

Apply big data analysis in substantive test in financial audit

Nguyen Vinh Khuong*, Nguyen Ngoc Tue Minh

University of Economics and Law, Vietnam National University HCMC

Received: 30/05/2019; Accepted: 04/12/2019

ABSTRACT

Big data has a significant impact on auditing and assurance service ensuring that companies are operating and generating more data than ever before. With the smart use of the latest technologies, along with the knowledge and experience available, auditors can achieve a better insight into the overall financial situation of customers. Hence, auditor can make better decisions, improve audit quality and ultimately create value for customers. The purpose of the paper is on the application of big data analysis at the stage of auditing, in particular, substantive test including analytical procedures and test of details of balances. In particular, the article discussed some common data analysis techniques and patterns to describe more precisely how to utilise in practice.

Keywords: *Big data, substantive test, auditing.*

**Corresponding author.*

Email: khuongnv@uel.edu.vn

Vận dụng phân tích dữ liệu lớn trong thử nghiệm cơ bản trong kiểm toán báo cáo tài chính

Nguyễn Vĩnh Khương*, Nguyễn Ngọc Tuệ Minh

Trường Đại học Kinh tế-Luật, ĐHQG TP.HCM

Ngày nhận bài: 30/05/2019; Ngày nhận đăng: 04/12/2019

TÓM TẮT

Dữ liệu lớn cũng ảnh hưởng đáng kể tới ngành nghề kiểm toán và dịch vụ đảm bảo khi các công ty đang vận hành và tạo ra nhiều dữ liệu hơn bao giờ hết. Với việc sử dụng một cách thông minh các công nghệ mới nhất, cùng với kiến thức và kinh nghiệm có sẵn, các kiểm toán viên có thể đạt được một cái nhìn sâu sắc hơn về toàn cảnh tình hình tài chính của khách hàng, từ đó đưa ra quyết định tốt hơn, nâng cao chất lượng kiểm toán và cuối cùng tạo ra giá trị cho khách hàng. Mục tiêu bài báo về vận dụng phân tích dữ liệu lớn vào giai đoạn thực hiện kiểm toán, cụ thể là thử nghiệm cơ bản bao gồm thủ tục phân tích và thử nghiệm chi tiết. Trong đó, bài báo đề cập đến một số kỹ thuật phân tích dữ liệu lớn thường gặp và ví dụ minh họa nhằm mô tả rõ ràng hơn cách thức vận dụng trong thực tiễn.

Từ khóa: Dữ liệu lớn, thử nghiệm cơ bản, kiểm toán.

1. ĐẶT VẤN ĐỀ

Phân tích dữ liệu để tạo ra giá trị kinh doanh không phải là một khái niệm mới, tuy nhiên, việc hiểu các dữ liệu một cách nhanh chóng và đưa ra các quyết định kinh doanh dựa trên kết quả phân tích dữ liệu thì ngày càng trở nên quan trọng hơn bao giờ hết. Các doanh nghiệp có khả năng tận dụng được lợi thế của dữ liệu lớn đều có thể thích ứng nhanh hơn với các xu hướng kinh doanh mới, thay đổi nhu cầu của khách hàng và hoạt động hiệu quả hơn. Điều này giúp nâng cao sự hài lòng của khách hàng và dẫn tới sự thành công của doanh nghiệp.

Thật vậy, dữ liệu lớn sẽ đem lại rất nhiều lợi ích đáng kể nếu chúng được sử dụng và kiểm soát tốt. Các chương trình này hỗ trợ trong việc hợp nhất và tiêu thụ khối lượng lớn dữ liệu có cấu trúc và phi cấu trúc, tạo cơ hội để phân tích và hiểu biết nhanh chóng. Trong khi trước đây, việc phân tích dữ liệu ngay lập tức như vậy có thể

là bất khả thi hoặc có thể tốn nhiều ngày, thậm chí hàng tuần để thực hiện. Bằng cách sử dụng thông tin chi tiết từ dữ liệu lớn, doanh nghiệp có thể đưa ra các quyết định đúng đắn, nhằm vào khách hàng mới theo những cách sáng tạo và khác biệt, phục vụ cho các khách hàng hiện tại với mô hình phân phối được cải tiến cho từng cá nhân và cung cấp những dịch vụ mới, phân biệt mình với các đối thủ cạnh tranh. Những doanh nghiệp có thể tận dụng các cơ hội mà dữ liệu lớn mang lại có thể phát triển lợi thế cạnh tranh lâu dài. Ngoài ra, dữ liệu lớn có thể nâng cao tính minh bạch của doanh nghiệp, cải thiện việc phân tích quản lý và báo cáo, và giảm chi phí quản lý.

