử, bao gồm dữ liệu về lịch sử giao dịch, dữ liệu về đặc điểm nhân khẩu, phản hồi của khách hàng trước, trong, và sau khi mua hàng. Hình 4 cho thấy hơn 50% đơn vị được khảo sát thường xuyên thu thập và lưu trữ thông tin về lịch sử giao dịch của khách hàng; 70% số đơn vị thường xuyên quan tâm đến phản hồi của khách hàng trước, trong và sau khi mua hàng. Ngoài ra, các công ty còn thu thập các thông tin liên quan đến (i) số lượng truy cập vào trang bán hàng; (ii) số lượt tương tác mỗi ngày; (iii) tỷ lệ chuyển đổi qua các kênh marketing khác nhau; và (iv) thời gian từ khi tiếp cận đến sản phẩm đến khi quyết định mua hàng qua các phần mềm như GetflyCRM.

Hình 3: Số lượng đơn hàng tiêu thụ qua kênh thương mại điện tử trung bình mỗi tháng của các đơn vị được khảo sát



Nguồn: Tác giả phân tích từ kết quả khảo sát bằng phần mềm SPSS và Excel.

Hình 4: Mức độ thường xuyên thu thập dữ liệu liên quan đến các mảng khác nhau trong quá trình kinh doanh thương mại điện tử của các đơn vị được khảo sát



(1: Không bao giờ → 5: Rất thường xuyên)

Nguồn: Tác giả biên tập từ kết quả phân tích khảo sát của Google Forms

Các đơn vị được khảo sát sử dụng khá đa dạng các kênh khác nhau để triển khai hoạt động kinh doanh, trong đó có (i) các sàn giao dịch thương mại điện tử trong nước như Lazada, Tiki, Shopee (60%); (ii) các trang mạng xã hội (Facebook, Instagram, Zalo,...) (76,7%); (iii) website của công ty (53,3%); (iv) sàn thương mại điện tử quốc tế (Amazon); và (v) các apps do công ty tự phát triển.

3.2. Thực trạng công tác thu thập dữ liệu phục vụ cho hoạt động kinh doanh thương mại điện tử

Để có thể khai thác được lợi ích từ dữ liệu trong hoạt động kinh doanh thương mại điện tử, điều quan trọng là các đơn vị kinh doanh phải thu thập và lưu trữ dữ liệu. Thông qua khảo sát, nhóm nghiên cứu nhận thấy, hầu hết các doanh nghiệp đều thực hiện việc thu thập dữ liệu liên quan đến hoạt động thương mại điện tử, bao gồm dữ liệu về lịch sử giao dịch, dữ liệu về đặc điểm nhân khẩu, phản hồi của khách hàng trước, trong, và sau khi mua hàng. Hình 4 cho thấy hơn 50% đơn vị được khảo sát thường xuyên thu thập và lưu trữ thông tin về lịch sử giao dịch của khách hàng; 70% số đơn vị thường xuyên quan tâm đến phản hồi của khách

