

MỘT SỐ CÔNG CỤ PHÂN TÍCH DỮ LIỆU THÔNG MINH VÀ ỨNG DỤNG TRONG CÁC BÀI TOÁN KINH TẾ

● NGUYỄN THỊ HỘI - BÙI QUANG TRƯỜNG

TÓM TẮT:

Sự phát triển mạnh mẽ của công nghệ thông tin và ứng dụng các phương tiện truyền thông xã hội vào các hoạt động sản xuất, kinh doanh đã làm cho bài toán phân tích dữ liệu trong các doanh nghiệp và tổ chức ngày càng trở nên quan trọng hơn bao giờ hết. Nghiên cứu nhằm giới thiệu và so sánh một số công cụ phân tích dữ liệu thông minh cùng các bài toán được ứng dụng trên thực tế hiện nay. Bài viết được chia thành 3 phần, gồm: giới thiệu chung về bài toán phân tích dữ liệu trong các doanh nghiệp và tổ chức; trình bày một số công cụ phân tích dữ liệu phổ biến hiện nay; đồng thời đưa ra một số so sánh và hàm ý đề xuất ứng dụng trong các bài toán kinh tế trong hoạt động sản xuất, kinh doanh của các doanh nghiệp, tổ chức và cơ quan nhà nước.

Từ khóa: phân tích dữ liệu, công cụ phân tích, dữ liệu kinh doanh, bài toán dự báo.

1. Đặt vấn đề

Cùng với sự phát triển của công nghệ thông tin và sự lan tỏa của cuộc cách mạng công nghệ lần thứ 4, các công cụ công nghệ thông tin ngày càng được ứng dụng nhiều hơn trong các bài toán kinh tế. Đặc biệt, với sự phát triển của Dữ liệu lớn (Big Data), Khoa học dữ liệu (Data Science) và các công cụ khai phá dữ liệu (Data Mining Tools) đã thúc đẩy sự phát triển quy trình phân tích dữ liệu lên một cấp độ mới. Quy trình không chỉ sử dụng các dữ liệu sơ cấp và phân tích bằng các công cụ thông thường như: SPSS, Eview, mà còn sử dụng thêm các công cụ mới có tính năng phù hợp hơn trong điều kiện mạng Internet phát triển như các công cụ: *Tableau Public*, *Power BI*, *FineReport*, *R* và *Python*. Các công cụ phân tích dữ liệu mới này được tích hợp thêm ngôn ngữ lập trình với các gói mô đun xử lý đơn giản, phù hợp với nhiều ngữ cảnh và nhiều kiểu dữ liệu khác nhau, không những đưa ra các kết

quả phân tích dữ liệu tốt, mà còn hỗ trợ trong các bài toán dự báo và hoạt động tốt trên các dữ liệu không đầy đủ.

2. Cơ sở lý thuyết

2.1. Một số khái niệm

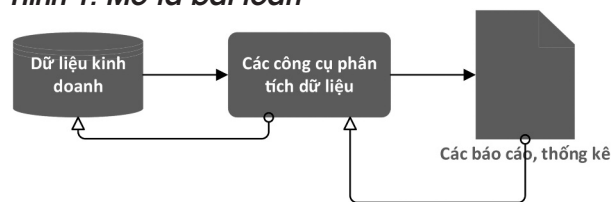
Phân tích dữ liệu: Phân tích dữ liệu (data analytics) (Oreily, (2017), là quá trình phát hiện, giải thích và truyền đạt các mô hình có ý nghĩa trong dữ liệu nhằm đưa ra các báo cáo, các bản thống kê tổng hợp theo mong muốn của người dùng. Các tổ chức, doanh nghiệp có thể áp dụng phân tích dữ liệu kinh doanh để mô tả, dự đoán và cải thiện hiệu suất kinh doanh.