Dữ liệu lớn cũng ảnh hưởng đáng kể tới ngành nghề kiểm toán và dịch vụ đảm bảo khi các doanh nghiệp, công ty đang vận hành và tạo ra nhiều dữ liệu hơn bao giờ hết. Với việc sử dụng một cách thông minh các công nghệ mới nhất, cùng với kiến thức và kinh nghiệm có sẵn,

*Tác giả liên hệ chính:

Email: khuongnv@uel.edu.vn

các kiểm toán viên có thể đạt được một cái nhìn sâu sắc hơn về toàn cảnh tình hình tài chính của khách hàng, từ đó đưa ra quyết định tốt hơn, nâng cao chất lượng kiểm toán và cuối cùng tạo ra giá trị cho khách hàng của họ. Cốt lõi của những thay đổi này là do khối lượng dữ liệu ngày càng tăng mà kiểm toán viên cần phải xử lý. Ngoài khối lượng tuyệt đối, các dữ liệu này còn có nhiều dạng khác nhau như văn bản, hình ảnh, âm thanh và video đòi hỏi dung lượng lưu trữ lớn (Mohanty và cộng sự, 2013). Các quy định pháp lý hiện hành, các chuẩn mực nghề nghiệp và các thói quen hành nghề dường như ngày càng tụt hậu trong một môi trường chứa nhiều thông tin và thay đổi nhanh chóng như thế này. Điều này khiến các công ty kiểm toán ngày càng khó khăn hơn trong việc phục vụ nhu cầu của khách hàng và có khả năng đe dọa sự tồn tại lâu dài của công ty ở hiện tại.

Tuy những tác động của dữ liệu lớn tới nền kinh tế nói chung và ngành nghề kiểm toán và dịch vụ đảm bảo nói riêng đã thể hiện rõ tại nhiều nơi trên thế giới, nhưng tại Việt Nam hiện nay lại có rất ít bài viết về cách thức vận dụng phân tích dữ liệu lớn trong bối cảnh kiểm toán và dịch vụ đảm bảo. Vì vậy, tác giả tổng hợp, phân tích cách thức vận dụng phân tích dữ liệu lớn trong kiểm toán báo cáo tài chính, cụ thể ở thử nghiệm cơ bản nhằm đáp ứng nhu cầu cấp thiết trên.

2. KHÁI NIỆM MỘT SỐ KỸ THUẬT DÙNG ĐỂ PHÂN TÍCH DỮ LIỆU LỚN

2.1. Cluster

Clustering (phân cụm) là quá trình phân bổ các đối tượng thành các nhóm có thành phần tương tự nhau thông qua việc tìm ra ranh giới giữa các nhóm này, bằng cách sử dụng một số thuật toán và phương pháp thống kê khác nhau. Phân tích cluster không tạo ra bất kỳ sự phân biệt nào giữa các biến phụ thuộc và độc lập. Phân tích cluster cho phép kiểm tra toàn bộ tập hợp các mối quan hệ phụ thuộc lẫn nhau để tìm ra mối quan hệ tương đồng giữa các đối tượng để xác định ra các cụm. Phân tích cluster cũng có

thể được sử dụng như một phương pháp giảm kích thước, trong đó số lượng đối tượng được nhóm thành một tập hợp các cụm, giúp làm giảm số biến lại và dùng cho mô hình dự đoán. Kỹ thuật phân tích này giúp giảm các vấn đề liên quan đến tính đa cộng tuyến (Đa cộng tuyến là một hiện tượng khá phổ biến trong các vấn đề của mô hình hồi quy đa biến. Đây là hiện tượng các biến giải thích có tương quan với các biến khác có trong mô hình. Đa cộng tuyến có thể làm cho kết quả nghiên cứu bị sai lệch ở nhiều khía cạnh khác nhau). Clustering có thể được nhóm thành ba loại: unsupervised, semisupervised và supervised.