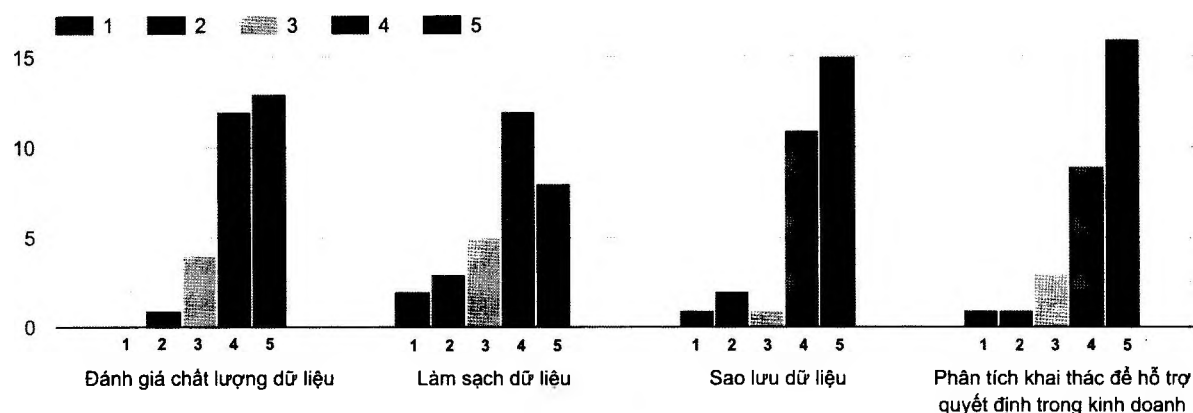
hàng trước, trong và sau khi mua hàng. Ngoài ra, các công ty còn thu thập các thông tin liên quan đến (i) số lượng truy cập vào trang bán hàng; (ii) số lượt tương tác mỗi ngày; (iii) tỷ lệ chuyển đổi qua các kênh marketing khác nhau; và (iv) thời gian từ khi tiếp cận đến sản phẩm đến khi quyết định mua hàng qua các phần mềm như GetflyCRM.

3.3. Thực trạng công tác xử lý và phân tích dữ liệu trong hoạt động kinh doanh thương mại điện tử

Dữ liệu sau khi được thu thập như đã phân tích ở trên, các đơn vị được khảo sát cũng khá thường xuyên thực hiện các công việc như (i) đánh giá chất lượng dữ liệu, (ii) làm sạch dữ liệu, (iii) sao lưu dữ liệu, và (iv) phân tích, khai thác để hỗ trợ quyết định trong kinh doanh với tỷ lệ % các đơn vị được khảo sát thực hiện thường xuyên hoặc rất thường xuyên tương ứng là 83,33%, 66,67%, 86,67%, và 83,33%). (Hình 5)

Bảng 1 thống kê tỷ lệ % các đơn vị được khảo sát theo mức độ thường xuyên thực hiện các kỹ thuật phân tích dữ liệu phục vụ hoạt động kinh doanh thương mại điện tử. Về nhóm kỹ thuật phân

Hình 5: Mức độ thường xuyên xử lý, khai thác dữ liệu thương mại điện tử của các đơn vị được khảo sát



(1: Không bao giờ → 5: Rất thường xuyên)

Nguồn: Tác giả biên tập từ kết quả phân tích khảo sát của Google Forms

Bảng 1. Mức độ thường xuyên thực hiện các kỹ thuật phân tích mô tả dữ liệu thương mại điện tử của các đơn vị được khảo sát

	Tỷ lệ % các đơn vị được khảo sát				
	Không bao giờ	Hiếm khi	Bình thường	Thường xuyên	Rất thường xuyên
Mức độ thường xuyên thực hiện trực quan hoá dữ liệu	3,3	3,3	20,0	33,3	40,0
Mức độ thường xuyên phân tích đặc điểm của khách hàng và phân khúc khách hàng, định vị thị trường, ...	13,3	20,0	26,7	16,7	23,3
Mức độ thường xuyên áp dụng kỹ thuật phân tích dữ liệu trực tuyến (OLAP)	3,3	0	36,7	23,3	36,7
Mức độ thường xuyên sử dụng dữ liệu để dự báo nhu cầu khách hàng	3,3	0	20,0	30,0	46,7
Mức độ thường xuyên sử dụng dữ liệu để dự báo hành vi mua lại của khách hàng	3,3	3,3	23,3	30,0	40,0
Mức độ thường xuyên phân tích dữ liệu để đề xuất các chính sách ưu đãi cho từng nhóm khách hàng	3,3	3,3	10,0	43,3	40,0
Mức độ thường xuyên phân tích dữ liệu để đề xuất các phương án cải tiến trong thương mại điện tử	3,3	0	20,0	30,0	46,7
Mức độ thường xuyên phân tích dữ liệu để đề xuất phát triển sản phẩm mới	6,7	3,3	13,3	33,3	43,3