Công cụ phân tích dữ liệu: Với sự ra đời của các phương tiện truyền thông xã hội (social media), các tổ chức, doanh nghiệp và người bán hàng đã có nhiều lựa chọn các kênh bán hàng, kênh truyền thông, kênh quảng bá và công cụ marketing mới thay thế cho các phương thức

truyền thống. Theo (D. Manning, et al., 2008), (Z. Abbassi, et al., 2015) và (Z. Yongzheng và P. Marco, 2013) thì các dữ liệu trên các phương tiện truyền thông xã hội thường có đặc trưng là không đầy đủ, không hoàn chỉnh, không có cấu trúc rõ ràng hoặc bởi nhiều ngôn ngữ trộn lẫn cũng theo đó xuất hiện nhiều, đòi hỏi các công cụ phân tích dữ liệu cũng có sự thay đổi và cải tiến để theo kịp các yêu cầu của người dùng.

Bài toán phân tích dữ liệu: Phân tích dữ liệu trong các tổ chức, doanh nghiệp là bài toán đã được chú trọng trong nhiều ứng dụng trong những năm gần đây. Bài toán phân tích dữ liệu không phải là bài toán mới nhưng các công cụ để phân tích dữ liệu càng ngày càng được cải tiến và đa dạng nhằm phù hợp với nhiều ngữ cảnh và nhiều kiểu dữ liệu khác nhau từ dữ liệu văn bản, dữ liệu số đến các dữ liệu đa phương tiện, các kiểu dữ liệu có yếu tố thời gian, hoặc phân tích dữ liệu theo thời gian thực. Một quy trình phân tích dữ liệu thông thường có 3 giai đoạn gồm: Tìm hiểu thu thập dữ liệu mà tổ chức, doanh nghiệp cần; Phân tích dữ liệu và phân loại dữ liệu; Cuối cùng là tạo các báo cáo và xác định các chiến lược kinh doanh dựa trên dữ liệu phân tích. Có thể mô tả bài toán phân tích dữ liệu trong các doanh nghiệp, tổ chức dựa các công cụ phân tích một cách tổng quát như Hình 1.

Hình 1: Mô tả bài toán



Trong các bài toán phân tích dữ liệu tại nghiên cứu này không tập trung vào cách thức thu thập dữ liệu và cách thức đưa ra các báo cáo và thống kê, mà còn tập trung giới thiệu các công cụ phân tích dữ liệu. Đặc biệt, các công cụ phân tích dữ liệu xã hội thu thập từ các phương tiện truyền thông xã hội.

2.2. Các nghiên cứu liên quan

2.2.1. Các nghiên cứu trong nước

Trong bài báo của tác giả Nguyễn Thanh Hải (2019) đăng trên PCW Việt Nam đề cập đến vai trò của phân tích dữ liệu trong các tổ chức, doanh nghiệp. Bài báo nhấn mạnh đến các công cụ mới,

với các gói tính năng hỗ trợ trong quá trình phân tích dữ liệu từ các dữ liệu trực quan biến đổi thành thông tin, sau đó dựa trên các công cụ phân tích dữ liệu thông minh đưa ra các thông tin hữu ích cho tổ chức, doanh nghiệp như doanh số bán hàng, sự bất thường trong doanh số kinh doanh, dự đoán các vấn đề tác động đến bất thường,... Bên cạnh đó, trong Trg (2020), Cisco Việt Nam (2019) đã nhấn mạnh đến các giải pháp Business Intelligence - Kinh doanh thông minh (BI) trong kinh doanh khi cách mạng công nghệ 4.0 và xu hướng chuyển đổi số đang bùng nổ. Các báo cáo này cho thấy kết quả phân tích dữ liệu kinh doanh, đặc biệt phân tích dữ liệu khách hàng ngày càng đóng vai trò quan trọng trong quá trình tồn tại và phát triển của các tổ chức, doanh nghiệp. Nghiên cứu của Nguyễn Anh Duy và Nguyễn Phúc Quỳnh Như (2019) giới thiệu đến cách thức khai thác dữ liệu trực tuyến với trường hợp của Amazone. Thông qua cách triển khai các nền tảng công nghệ của Big Data, cộng với sự thấu hiểu khách hàng, Amazon đã và đang đạt được các lợi thế trong ngành kinh doanh bán lẻ trực tuyến.