2.2. K-means

K-means là một trong những kỹ thuật clustering (phân cụm) được sử dụng rộng rãi nhất vì tính đơn giản và tốc độ của nó. Nó phân vùng dữ liệu thành các cụm bằng cách gán từng đối tượng cho trung tâm cụm gần nhất của nó (trung tâm cụm là giá trị trung bình của các biến cho tất cả các đối tượng trong cụm đó) dựa trên thước đo khoảng cách được sử dụng. Ngoài ra, k-means hoạt động nhanh chóng hơn cho các tập dữ liệu lớn, thứ rất phổ biến hiện nay. Thuật toán cơ bản cho k-means hoạt động như sau:

1. Chọn số cụm, đặt là k .
2. Chọn k điểm làm trung tâm cụm (ví dụ: các đối tượng k được chọn ngẫu nhiên từ tập dữ liệu).
3. Gán từng đối tượng cho trung tâm cụm gần nhất.
4. Tính toán lại trung tâm cụm mới.
5. Lặp lại bước 3 và 4 cho đến khi tiêu chí hội tụ được đáp ứng (ví dụ: việc gán đối tượng cho các cụm không còn thay đổi qua nhiều lần lặp) hoặc đạt được số lần lặp tối đa.

Nhiều vấn đề cần được xem xét trong phân cụm k-means:

- Thuật toán k-means yêu cầu số lượng k cụm làm đầu vào. Phương pháp ABC có thể được sử dụng để ước tính số lượng cụm.

- Nên chọn độ tương tự/khoảng cách tùy theo yêu cầu của nhiệm vụ.

- Các cụm có thể hội tụ đến mức tối thiểu cục bộ. Do vấn đề này, các cụm thu được có thể không phải là đúng. Để tránh điều này, có thể hữu ích khi chạy thuật toán với các trung tâm cụm ban đầu khác nhau và so sánh kết quả.

2.3. Decision tree (cây quyết định)

Thuật toán cây quyết định là công cụ được dùng để phân lớp dữ liệu, mỗi cây quyết định tượng trưng cho một sự quyết định của một lớp các dữ kiện nào đó. Mỗi nút trong cây là tên của một lớp hay một phép thử thuộc tính cụ thể nào đó, phép thử này phân chia không gian trạng thái các dữ kiện tại nút đó thành các kết quả có thể đạt được của phép thử. Mỗi tập con được phân chia của phép thử là không gian con của các sự kiện, nó tương ứng với một vấn đề con của sự phân lớp. Các cây quyết định được dùng để hỗ trợ quá trình ra quyết định kinh doanh.

Cây quyết định có thể định nghĩa, diễn giải bằng một tập các luật IF– THEN, với cách trình bày như vậy nó sẽ giúp cho người đọc dễ đọc và dễ hiểu. Cây quyết định có thể thực hiện được cả với các dữ liệu chứa lỗi (noisy data). Về bản chất, cây quyết định là một trong các phương pháp quy nạp được dùng phổ biến nhất trong quá trình xử lý dữ liệu. Một cách tổng thể, cây quyết định có các tính chất sau:

- Mỗi nút trong (internal node) biểu diễn một thuộc tính cần kiểm tra giá trị (attribute to be tested) đối với các các tập thuộc tính.

- Nút lá (leaf node) hay còn gọi là nút trả lời biểu thị cho một lớp các trường hợp mà nhãn của nó là tên của lớp, nó biểu diễn một lớp (a classification).

- Nút nhánh (branch) từ một nút sẽ tương ứng với một giá trị có thể của thuộc tính gắn với nút đó.

- Nhãn (label) của nút này là tên của thuộc tính và có một nhánh nối nút này đến các cây con ứng với mỗi kết quả có thể có phép thử. Nhãn của nhánh này là các giá trị của thuộc tính đó. Nút trên cùng gọi là nút gốc.

2.4. Artificial Neuron Network (ANN)

Artificial Neural Network (ANN), hay còn gọi là mạng nơron nhân tạo, là mô hình xử lý thông tin được mô phỏng dựa trên hoạt động của hệ thống thần kinh của sinh vật, bao gồm số lượng lớn các nơron được gắn kết để xử lý thông tin. ANN giống như bộ não con người, được học bởi kinh nghiệm (thông qua huấn luyện), có khả năng lưu giữ những kinh nghiệm hiểu biết (tri thức) và sử dụng những tri thức đó trong việc dự đoán các dữ liệu chưa biết (unseen data).

Cấu trúc tổng quát của một ANN gồm 3 thành phần đó là: Input Layer, Hidden Layer và Output Layer. Trong đó, lớp ẩn (Hidden Layer) gồm các Nơron nhận dữ liệu input từ các Nơron ở lớp (Layer) trước đó và chuyển đổi các input này cho các lớp xử lý tiếp theo. Trong một ANN có thể có nhiều lớp ẩn. Trong đó các Processing Elements (PE) của ANN gọi là Nơron, mỗi Nơron nhận các dữ liệu vào (Inputs) xử lý chúng và cho ra một kết quả (Output) duy nhất. Kết quả xử lý của một Nơron có thể làm Input cho các Nơron khác.

