ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 225(06): 536 - 540

TỔNG QUAN PHÂN TÍCH DỮ LIỆU LỚN TRONG THƯƠNG MẠI ĐIỆN TỬ

Lê Triệu Tuấn, Lý Thu Trang*
Trường Đại học Công nghệ thông tin & Truyền thông - ĐH Thái Nguyên

TÓM TẮT

Phân tích dữ liệu lớn mang lại nhiều lợi ích cho các doanh nghiệp thương mại điện tử. Nó không chỉ cho phép họ hiểu sâu hơn về hành vi khách hàng của họ và xu hướng phát triển lĩnh vực kinh doanh mà còn cho phép họ đưa ra những quyết định chính xác hơn để cải thiện việc bán hàng, tiếp thị, giữ chân khách hàng và mọi khía cạnh khác trong kinh doanh. Tuy nhiên, phân tích dữ liệu lớn có thể gặp nhiều khó khăn và đặc biệt là khi cơ sở hạ tầng phục vụ cho nó không hoạt động tối ưu dẫn đến các thông tin quan trọng không khả dụng hoặc bị chậm trễ. Bài báo nghiên cứu một cách tổng quan về lợi ích của việc phân tích dữ liệu lớn và đề xuất một mô hình phân tích dùng cho các doanh nghiệp thương mại điện tử, giúp các doanh nghiệp này có thêm cái nhìn về việc sử dụng dữ liệu lớn để cải thiện hiệu suất kinh doanh. Mô hình này có thể làm tài liệu tham khảo cho các nghiên cứu tiếp theo về dữ liệu lớn.

Từ khóa: *Dữ liệu lớn; thương mại điện tử; phân tích dữ liệu lớn; hiệu quả kinh doanh; Spark.*

Ngày nhận bài: 09/3/2020; Ngày hoàn thiện: 31/5/2020; Ngày đăng: 31/5/2020

OVERVIEW OF BIG DATA ANALYTICS IN E-COMMERCE

Le Trieu Tuan, Ly Thu Trang*
TNU - University of Information and Communication Technology

ABSTRACT

Big data analytics brings many benefits to e-commerce businesses. It not only allows them to better understand their customer behavior and business development trends, but also allows them to make more accurate decisions to improve sales, marketing, retention customers and every other aspect of business. However, analyzing big data can be difficult, especially when the infrastructure that is serving it does not work optimally, leading to important information being unavailable or delayed. This paper reviews an overview of the benefits of big data analytics and proposes an analytics model for e-commerce businesses, giving them an insight into their use. Big data to improve sales performance. This model can serve as a reference for subsequent studies on big data.

Keywords: *Big data; E-commerce; big data analytics; business performance; Spark.*

Received: 09/3/2020; Revised: 31/5/2020; Published: 31/5/2020

* Corresponding author. *Email: tranglt@ictu.edu.vn*

1. Giới thiệu

Dữ liệu lớn (Big data) đã và đang làm thay đổi cách mà các công ty thương mại điện tử đưa ra quyết định trong kinh doanh, nó đòi hỏi tư duy mới với các khái niệm về công nghệ và giá trị mới. Thuật ngữ 'big data' đã trở thành từ thông dụng trên internet từ năm 2012. Cũng từ thời gian đó, dữ liệu lớn hứa hẹn sẽ được ứng dụng nhiều hơn trong tương lai. Ngày nay, các doanh nghiệp thương mại điện tử quy mô nhỏ hay lớn đều có xu hướng sử dụng phân tích dữ liệu lớn để tạo ra ưu thế cạnh tranh. Khi đề cập tới big data thì không chỉ đề cập tới chính nó, mà đó còn là về khả năng phân tích sử dụng công nghệ có giá cả phải chăng. Nhiều công ty thương mại điện tử thực hiện phân tích dữ liệu lớn theo thời gian thực và đạt được những kết quả có giá trị thúc đẩy lợi nhuận và đưa ra những quyết định kinh doanh tốt hơn. Tuy nhiên, phân tích dữ liệu lớn có thể gặp khó khăn khi cơ sở hạ tầng của dữ liệu lớn không hoạt động tối ưu và khi đó thuật toán phân tích sẽ hoạt động không hiệu quả hoặc bị trì hoãn [1].

2. Dữ liệu lớn

Dữ liệu lớn được định nghĩa là dữ liệu vượt quá khả năng xử lý của hệ thống xử lý dữ liệu thông thường do khối lượng, vận tốc và tính biến đổi của nó [2]. Dữ liệu lớn thường tồn tại dưới dạng có cấu trúc và phi cấu trúc:

Dữ liệu có cấu trúc: những dữ liệu này tồn tại ở dạng nhân khẩu học bao gồm tên, tuổi, địa chỉ, giới tính, ngày sinh, sở thích... có thể được phân tích dễ dàng.