2.2.2. Nghiên cứu ở nước ngoài

Nghiên cứu của Allahyari, Mehdi, et al. (2017) đưa ra một thống kê tổng quát các cách phân tích và các kỹ thuật trích chọn dữ liệu đặc trưng đối với dữ liệu văn bản trên các phương tiện truyền thông xã hội. Trong đó chia thành 2 nhóm văn bản quy chuẩn và văn bản không quy chuẩn. Trong các nghiên cứu của Akhtar et al., (2020); Mandava, Geetha Bhargava. (2018); N. Couldry, and J. Turow (2014), N Balaji et al, (2021) giới thiệu các công cụ phân tích dữ liệu thông minh được ứng dụng trong các bài toán cụ thể, như: các bài toán dự báo, các bài toán cho phân tích ra kết quả dựa trên nhu cầu, các bài toán xác định mô hình hóa dữ liệu cho các tổ chức, doanh nghiệp. Nghiên cứu của D. Manning, et al., (2008); Z. Abbassi, et al., (2015); Z. Yongzheng and P. Marco, (2013) đưa ra các cách trích chọn thông tin đặc trưng trên các phương tiện truyền thông xã hội như website, cổng thông tin giải trí, trang mạng xã hội, các nền tảng bán hàng trực tuyến,...

Như vậy, không chỉ có các nghiên cứu về cách phân tích dữ liệu, mà còn có các nghiên cứu về các công cụ phân tích, thuật toán ứng dụng trong phân tích và so sánh chúng khi thực hiện trên dữ

liệu đặc biệt, dữ liệu trên các phương tiện truyền thông xã hội. Vì vậy, các công cụ phân tích dữ liệu thông minh đang được xem là một xu hướng ứng dụng mới trong các bài toán phân tích dữ liệu, đặc biệt đối với hệ thống thông tin thương mại điện tử.

3. Phương pháp nghiên cứu

Nhóm nghiên cứu sử dụng kết hợp giữa phương pháp nghiên cứu định tính và định lượng. Nghiên cứu định tính: Nhóm nghiên cứu sử dụng nhiều phương pháp khác nhau để thu thập, tìm hiểu, nghiên cứu các tài liệu về công cụ phân tích dữ liệu phổ biến hiện nay. Nghiên cứu định lượng: Nhóm nghiên cứu sử dụng một số bộ dữ liệu để tiến hành thực nghiệm và minh họa nhằm so sánh giữa các công cụ phân tích dữ liệu trực tuyến trong nghiên cứu. Sau đó dựa trên các nghiên cứu định tính và định lượng, nhóm nghiên cứu sử dụng phương pháp luận khoa học để đề xuất lựa chọn kỹ thuật phù hợp cho phân tích dữ liệu khách hàng trực tuyến trong các hệ thống thông tin thương mại điện tử của các tổ chức, doanh nghiệp.

4. Kết quả nghiên cứu

4.1. Một số công cụ phân tích dữ liệu hiện nay

Công cụ lập trình R: Lập trình R là một công cụ rất mạnh cho học máy, thống kê và phân tích dữ liệu. Ưu điểm của R là phân tích được hầu hết các kiểu dữ liệu, rất phù hợp cho việc mô hình hóa dữ liệu, thao tác khá đơn giản, dễ dàng với nhiều kiểu dữ liệu khác nhau đặc biệt trong việc mô hình hóa dữ liệu.