3. VẬN DỤNG PHÂN TÍCH DỮ LIỆU LỚN VÀO THỦ TỤC PHÂN TÍCH

Phân tích dữ liệu lớn là (những) kỹ thuật phân tích dữ liệu có thể được sử dụng để thực hiện các bước của một cuộc kiểm toán như đánh giá rủi ro, thử nghiệm kiểm soát và thủ tục cơ bản (gồm phân tích cơ bản và kiểm tra chi tiết). Trong phạm vi bài viết này, tác giả sẽ dựa trên những tài liệu, công trình nghiên cứu trước đây về phân tích dữ liệu lớn để trình bày những ứng dụng của phân tích dữ liệu lớn vào việc thực hiện thủ tục phân tích và thử nghiệm chi tiết. Dựa vào hướng dẫn phân tích dữ liệu kiểm toán được ban hành bởi Viện Kế toán công chứng Mỹ và các nghiên cứu liên quan về áp dụng các kỹ thuật này trong quá trình thực hiện cuộc kiểm toán (AICPA, 2017, GTAG, 2017).

3.1. Định nghĩa thủ tục phân tích

Thủ tục phân tích bao gồm việc đánh giá thông tin tài chính qua việc phân tích các mối quan hệ giữa dữ liệu tài chính và phi tài chính; kiểm tra các biến động và các mối quan hệ xác

định không nhất quán với các tài liệu, thông tin liên quan khác hoặc có sự chênh lệch lớn so với các giá trị dự kiến. Như vậy, thủ tục phân tích ngoài mục đích xác định nội dung, lịch trình và phạm vi của các thủ tục kiểm toán khác, nó còn được thực hiện như là một thử nghiệm cơ bản khi việc sử dụng thủ tục này có hiệu quả hơn so với kiểm toán chi tiết nhằm giảm bớt rủi ro phát hiện liên quan đến cơ sở dữ liệu của báo cáo tài chính.

3.2. Trình tự khi thực hiện thủ tục phân tích bằng phân tích dữ liệu lớn

Dưới đây là năm bước mà kiểm toán viên có thể sử dụng trong việc lập kế hoạch, thực hiện và đánh giá kết quả của một thủ tục phân tích cơ bản có áp dụng dữ liệu lớn

Bước 1: Lập kế hoạch thủ tục phân tích cơ bản.

- Xác định các khoản mục trên báo cáo tài chính, các tài khoản hoặc thuyết minh cần phân tích.

- Xác định các rủi ro có sai sót trọng yếu đã được đánh giá.

- Xác định các kỳ vọng của kiểm toán viên, bao gồm:

- Các biến độc lập (dự đoán) sẽ được sử dụng.

- Nguồn dữ liệu cho các biến đó.

- Mức độ phân tổ của dữ liệu.

- Xác định độ chính xác mong muốn của kiểm toán viên.

- Xác định ngưỡng chênh lệch có thể bỏ qua.

- Xác định thủ tục phân tích cơ bản (ví dụ: phân tích xu hướng, phân tích tỷ lệ, mô hình dự báo phi thông kê, phân tích hồi quy,...) có thể đáp ứng các mục tiêu của kiểm toán viên.

- Phát triển mô hình sẽ được sử dụng, nếu có.

- Xác định các loại bảng biểu minh họa sẽ được sử dụng.

Bước 2: Thu thập dữ liệu

Bước 3: Đánh giá độ tin cậy của dữ liệu.

Xem xét đến nguồn dữ liệu, tính so sánh, bản chất và mức độ phù hợp của thông tin và kiểm soát việc chuẩn bị dữ liệu.

Bước 4: Thực hiện thủ tục phân tích cơ bản

- Thiết lập kỳ vọng của kiểm toán viên về dữ liệu ghi nhận hoặc tỷ lệ được tính toán.

- Đánh giá xem kỳ vọng của kiểm toán viên có đủ chính xác không, nếu không thì cần có các phương án bổ sung để tăng độ chính xác.

- Thực hiện quy trình phân tích cơ bản và so sánh với kỳ vọng của kiểm toán viên

Bước 5: Đánh giá kết quả của thủ tục phân tích cơ bản.

- Xác định xem sự khác biệt giữa kỳ vọng của kiểm toán viên và kết quả phân tích có đáng kể hay không.

- Điều tra bất kỳ sự khác biệt đáng kể, xác định nguyên nhân.

- Xác định xem liệu thủ tục phân tích cơ bản có xác định được sai sót nào không và đánh giá sai sót đó.