Dữ liệu phi cấu trúc: dữ liệu dạng này chứa các thông tin phức tạp như: các nội dung đính kèm email, hình ảnh, nội dung bình luận trên mạng xã hội… dữ liệu dạng này khó phân tích hơn dữ liệu dạng có cấu trúc, dữ liệu lớn đa số tồn tại dưới dạng này.

Big data có 5 đặc trưng (5V) như sau:

Khối lượng (Volume): Đây là đặc trưng cơ bản nhất của big data, khối lượng của big data đang tăng lên từng ngày. Dữ liệu truyền thống có thể lưu trữ trên các thiết bị đĩa mềm, đĩa cứng. Nhưng với big data thì ta phải dùng công nghệ mới như công nghệ điện toán đám mây thì mới đáp ứng khả năng lưu trữ.

Tốc độ (Velocity): Tốc độ có thể được hiểu theo 2 khía cạnh: thứ nhất, khối lượng dữ liệu gia tăng rất nhanh (mỗi giây có tới 72,9 triệu các yêu cầu truy cập tìm kiếm trên web bán hàng của Amazon); thứ hai, xử lý dữ liệu nhanh ở mức thời gian thực (real-time), có nghĩa dữ liệu được xử lý ngay tức thời ngay sau khi chúng phát sinh (tính bằng mili giây). Các ứng dụng phổ biến trên lĩnh vực Internet, tài chính, ngân hàng, hàng không, quân sự, y tế… như hiện nay phần lớn dữ liệu được xử lý real-time.

Đa dạng (Variety): Các dữ liệu được sinh ra là phi cấu trúc (tài liệu, blog, hình ảnh, video, bài hát, dữ liệu từ thiết bị cảm biến vật lý, thiết bị chăm sóc sức khỏe…). Big data cho phép liên kết và phân tích nhiều dạng dữ liệu khác nhau.

Độ chính xác (Veracity): Bài toán phân tích và loại bỏ dữ liệu thiếu chính xác và nhiễu đang là tính chất quan trọng của big data. Với xu hướng phương tiện truyền thông xã hội (Social Media) và mạng xã hội (Social Network) ngày nay và sự gia tăng mạnh mẽ tính tương tác và chia sẻ của người dùng Mobile làm cho bức tranh xác định về độ tin cậy và chính xác của dữ liệu ngày càng khó khăn hơn.

Giá trị (Value): Giá trị là một đặc trưng quan trọng của big data, nó cho chúng ta thông tin để ra quyết định xem có nên triển khai big data hay không. Nếu chúng ta có big data mà chỉ nhận được phần trăm lợi ích rất nhỏ từ nó thì không nên đầu tư phát triển.

3. Phân tích dữ liệu lớn và lợi ích đối với doanh nghiệp thương mại điện tử

Phân tích dữ liệu lớn mang lại nhiều lợi ích cho doanh nghiệp kinh doanh thương mại điện tử, giúp doanh nghiệp giảm chi phí và nâng cao hiệu quả kinh doanh, giúp các nhà

quản lý của công ty ra quyết định chính xác hơn. Các công ty thương mại điện tử hàng đầu như: Amazon, Alibaba, Lazada là những nhà tiên phong trong việc tạo ra giá trị từ việc ứng dụng phân tích dữ liệu lớn. Dưới đây là những lợi ích cụ thể mà phân tích dữ liệu lớn có thể mang lại cho công ty thương mại điện tử:

(1) Phân tích dữ liệu lớn cho phép các công ty tìm kiếm các thông tin kinh doanh có giá trị

Có nhiều công trình nghiên cứu được công bố gần đây đã cho thấy nhiều công ty thương mại điện tử đã ứng dụng phân tích dữ liệu lớn để có được những thông tin kinh doanh có giá trị vượt trội hơn so với các công ty kinh doanh cùng lĩnh vực từ 5% đến 6% về năng suất lợi nhuận [3]. Theo bài báo "Big data and its technical challenges" [4] trên các kết quả khảo sát các giám đốc điều hành cấp cao từ các công ty trên 19 quốc gia và 7 ngành công nghiệp cho kết quả là 92% giám đốc điều hành có ứng dụng phân tích dữ liệu lớn hài lòng với nhận định này. Ví dụ: Amazon, gã khổng lồ trong thương mại điện tử là một ví dụ điển hình về việc tạo ra giá trị kinh doanh từ việc ứng dụng phân tích dữ liệu lớn. Trong vài năm trở lại đây, công ty đã có thể tạo ra sự tăng trưởng hơn 10% hàng năm về doanh số [5].