Nguồn: Tác giả tổng hợp từ kết quả phân tích bằng phần mềm SPSS và Excel.

tích mô tả, đa số các đơn vị có thực hiện phân tích, tuy nhiên chủ yếu tập trung ở việc phân tích mô tả và trực quan hóa dữ liệu, khá ít doanh nghiệp thường xuyên tận dụng được các công cụ phân tích dữ liệu trực tuyến (OLAP). Về nhóm kỹ thuật phân tích dự báo, 76,7% và 70% các đơn vị được khảo sát thường xuyên hoặc rất thường xuyên thực hiện các phân tích nhằm dự báo nhu cầu khách hàng hoặc dự báo về hành vi mua lại của khách hàng.

Về nhóm kỹ thuật phân tích đề xuất, các đơn vị được khảo sát cũng đang khai thác dữ liệu để đề xuất các phương án cải tiến hoạt động kinh doanh, phát triển sản phẩm mới, hay đề xuất các chính sách ưu đãi riêng biệt cho từng nhóm đối tượng khách hàng khác nhau. Vẫn còn tồn tại một số doanh nghiệp không bao giờ hoặc hiếm khi thực hiện phân tích dữ liệu trong hoạt động kinh doanh thương mại điện tử của mình.

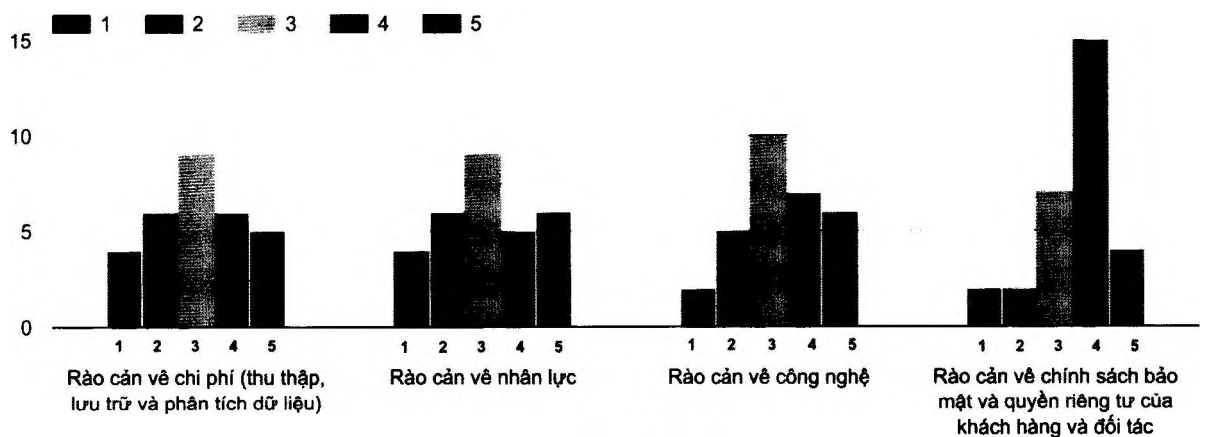
Bên cạnh các công ty/ sản phẩm thương mại điện tử lớn có riêng bộ phận rất lớn thực hiện hoạt động phân tích dữ liệu với các công nghệ chuyên nghiệp như Hadoop, Apache Spark thì các đơn vị kinh doanh thương mại điện tử thường sử dụng các công cụ để thu thập, phân tích dữ liệu và dữ liệu lớn như: Power BI, Tableau, Excel, SAS, hoặc công cụ tự phát triển, ERP, Kiotviet, nhanh.vn, Base+,...

hoặc tận dụng công cụ Analytics của các nền tảng như Google hay Facebook.