- Kết luận xem thủ tục phân tích đã đạt được mục tiêu hay chưa. Nếu các mục tiêu chưa đạt được, cần lập kế hoạch và thực hiện các thủ tục khác để đạt được các mục tiêu đó.

❖ Tính hợp lý và khả năng dự đoán

Tiền đề cơ bản trong việc áp dụng các thủ tục phân tích cơ bản là mối quan hệ hợp lý giữa các dữ liệu được dự kiến sẽ tồn tại và tiếp tục. Điều quan trọng là phải xem xét những yếu tố nào làm cho mối quan hệ trở nên hợp lý bởi vì có nhiều loại dữ liệu khác nhau, đôi khi dường như chúng có vẻ liên quan nhưng thực sự thì lại là không. Điều này có thể dẫn đến những kết luận sai lầm của kiểm toán viên. Ngoài ra, sự hiện diện của một mối quan hệ bất ngờ có thể cung cấp bằng chứng quan trọng khi được xem xét kỹ lưỡng.

Sự hiểu biết của kiểm toán viên về doanh nghiệp và môi trường của doanh nghiệp sẽ rất hữu ích trong việc giúp xác định sự tồn tại và tiếp tục của các mối quan hệ giữa các dữ liệu. Các mối quan hệ càng dễ dự đoán thì kỳ vọng của kiểm toán viên sẽ càng chính xác. Sau đây là các yếu tố kiểm toán viên có thể xem xét khi dự đoán số về các mối quan hệ:

- Bản chất của các số dư khoản mục (ví dụ: tài khoản bao gồm ước tính kế toán hoặc là lũy kế số phát sinh thực tế được).
- Sự kết hợp các sản phẩm bán ra.
- Hồ sơ doanh nghiệp (ví dụ: số lượng, kích cỡ và vị trí của các đơn vị sản xuất).
- Quyết định của ban quản lý (ví dụ: trong việc lập dự toán, mua tài sản cố định).
- Sự ổn định của môi trường mà doanh nghiệp hoạt động.
- Khoản mục là một tài khoản trên bảng cân đối kế toán hay thuộc báo cáo kết quả hoạt động kinh doanh.

Ví dụ: kỳ vọng được hình thành cho tài khoản trên báo cáo kết quả hoạt động kinh doanh có xu hướng chính xác hơn kỳ vọng cho tài khoản trên bảng cân đối kế toán vì các mối quan hệ trên báo cáo thu nhập thường dễ dự đoán hơn, vì chúng xảy ra trong một khoảng thời gian, thay vì tại một thời điểm. Ngoài ra, kỳ vọng hình thành trong điều kiện kinh tế ổn định (ví dụ: lãi suất ổn định) hoặc các yếu tố môi trường ổn định (ví dụ: không có thay đổi về quy định) có xu hướng chính xác hơn kỳ vọng hình thành trong nền kinh tế hoặc môi trường không ổn định.

❖ Mức độ phân tán của dữ liệu

Mức độ phân tán của dữ liệu liên quan đến việc các số dư tài khoản, giao dịch hoặc các nhóm khác được kết hợp để sử dụng trong thủ tục phân tích cơ bản. Sau đây là một vài ví dụ:

- Dữ liệu hàng năm có thể được phân chia thành cấp độ hàng tháng hoặc hàng tuần.
- Doanh thu và chi phí bán hàng có thể được phân chia theo dòng sản phẩm.

- Dữ liệu cho toàn bộ nhà máy có thể được phân chia theo vị trí của phân xưởng.

Một sai sót trọng yếu có thể dễ dàng được xác định thông qua việc sử dụng dữ liệu được phân tách. Ví dụ, rủi ro mà sai sót trọng yếu có thể bị che khuất bởi các yếu tố bù đắp tăng lên khi hoạt động của một doanh nghiệp trở nên phức tạp và đa dạng hơn. Phân tán dữ liệu sẽ giúp làm giảm rủi ro này. Ngoài ra, kỳ vọng của kiểm toán viên có thể sẽ chính xác hơn khi dữ liệu phân tách được sử dụng. Kiểm toán viên cũng có thể xem xét dữ liệu được phân tách theo các cách khác nhau (ví dụ: doanh thu cho thuê theo tháng và theo tài sản cho thuê) vì trong một số trường hợp, điều này có thể làm tăng khả năng xác định sai sót trọng yếu.