(2) Phân tích dữ liệu lớn cho phép doanh nghiệp hiểu rõ hơn thông tin từ nhiều nguồn khác nhau từ đó đưa ra quyết định chính xác hơn

Việc đưa ra quyết định của các đối tượng quản lý phụ thuộc lớn vào thông tin có được. Thông tin có chính xác và đầy đủ thì hiệu quả của quyết định đó càng lớn. Các nhà quản lý của các công ty kinh doanh trong lĩnh vực thương mại điện tử dựa vào thông tin khách hàng, thông tin bán hàng, thông tin về các đối thủ cạnh tranh, thông tin về thị trường... để đưa ra quyết định kinh doanh cho công ty mình. Do đó, phân tích dữ liệu lớn cho phép các doanh nghiệp hiểu rõ hơn về khách hàng của mình, về nhu cầu thị trường và thực hiện các điều chỉnh để thúc đẩy nhanh hơn quá trình ra quyết định. Theo nghiên cứu của Weiland [6], ngày nay hầu hết các công ty thương mại điện tử lớn đã ứng dụng phân tích dữ liệu lớn để hỗ trợ trong việc đưa ra các quyết định trong kinh doanh.

(3) Việc phân tích dữ liệu lớn cũng có thể mang lại lợi ích cho khách hàng vì họ có thể nhận các thông tin hoặc các dịch vụ được thiết kế, tùy chỉnh dành riêng cho khách hàng

Dựa vào dữ liệu lịch sử bán hàng người ta có thể thay đổi giá bán theo thời vụ hoặc theo đối thủ cạnh tranh. Việc theo dõi hành vi người tiêu dùng cũng có thể làm thay đổi giá bán dựa trên lưu lượng truy cập và chuyển đổi để tối ưu hóa doanh thu và lợi nhuận. Phân tích dữ liệu lớn cung cấp một cái nhìn sâu sắc chi tiết liên quan đến hành vi của khách hàng, điều này có thể giúp doanh nghiệp thương mại điện tử điều chỉnh các sản phẩm và dịch vụ của họ theo sở thích và nhu cầu của khách hàng. Ngoài ra, mỗi thông tin của khách hàng liên quan đến các sản phẩm hoặc dịch vụ cụ thể giúp công ty thực hiện việc quảng cáo và tiếp thị hiệu quả hơn. Việc này sẽ giúp doanh nghiệp tiết kiệm chi phí quảng cáo cũng như tăng doanh thu thông qua những khách hàng có giá trị cao. Khi khách hàng có thông tin được lưu trữ tại công ty thì có thể nhận được các thông tin dựa trên hoàn cảnh và yêu cầu của họ. Sự giao tiếp giữa khách hàng và công ty được tăng cường, trong đó khách hàng dễ dàng bày tỏ mối quan tâm và nhu cầu của họ tới công ty [7].

(4) Phân tích dữ liệu lớn cung cấp một cách hiệu quả và nhanh hơn để thu thập, quản lý, phân phối và kiểm soát thông tin có cấu trúc và không có cấu trúc trên các kênh thương mại điện tử

Nội dung trên trang web chính xác sẽ tạo niềm tin và hỗ trợ cải thiện hình ảnh thương hiệu. Trong trường hợp khi thông tin về một sản phẩm không đầy đủ sẽ khó tìm hoặc không nhất quán thì khách hàng có xu hướng sẽ tìm các lựa chọn thay thế khác. Do đó, điều này có thể ảnh hưởng tiêu cực đến hiệu suất

doanh số và thương hiệu của công ty [4]. Bằng cách sử dụng phân tích thống kê, xử lý phân tán, thuật toán động và khả năng theo thời gian thực, các thực thể thương mại điện tử có thể thiết lập các cơ chế định giá thực tế trên mỗi hồ sơ khách hàng. Điều này có thể giúp các nhân viên bán hàng tốt hơn vì giờ đây, họ có dữ liệu thực tế và theo thời gian thực. Thông qua phân tích dữ liệu, các công ty có thể ước tính được độ co giãn của giá. Chẳng hạn, trong trường hợp khi giá được tăng cho một sản phẩm hoặc dịch vụ cụ thể, hoặc khi có nhu cầu giảm giá [8].