3.4. Các nhân tố ảnh hưởng đến hoạt động phân tích dữ liệu và dữ liệu lớn trong thương mại điện tử

Thông qua các kết quả khảo sát cho thấy một số doanh nghiệp, nhất là doanh nghiệp nhỏ không thường xuyên thu thập và phân tích dữ liệu phục vụ hoạt động kinh doanh thương mại điện tử của mình do còn gặp nhiều khó khăn. Hình 6 đã trực quan kết quả khảo sát về những rào cản đối với hoạt động thu thập và phân tích dữ liệu. Hầu hết các đơn vị được khảo sát đều cho rằng có 4 rào cản chính, bao gồm: Rào cản về chi phí, rào cản về nhân lực, rào cản về công nghệ và rào cản về chính sách bảo mật, quyền riêng tư của khách hàng và đối tác. Trong đó rào cản về nhân lực là có tỷ lệ số đơn vị cao nhất đánh giá có ảnh hưởng rất lớn. Rào cản về công nghệ và chính sách bảo mật được đánh giá ở mức 4 và mức 5 cũng tương đối cao, cho thấy vấn đề hạn chế trong tiếp cận công nghệ của các đơn vị kinh doanh thương mại điện tử, nhất là những đơn vị có quy mô nhỏ. Vấn đề về chính sách bảo mật và quyền riêng tư là vấn đề chung của tất cả các doanh nghiệp không phân biệt quy mô, điều này đòi hỏi các doanh nghiệp cần thận trọng

Hình 6: Mức độ ảnh hưởng của các nhân tố/ rào cản lên hoạt động thu thập và phân tích dữ liệu trong thương mại điện tử của các đơn vị được khảo sát



(1: Hoàn toàn không ảnh hưởng → 5: Ảnh hưởng rất lớn)

trong quá trình thu thập, chia sẻ và phân tích dữ liệu phục vụ cho hoạt động kinh doanh.

4. Kết luận

Phân tích dữ liệu nói chung và dữ liệu lớn nói riêng trong kinh doanh thương mại điện tử ngày càng quan trọng và cấp thiết đối với các doanh nghiệp. Nghiên cứu đã thực hiện khảo sát và phân tích thực trạng tình hình thu thập và phân tích dữ liệu của các đơn vị kinh doanh thương mại điện tử thông qua các tiêu chí chính, như (i) mức độ thường xuyên thực hiện hoạt động thu thập, xử lý và phân tích dữ liệu; (ii) các công cụ, kỹ thuật được sử dụng để khai thác nguồn tài nguyên dữ liệu quý giá này; (iii) những rào cản đối với hoạt

động thu thập và phân tích dữ liệu trong hoạt động kinh doanh thương mại điện tử.

Hạn chế của nghiên cứu này là cỡ mẫu khảo sát mặc dù đại diện cho các nhóm quy mô doanh nghiệp, lĩnh vực kinh doanh khác nhau, nhưng vẫn còn khá nhỏ. Các giải pháp để khắc phục những rào cản đối với hoạt động thu thập và phân tích dữ liệu cũng chưa được đưa ra. Các nghiên cứu trong tương lai có thể thực hiện khảo sát trên quy mô rộng hơn và thực hiện phỏng vấn sâu các nhóm doanh nghiệp khác nhau, phân tích và tìm ra giải pháp để hoạt động thu thập và phân tích dữ liệu trong hoạt động kinh doanh thương mại điện tử được thực hiện một cách hiệu quả■

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Bách khoa Hà Nội trong đề tài mã số T-2020-PC-045.

TÀI LIỆU THAM KHẢO:

1. Fu, L. (2019). Precision Marketing Method and Strategy Based on Big Data Analysis in E-Commerce Environment. *In International conference on Big Data Analytics for Cyber-Physical-Systems, Singapore, 2019* (1075-1080). Singapore: Springer.
2. Kulkarni, Chhaya, Kavitha Loganathan Sundara Rajan, and Xin Wang. (2020). Efficient Product Inventory Maintenance for Black Friday Sale via Spark Big Data System. *UMBC Student Collection*.
3. Le, Thi Mai, and Shu-Yi Liaw. (2017). Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an e-Commerce Context. *Sustainability* 9(5): 798.
4. Liu, Yi, Jiahuan Lu, Feng Mao, and Kaidi Tong. (2019). The Product Quality Risk Assessment of E-Commerce by Machine Learning Algorithm on Spark in Big Data Environment. *Journal of Intelligent & Fuzzy Systems* (Preprint): 1-11.
5. Mishra, Monika, Mingoo Kang, and Jongwook Woo. 2019. Rating Prediction Using Deep Learning and Spark. [Online] Available at: <https://www.calstatela.edu/sites/default/files/groups/High%20Performance%20Information%20Computing%20Center%20%28HiPIC%29/papers/ratingpredictiondlsarkiconi2019.pdf>
6. Shaikh, E., Mohiuddin, I., Alufaisan, Y., & Nahvi, I. (2019). Apache Spark: A Big Data Processing Engine. *2nd IEEE Middle East and North Africa COMMunications Conference (MENACOMM), Bahrain, 2019* (1-6). Bahrain: IEEE.
7. Suguna, S., M. Vithya, and JI Christy Eunaicy. (2016). "Big Data Analysis in E-Commerce System Using HadoopMapReduce." *In 2016 International Conference on Inventive Computation Technologies (ICICT), India, 2016*, (1-6). India: IEEE
8. Trịnh Thu Trang. (2019). Nghiên cứu về lợi ích của dữ liệu lớn - Big Data với doanh nghiệp thương mại điện tử trong nước và thế giới. *Tạp chí Công Thương*. Truy cập tại: <http://tapchicongthuong.vn/bai-viet/nguyen-cuu-ve-loi-ich-cua-du-lieu-lon-big-data-voi-doanh-nghiep-thuong-mai-dien-tu-trong-nuoc-va-the-gioi-64331.htm>.
9. Vũ Đức Thi. (2017). "Nghiên cứu phát triển một số phương pháp xử lý dữ liệu lớn và ứng dụng trong thương mại điện tử. Truy cập tại: https://repository.vnu.edu.vn/handle/VNU_123/62167

10. Xylogiannopoulos, Konstantinos F., Panagiotis Karampelas, and Reda Alhaji. (2020). Simplifying E-Commerce Analytics by Discovering Hidden Knowledge in Big Data Clickstreams. In Kaya M., Birinci Ş., Kawash J., Alhaji R. (Eds). *Putting Social Media and Networking Data in Practice for Education, Planning, Prediction and Recommendation. Lecture Notes in Social Networks* (pp. 51-74). Cham, Switzerland: Springer.
11. Zhou, Hangjun et al. (2019). A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics. [online] Available at: <https://www.techscience.com/cmc/v60n1/28371>

Ngày nhận bài: 12/4/2021

Ngày phản biện đánh giá và sửa chữa: 25/4/2021

Ngày chấp nhận đăng bài: 7/5/2021

Thông tin tác giả:

TRẦN THỊ HƯƠNG - VŨ NGỌC THÀNH - PHẠM MAI CHI

Viện Kinh tế và Quản lý, Trường Đại học Bách khoa Hà Nội

APPLICATION OF DATA AND BIG DATA ANALYSIS IN E-COMMERCE ACTIVITIES IN HANOI CITY

● TRAN THI HUONG - VU NGOC THANH - PHAM MAI CHI

**School of Economics and Management / Hanoi University
of Science and Technology**

ABSTRACT:

Along with the trend of digital transformation in different socio-economic aspects, e-commerce is growing rapidly. The transition from traditional business methods to e-commerce models has brought benefits to both businesses and consumers. This trend also generates a vast amount of data which is considered a new resource for businesses in the new development era. The successful exploitation of this new resource through a big data collection and analysis system will bring significant benefits to both business entities and customers in e-commerce activities. This study analyzes the current situation of collecting, storing, cleaning, and mining data in e-commerce and points out barriers to these activities in Vietnam.

Keywords: data analytics, big data, big data analytics, e-commerce.