Mặt khác, việc phân tán dữ liệu có thể đưa ra các vấn đề để kiểm toán viên xem xét. Trong một số trường hợp, các kiểm soát của một doanh nghiệp có thể hoạt động hiệu quả cho dữ liệu ở mức tổng hợp cao nhưng lại không hiệu quả đối với dữ liệu phân tách. Ví dụ: dữ liệu hàng quý có thể không chịu sự kiểm soát giống như dữ liệu hàng năm. Hơn nữa, dữ liệu giữa niên độ thường sẽ không được kiểm toán mà chỉ được soát xét.

3.3. Minh họa trường hợp công ty thuộc lĩnh vực sản xuất

Kiểm toán viên đang kiểm toán báo cáo tài chính của một doanh nghiệp tư nhân sở hữu và quản lý 10 khu chung cư. Có 1.200 căn hộ có kích thước khác nhau và nằm ở các khu vực khác nhau của cùng một thành phố. Doanh thu cho thuê hàng năm từ các căn hộ này trung bình khoảng 14 triệu đô la mỗi năm trong vài năm qua.

❖ Lập kế hoạch thủ tục phân tích cơ bản.

• Xác định các khoản mục và cơ sở dẫn liệu liên quan

- Thủ tục phân tích cơ bản này sẽ là một nguồn bằng chứng kiểm toán liên quan đến doanh thu cho thuê. Các cơ sở dẫn liệu được xác định như sau:

- Phát sinh. Tất cả các giao dịch doanh thu cho thuê được ghi nhận đã xảy ra và liên quan đến công ty.

- Đầy đủ. Tất cả các giao dịch cho thuê cần được ghi lại đã được ghi lại.

- Chính xác. Các khoản tiền và dữ liệu khác liên quan đến các giao dịch cho thuê đã được ghi lại đúng.

- Đúng kỳ. Giao dịch cho thuê đã được ghi nhận trong đúng kỳ.

• Các rủi ro có sai sót trọng yếu đã được xác định

Thủ tục phân tích cơ bản này được thiết kế để xử lý với mức độ rủi ro sai lệch vừa phải của doanh thu cho thuê. Mức độ đảm bảo được cung cấp bởi thủ tục này bị ảnh hưởng bởi các vấn đề sau:

- Kết quả của công việc kiểm toán được thực hiện trong một môi trường kiểm soát tốt. Không có bất kỳ dấu hiệu nào về động cơ hoặc xu hướng của chủ sở hữu hoặc người quản lý cố tình làm sai lệch doanh thu cho thuê.

- Kết quả của các thủ tục liên quan đến việc thiết kế các kiểm soát của công ty đối với doanh thu cho thuê và việc thực hiện các kiểm soát đó.

- Kết quả của các thủ tục kiểm toán khác trực tiếp hoặc gián tiếp cung cấp bằng chứng kiểm toán liên quan đến doanh thu cho thuê, ví dụ như xác minh các khoản thu tiền mặt từ người thuê, gửi thư xác nhận cho người thuê về khoản phải thu và các điều khoản của hợp đồng cho thuê.

• Kỳ vọng của kiểm toán viên

Kỳ vọng của kiểm toán viên là ước tính doanh thu cho thuê cho mỗi tháng được tổng hợp để cung cấp ước tính doanh thu cho thuê trong năm được kiểm toán. Các biến độc lập (dự đoán) được sử dụng để thực hiện ước tính này bao gồm:

- Số lượng căn hộ trong mỗi 10 khu chung cư của doanh nghiệp .

- Kích thước và số phòng của các căn hộ.

- Ngày hết hạn của hợp đồng thuê, đặc biệt là ngày hết hạn trong năm hiện tại.

- Giá cho thuê trung bình hàng tháng trên thị trường mà doanh nghiệp hoạt động

• Mức độ chính xác mong muốn của kỳ vọng và các sai lệch đáng kể

Độ chính xác mong muốn cho thủ tục phân tích cơ bản này là mức trọng yếu thực hiện được xác định bởi kiểm toán viên bằng cách sử dụng xét đoán chuyên môn.

Sự khác biệt giữa kỳ vọng của kiểm toán viên về doanh thu cho thuê trong một tháng và số doanh thu cho thuê được ghi nhận trong một tháng được coi là đáng kể nếu chênh lệch đó đơn lẻ hoặc khi tổng hợp với các sai sót khác, vượt quá mức trọng yếu thực hiện.

• Loại kỹ thuật phân tích được sử dụng

Một mô hình ANN (mạng lưới nơron nhân tạo) được sử dụng để phát triển kỳ vọng về doanh thu cho thuê của kiểm toán viên. Kiểm toán viên sử dụng bộ dữ liệu về doanh thu của các năm trước cho quá trình học tập (learning process) để tìm ra xu hướng và từ đó xây dựng mô hình dự đoán về doanh thu năm nay.