Mạng xã hội là công cụ giúp cho các công ty gắn kết với khách hàng cung cấp các cơ hội thị trường cũng như việc phát triển khách hàng tiềm năng. Sự phát triển của các phương tiện truyền thông xã hội ngày càng được nhiều công ty sử dụng để xây dựng nên các chiến dịch tiếp thị sản phẩm. Sự sẵn có của số lượng lớn dữ liệu được sinh ra từ các tương tác khách hàng trực tuyến có thể hữu ích cho các nhân viên bán hàng đạt hiệu quả hơn [9].

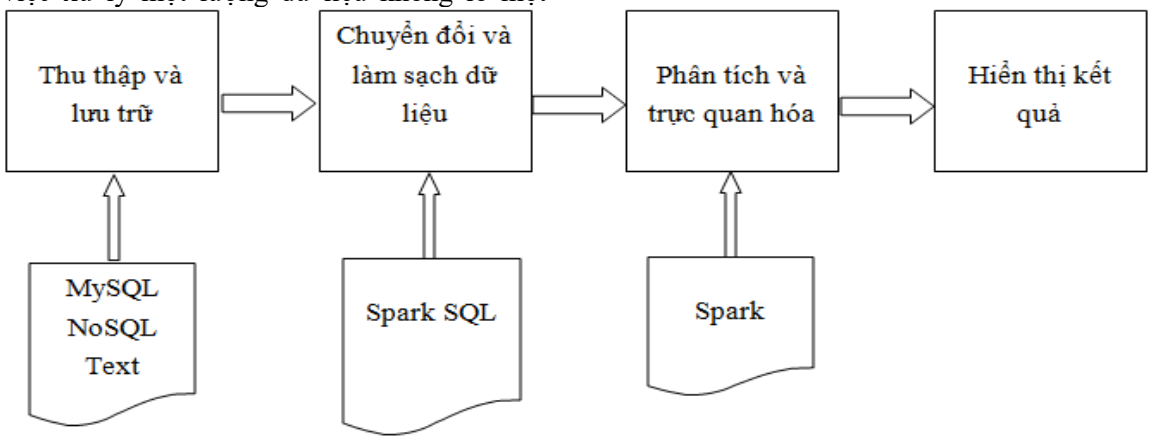
4. Công nghệ phân tích dữ liệu lớn

Hiện nay có nhiều công nghệ phân tích dữ liệu lớn đã và đang được nghiên cứu và phát triển bởi các nhà nghiên cứu trong nước và trên thế giới. Các công nghệ này giúp cho việc xử lý một lượng dữ liệu khổng lồ một cách nhanh chóng, giúp cho người dùng dễ dàng tìm được thông tin cần thiết nhanh chóng trong thời gian thực. Bài báo giới thiệu một công nghệ mới hiện nay được rất nhiều cá nhân và tổ chức sử dụng để phân tích dữ liệu lớn, đó là Apache Spark.

Apache Spark là một công nghệ xử lý rất nhanh một lượng lớn dữ liệu, hỗ trợ rất mạnh cho lập trình Java, Scala và Python. Ngoài ra, còn có một tập hợp các thư viện hỗ trợ xử lý stream, machine và phân tích hình ảnh. Tốc độ xử lý của Apache Spark có được do việc tính toán được thực hiện cùng lúc trên nhiều máy khác nhau. Đồng thời việc tính toán được thực hiện hoàn toàn ở bộ nhớ trong của máy tính. Apache Spark cho phép xử lý dữ liệu theo thời gian thực, vừa nhận dữ liệu từ các nguồn khác nhau đồng thời thực hiện ngay việc xử lý dữ liệu vừa nhận được. Apache Spark có thể nhanh hơn gấp 10 lần so với các công nghệ trước nó [10]. Với những ưu điểm vượt trội của Spark, việc sử dụng công nghệ này vào phân tích dữ liệu lớn sẽ mang lại những lợi ích nhất định cho doanh nghiệp.

5. Mô hình xử lý, phân tích dữ liệu lớn

Dựa trên các hình thức kinh doanh thương mại điện tử và lý thuyết về dữ liệu lớn, nhóm tác giả xây dựng mô hình xử lý, phân tích dữ liệu lớn được mô tả như hình 1.



Hình 1. *Mô hình xử lý, phân tích và xử lý dữ liệu lớn*

Trích xuất và thu thập: Theo truyền thống, dữ liệu tồn tại dưới dạng có cấu trúc và thường được lưu trữ trong kho dữ liệu tĩnh sẽ được truy vấn định kỳ. Đối với dữ liệu lớn, nguồn vào linh động và dữ liệu được trích xuất ra thường tồn tại dưới dạng không có cấu trúc. Một doanh nghiệp thương mại điện tử có nhiều nguồn dữ liệu, như: thông tin về đối thủ cạnh tranh, dữ liệu về giá cả,

dữ liệu bán hàng, dữ liệu về tài chính, cổ phiếu, dữ liệu quảng cáo, các phản hồi của khách hàng… những dữ liệu này sẽ được thu thập và lưu trữ dưới dạng SQL, NoSQL hoặc văn bản.