Công cụ Tableau: Tableau là công cụ thực hiện các nghiệp vụ phân tích một cách nhanh chóng, đơn giản và trực quan dành cho tất cả mọi người. Ưu điểm của Tableau là có phiên bản miễn phí và dễ dàng tương tác với bất kỳ loại dữ liệu nào từ các công cụ văn phòng như Excel, Data Warehouse cho tới Dữ liệu trực tuyến trên Website hoặc các phương tiện truyền thông xã hội. Đặc biệt, Tableau có khả năng cập nhật hệ thống dữ liệu đầu vào theo thời gian thực, trực quan hóa dữ liệu bằng nhiều hình thức như các biểu đồ hay thậm chí là cả một Dashboard, có thể hỗ trợ phân tích dữ liệu lớn Big Data.

Python: Python là một ngôn ngữ lập trình theo kịch bản (scripting) rất phổ biến hiện nay và hết sức thú vị. Ưu điểm là dễ học, dễ viết mã, dễ duy trì và

được cung cấp dưới dạng mã nguồn mở miễn phí cho học tập và nghiên cứu. Các thư viện học máy phong phú và hỗ trợ rất tốt trong xử lý dữ liệu văn bản như các thư viện: Scikitlearn, Theano, Tensorflow và Keras.

Power BI: Power BI là công cụ “phân tích kinh doanh” phân tích dữ liệu chia sẻ thông tin chi tiết. Kết nối dữ liệu dễ dàng, hiển thị nhanh trên Dashboard - bảng điều khiển, Reports. Power BI là tập hợp nhiều ứng dụng và connectors. Biến các nguồn dữ liệu không liên quan, thành thông tin chi tiết mạch lạc, trực quan và tương tác. Nguồn dữ liệu đa dạng, có thể giản tệp Excel hoặc bảng của website, đến Azure hoặc AWS.

Excel Advance: Excel nằm trong bộ công cụ văn phòng Microsoft Office gồm nhiều phần mềm hỗ trợ viết văn bản, thuyết trình, quản lý email hay bảng tính như Excel. Mặc dù tưởng chừng đơn giản nhưng Excel giúp tạo ra các bảng tính, cùng các tính năng, công cụ hỗ trợ người dùng tính toán dữ liệu nhanh, chính xác với số lượng hàng triệu ô tính.

4.2. Các bài toán ứng dụng

Ứng dụng trong hệ thống khuyến nghị: Hệ thống khuyến nghị (Recommender System - RS), hay còn gọi là hệ thống tư vấn là một hệ thống lọc thông tin nhằm dự đoán đánh giá sở thích, mối quan tâm, nhu cầu của người dùng để đưa ra một hoặc nhiều mục, sản phẩm, dịch vụ mà người dùng có thể sẽ quan tâm với xác suất lớn nhất. Trong những năm gần đây, hệ thống khuyến nghị đã trở nên phổ biến và được sử dụng trong nhiều lĩnh vực khác nhau như truyền hình, tin tức, dịch vụ tài chính, viễn thông, thương mại điện tử và mạng xã hội,...

Ứng dụng trong phân nhóm khách hàng: Thị trường cạnh tranh, đời sống cải thiện, sản phẩm/dịch vụ phát triển đa dạng, điều này tạo ra những sự khác biệt trong nhu cầu của mỗi người. Chính vì thế cần hướng đến phân loại ra các nhóm khách hàng, doanh nghiệp sẽ dễ dàng trong khâu quản lý khách hàng nhằm thỏa mãn được sự hài lòng của khách hàng. Phân loại khách hàng chính là nắm bắt những đặc điểm chung trong dữ liệu khách hàng, thống kê những đặc tính tương đồng của những khách hàng trước và sau khi mua/sử dụng dịch vụ/sản phẩm của doanh nghiệp, từ đó chia các đối tượng khách hàng thành những nhóm nhỏ.