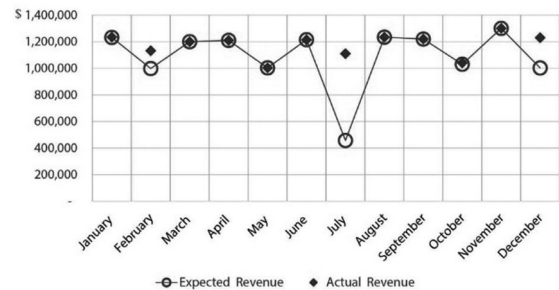
Kiểm toán viên hình thành kỳ vọng về số doanh thu hàng tháng từ mỗi tòa nhà chung cư của công ty. Các khu chung cư có kích thước và số lượng phòng tương tự được nhóm lại. Số lượng căn hộ trong mỗi nhóm được nhân với tỷ lệ cho thuê trên thị trường trung bình hàng tháng cho loại căn hộ đó. Việc tính toán được điều chỉnh bởi các yếu tố phản ánh tỷ lệ trống trung bình hàng tháng trên thị trường và ngày hết hạn của hợp đồng thuê. Các tính toán được tổng hợp sẽ biểu thị tổng doanh thu cho thuê dự kiến theo tháng và tổng doanh thu hàng năm cho mỗi trong số 10 khu chung cư.

• Các bảng biểu

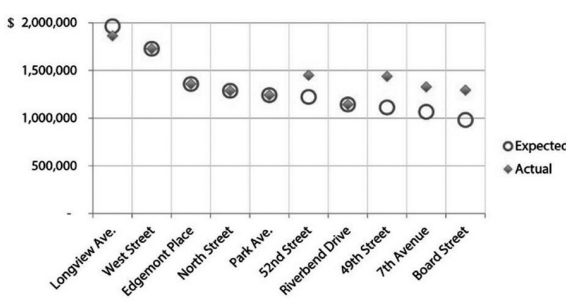
Các biểu đồ sau đây được sử dụng:

- Biểu đồ 1 so sánh tổng doanh thu dự kiến với tổng doanh thu cho thuê thực tế theo tháng. Mục đích của biểu đồ này là giúp xác định các tháng cụ thể, nếu có, để phân tích sâu hơn.

- Biểu đồ 2 cho thấy tổng doanh thu trên mỗi khu chung cư. Phân chia doanh thu theo từng khu sẽ cung cấp cho kiểm toán viên cái nhìn rõ ràng hơn về sự tồn tại của một sai sót trọng yếu. Ví dụ, nó có thể cho phép kiểm toán viên phát hiện ra những sai sót đã bị bù đắp khi doanh thu cho thuê quá mức của một khu có thể được bù đắp bằng doanh thu cho thuê cho một khu khác.



Biểu đồ 1. Doanh thu dự kiến và thực tế theo tháng



Biểu đồ 2. Doanh thu dự kiến và thực tế theo từng khu chung cư

❖ Thực hiện thủ tục phân tích cơ bản

Kiểm toán viên đã thực hiện một ANN và áp dụng mô hình dự đoán cho doanh thu cho thuê theo kế hoạch. Như thể hiện trong biểu đồ 1, doanh thu cho thuê trong các tháng 2, 7 và 12 được tìm thấy cao hơn dự kiến. Biểu đồ 2 lại cho thấy một vài khu chung cư cụ thể có doanh thu cho thuê vượt quá mong đợi. Những khác biệt so với kỳ vọng được coi là đáng kể và do đó cần được tìm hiểu sâu hơn.

Kiểm toán viên thực hiện các thủ tục tiếp theo để điều tra sự khác biệt đáng kể so với kỳ vọng. Các thủ tục này bao gồm phỏng vấn quản lý và thu thập bằng chứng kiểm toán để chứng thực hoặc phản bác các phản hồi nhận được từ ban quản lý.