Chuyển đổi và làm sạch dữ liệu: Các dữ liệu lúc này vẫn còn nhiều thành phần dư thừa hoặc tồn tại ở dạng chưa được chuẩn hóa, do vậy cần được chuyển đổi và làm sạch. Sau đó sẽ được lưu vào Spark SQL - một thành phần của Spark.

Phân tích và trực quan hóa: Giai đoạn này, dữ liệu sẽ được phân tích bởi Spark. Các kết quả phân tích có thể được áp dụng trực tiếp vào hệ thống quản lý dữ liệu phân tán Mllib - một thành phần của Spark. Đây thường sẽ là một cơ sở dữ liệu SQL hoặc NoSQL. Công việc sử dụng, phân tích sau đó được áp dụng cho cơ sở dữ liệu này.

Hiển thị kết quả: Ở giai đoạn này, doanh nghiệp nhận được kết quả xử lý và sẽ sử dụng vào các công việc như: định giá sản phẩm, thiết lập lên chiến dịch quảng cáo, dùng để ra quyết định.

6. Kết luận

Phân tích dữ liệu lớn có thể hỗ trợ các doanh nghiệp thương mại điện tử nâng cao hiệu suất kinh doanh và chăm sóc khách hàng. Điều này giúp doanh nghiệp duy trì và thu hút thêm khách hàng tiềm năng bên cạnh việc cải thiện chất lượng kinh doanh và nâng cao hình ảnh thương hiệu. Với sự tăng trưởng của dữ liệu trong lĩnh vực thương mại điện tử khoảng 30% đến 60% hàng năm thì đây là một cơ hội rõ ràng để cho các doanh nghiệp này nắm bắt, áp dụng công nghệ phân tích dữ liệu lớn vào doanh nghiệp mình. Các doanh nghiệp thương mại điện tử cần có một hệ thống quản lý dữ liệu hiện đại để thay thế những hệ thống lỗi thời. Các hệ thống này phải cung cấp một sự bảo đảm mạnh mẽ, nhất quán trên cơ sở tỷ lệ lưu trữ và tỷ lệ tính toán. Các cơ sở dữ liệu mới cũng cần có khả năng mở rộng trong việc sử dụng các cơ chế dữ liệu đơn giản hơn về tính nhất quán dữ liệu với chi phí thấp trong việc tăng cường tính khả dụng và hiệu suất. Mô hình bài báo đã trình bày có thể là một yếu tố để cho các doanh nghiệp tham khảo.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1]. N. Couldry, and J. Turow, "Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy," *International Journal of Communication*, vol. 8, pp. 1710-1726, 2014.

[2]. M. Graham, "Big data and the end of theory?," 2012. [Online]. Available: https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory. [Accessed March, 2020].

[3]. H. Hu, Y. Nen, T. S. Chua, and X. Li, "Towards Scalable System for Big Data Analytics: A Technology Tutorial," *IEEE Access*, vol. 2, pp. 652-687, June 2014.

[4]. H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrisshnan, and C. Shahabi "Bigdata and íts technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86-94, July, 2014.

[5]. Danziger, "How Amazon used big data to rule e-commerce", 2019. [Online]. Available: https://insidebigdata.com/2019/11/30/how-amazon-used-big-data-to-rule-e-commerce/ [Accessed November 30, 2019].

[6]. Weiland, "Big data as an information source in the decision making-processes of the e-commerce companies," *Research journal of the university of Gdansk*, vol. 71, pp. 179-190, 2017, doi: 10.5604/01.3001.0010.5734.

[7]. O. P. Rud, *Data mining cookbook: modeling data for marketing, risk, and customer relationship management*. New York: John Wiley & Sons, 2011.

[8]. J. Manyika, M. Chui, B. Brown, J. Bhuhin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big Data: The next frontier for innovation, competition and productivity," June 2011. [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx [Accessed May 2020]

[9]. Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh, and A. V. Vasilakos, "Big Data Analytics: A Survey," 2015. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0030-3 [Accessed March, 2020]

[10]. N. N. Phung, "About the Spark", 2020. [Online]. Available: https://viblo.asia/p/tim-hieu-ve-apache-spark-ByEZkQQW5Q0, [Accessed May 2020].