4.3. Một số kết quả so sánh thực nghiệm

Để thực hiện so sánh các công cụ phân tích dữ liệu, nhóm nghiên cứu đã sử dụng một bộ dữ liệu thực của một hệ thống siêu thị và áp dụng 3 công cụ để so sánh là Excel Advance, Power BI và Python.

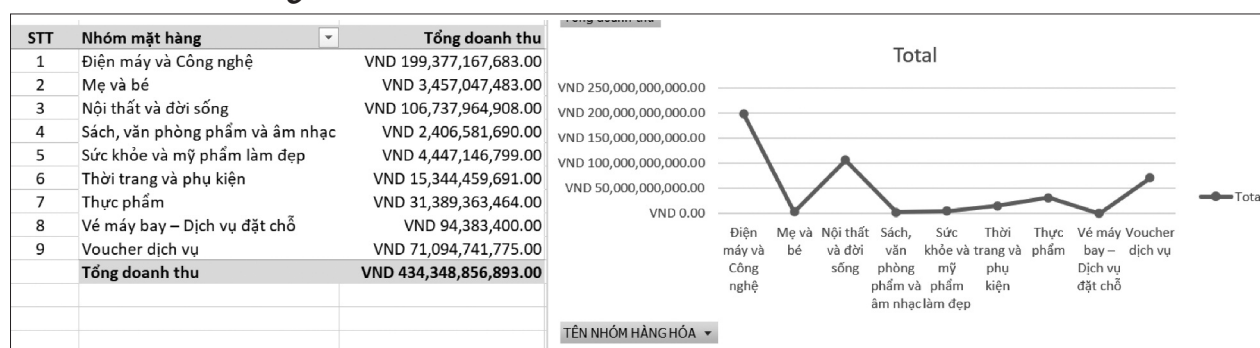
Bộ dữ liệu mẫu:

Bộ dữ liệu mẫu gồm thông tin về tình trạng bán hàng và thống kê theo các hóa đơn, bao gồm tổng cộng 92.000 hóa đơn của khách hàng từ 9 nhóm mặt hàng của công ty, bao gồm: Điện máy và

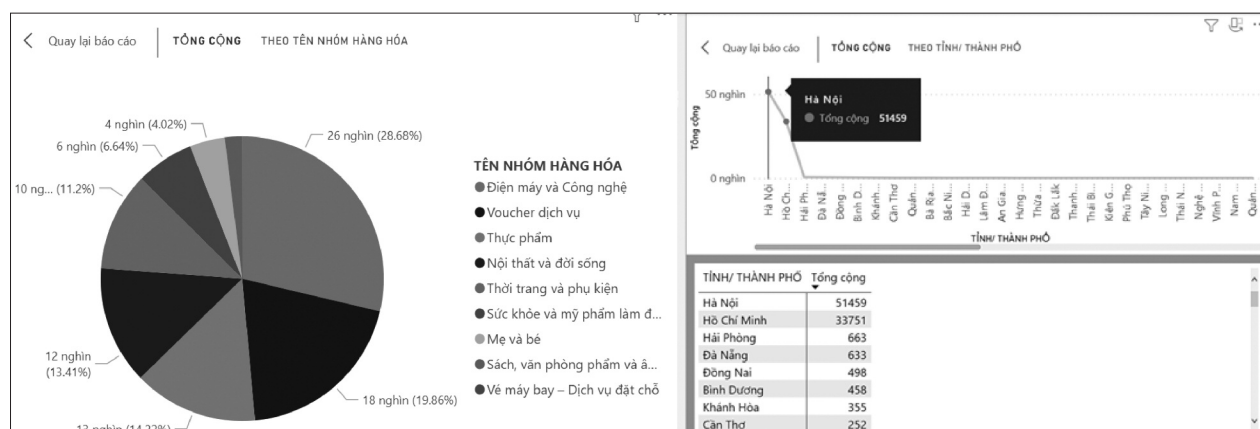
Công nghệ; Mẹ và bé; Nội thất và đời sống; Sách, văn phòng phẩm và âm nhạc; Sức khỏe và mỹ phẩm làm đẹp; Thời trang và phụ kiện; Thực phẩm; Vé máy bay - Dịch vụ đặt chỗ; Voucher dịch vụ. Mục tiêu của chúng tôi là phân tích dữ liệu để đưa ra dự báo về nhóm mặt hàng, về xu hướng, hoặc phân loại khách hàng cho tổ chức. (Hình 2, 3, 4)

Bộ dữ liệu mẫu gồm thông tin về tình trạng bán hàng và thống kê theo các hóa đơn, bao gồm tổng

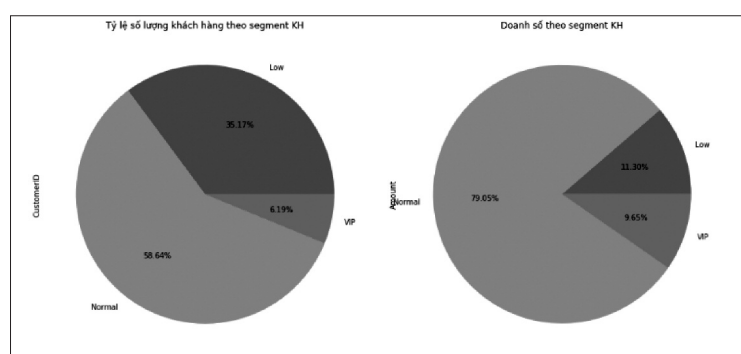
Hình 2: Phân tích bằng Excel Advance



Hình 3: Phân tích bằng Power BI



Hình 4: Phân tích bằng Python



cộng 92.000 hóa đơn của khách hàng từ 9 nhóm mặt hàng. Nghiên cứu cho thấy: Nếu chỉ so sánh về số liệu cố định, cả 3 công cụ đều có thể phân tích, vẽ biểu đồ và thực hiện các phép tính trung gian như nhau. Tuy nhiên, nếu muốn xem xét trên cùng giao diện và thay đổi các tham số như thị trường (theo tỉnh/ thành phố) thì Power BI cho kết quả trực quan sinh động, dễ dàng cập nhật và cho giao

Bảng 1. So sánh 3 công cụ

Nội dung	Excel Advance	Power BI	Python
Phân tích trực quan	Tốt	Rất tốt	Tốt
Tính dễ sử dụng	Rất tốt	Tốt	Thấp
Biểu diễn sự phức tạp, nhiều tham số	Thấp	Khá cao	Cao
Biểu diễn biểu đồ, đồ thị, mô tả	Tốt	Rất tốt	Khá
Hỗ trợ truy vấn lồng nhau	Không	Có nhưng yếu	Tốt
Bài toán phân loại	Phân loại đơn chỉ có phép AND	Phân loại đơn, phân loại kép, có thể thực hiện AND, OR, NOT	Phân loại đơn, phân loại kép, có thể thực hiện AND, OR, NOT và các phép toán phân cụm dựa trên học máy
Bài toán dự báo	Dự báo dựa trên hàm có sẵn	Dự báo dựa trên mô hình	Dự báo theo học máy và tùy biến theo yêu cầu người dùng
Bài toán sắp xếp	Thủ công	Thủ công và xây dựng trên hàm	Dựa trên các hàm và có thể sắp xếp theo cụm hoặc theo các nhóm với yêu cầu khác nhau.

diện kết nối thông minh hơn Excel Advance. Tuy nhiên, để phân cụm và đưa ra mô hình dự báo tốt nhất là Python với các thư viện hỗ trợ rất phong phú từ xử lý dữ liệu, phân tích văn bản, phân tích số liệu đến đưa ra các mức độ. Nhược điểm của Python là khó sử dụng (phải có kiến thức về lập trình), biết cài đặt thêm các thư viện. So sánh của 3 công cụ chi tiết như Bảng 1.

5. Kết luận

Nghiên cứu với mục tiêu nghiên cứu, tìm hiểu,

so sánh một số công cụ phân tích dữ liệu thông minh ứng dụng trong các hoạt động sản xuất, kinh doanh của các doanh nghiệp, tổ chức hiện nay, như: Tableau Public, Power BI, FineReport, R và Python,... nhằm đưa ra một số hàm ý ứng dụng khi sử dụng công cụ phân tích dữ liệu trong hoạt động trên thực tế. Qua kết luận thấy rằng, có nhiều công cụ hữu ích cho phân tích dữ liệu, tuy nhiên mỗi công cụ có một mức độ ảnh hưởng và ứng dụng trên các bài toán khác nhau ■

TÀI LIỆU THAM KHẢO:

1. Nguyễn Anh Duy, Nguyễn Phúc Quỳnh Như (2019). Dữ liệu lớn: Cách thức khai thác cơ hội từ dữ liệu? Trường hợp Amazon. *Phát triển và Hội nhập*, số 46 (56), tháng 05 - 06/2019.
2. Phan Thanh Đức và các tác giả (2019). Ứng dụng dữ liệu lớn trong hoạt động quản trị quan hệ khách hàng tại các ngân hàng thương mại Việt Nam. *Tạp chí Khoa học và Đào tạo Ngân hàng*, số 203, tháng 4/2019.
3. Oreily (Trần Mạnh Hoàng dịch) (2017). *Phân tích dữ liệu tinh gọn*. NXB Công Thương.
4. Lê Triệu Tuấn, Lý Thu Trang, (2020). Tổng quan dữ liệu lớn trong thương mại điện tử. *Tạp chí Khoa học và Công nghệ - Đại học Thái Nguyên*, 225(06): 536 - 540, 202.
5. Trg (2020). Áp dụng hiệu quả giải pháp BI vào các hoạt động kinh doanh. Truy cập tại: trginternational.com/solutions.
6. D. Manning, et al., (2008). *Introduction to Information Retrieval*. New York, USA: Cambridge University Press, ISBN: 0521865719, 9780521865715, 2008.

7. Mandava, Geetha Bhargava (2018). Analysis and Design of Visualization of Educational Institution Database using Power BI Tool, Software & Data Engineering. *Global Journal of Computer Science and Technology*, Vol 1(4), Ver 1.0, 2018.

8. Z. Yongzheng and P. Marco (2013). *Predicting Purchase Behaviors from Social Media*. in *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, Rio de Janeiro, Brazil, 2013.

Ngày nhận bài: 11/1/2022

Ngày phản biện đánh giá và sửa chữa: 11/2/2022

Ngày chấp nhận đăng bài: 21/2/2022

Thông tin tác giả:

1. TS. NGUYỄN THỊ HỘI

2. ThS. BUI QUANG TRƯỜNG

Khoa Hệ thống thông tin Kinh tế và Thương mại điện tử

Trường Đại học Thương mại

SOME INTELLIGENT DATA ANALYSIS TOOLS AND THE USE OF THESE TOOLS TO SOLVE ECONOMIC PROBLEMS

● Ph.D **NGUYEN THI HOI**¹

● Master. **BUI QUANG TRUONG**¹

¹Faculty of Economic Information Systems and E-commerce
Thuongmai University

ABSTRACT:

The rapid development in information technologies and the use of social media networks in business activities have made the data analysis increasingly important in businesses and organizations. This paper introduces and compares some intelligent data analysis tools and problems that are applied in practice today. This paper has three parts in order to generally introduce the problem of data analysis in businesses and organizations, present some popular data analysis tools, and present comparisons and make some implications for the use of intelligent data analysis tools in production and business activities of enterprises, organizations and state agencies.

Keywords: data analysis, analytical tools, business data, forecasting problems.