❖ Đánh giá kết quả của thủ tục phân tích cơ bản

Mô hình của kiểm toán viên đã tính đến việc một số lượng lớn các hợp đồng thuê sẽ hết hạn vào cuối tháng sáu. Thông thường, các căn hộ sau khi hết hợp đồng thuê sẽ bị trống ít nhất hai tuần. Tuy nhiên, trong trường hợp này, những người thuê mới được tìm thấy nhanh hơn. Ngoài ra, kiểm toán viên còn tìm hiểu được rằng các nỗ lực làm mới khu phố đã khiến cho thành phố sạch đẹp và sống động hơn, đặc biệt là các khu vực xung quanh các khu chung cư của doanh nghiệp, điều này làm cho doanh thu cao hơn dự kiến. Tất cả các hợp đồng thuê mới đã được kiểm tra để xác minh rằng trên thực tế, những người thuê mới đã được tìm thấy nhanh hơn và doanh nghiệp có thể tính giá thuê cao hơn đáng kể cho các căn hộ trong các khu này so với mô hình của kiểm toán viên sử dụng để dự đoán. Dựa trên việc thu thập bằng chứng thích hợp, kiểm toán viên kết luận rằng không có sự sai lệch trọng yếu nào về doanh thu cho thuê.

4. VẬN DỤNG PHÂN TÍCH DỮ LIỆU LỚN VÀO THỦ NGHIỆM CHI TIẾT

4.1. Định nghĩa thử nghiệm chi tiết

Thử nghiệm chi tiết (các nhóm giao dịch, số dư tài khoản và thuyết minh) là một loại thử nghiệm cơ bản. Nếu kiểm toán viên đã xác định rủi ro có sai sót trọng yếu đã được đánh giá ở cấp độ cơ sở dẫn liệu là rủi ro đáng kể thì kiểm toán viên phải thực hiện các thử nghiệm cơ bản để xử lý rủi ro này. Nếu chỉ thực hiện thử nghiệm cơ bản đối với một rủi ro đáng kể thì thử nghiệm cơ bản phải bao gồm kiểm tra chi tiết.

4.2. Trình tự khi thực hiện thử nghiệm chi tiết bằng phân tích dữ liệu lớn

Dưới đây năm bước cơ bản và các quy trình liên quan để sử dụng khi lập kế hoạch, thực hiện và đánh giá kết quả của một phép phân tích được sử dụng để thực hiện kiểm tra chi tiết.

Bước 1: Lập kế hoạch cho phép phân tích.

- Xác định các khoản mục, tài khoản, thuyết minh, và các cơ sở dẫn liệu liên quan,

đồng thời xác định bản chất, thời gian và mức độ của tổng thể mà kiểm toán viên sẽ phân tích.

- Xác định mục đích chung của việc phân tích dữ liệu

- Xác định các mục tiêu cụ thể của việc phân tích dữ liệu

- Xác định tổng thể dữ liệu sẽ được phân tích hoặc kiểm tra, bao gồm cho mục đích lập kế hoạch, xem xét sơ bộ các vấn đề có thể ảnh hưởng đến sự liên quan, tính sẵn có và độ tin cậy của dữ liệu.

- Chọn phép phân tích dữ liệu phù hợp mục đích cũng như các kỹ thuật, công cụ, loại bảng biểu sẽ được sử dụng.

Bước 2: Truy cập và chuẩn bị dữ liệu cho phép phân tích.

Bước 3: Xem xét sự liên quan và độ tin cậy của dữ liệu được sử dụng.

Bước 4: Thực hiện phép phân tích.

- Nếu kết quả ban đầu của phép phân tích cho thấy việc thiết kế hoặc hiệu quả của nó cần phải được sửa đổi thì phải điều chỉnh lại và thực hiện lại phân tích.

- Nếu kiểm toán viên kết luận rằng phép phân tích đã được thiết kế và thực hiện phù hợp và phép phân tích đã xác định được các khoản mục cần được xem xét, thì việc lập kế hoạch và thực hiện các thủ tục bổ sung cho các khoản mục đó cần phù hợp với mục tiêu của phép phân tích.

Bước 5: Đánh giá kết quả và kết luận về mục đích và mục tiêu cụ thể của việc thực hiện phép phân tích đã đạt được hay chưa.

Phép phân tích dữ liệu được sử dụng để thực hiện kiểm tra chi tiết có thể giúp kiểm toán viên xác định được sai sót. Sai sót này có thể cho thấy sự thiếu hụt của một kiểm soát có liên quan hoặc sự sai sót của một kiểm soát mà kiểm toán viên ý định tin tưởng vào. Một sai sót như vậy thường sẽ được coi là một sự thiếu sót trong kiểm soát nội bộ, mức độ nghiêm trọng của sai sót này sẽ yêu cầu sự đánh giá của kiểm toán viên. Khi đánh giá hiệu quả hoạt động của các

kiểm soát có liên quan, kiểm toán viên cần đánh giá xem các sai sót đã được phát hiện bởi các thủ tục phân tích cơ bản cho thấy các kiểm soát có hoạt động hiệu quả hay không. Tuy nhiên, việc không có sai sót được phát hiện bởi các thủ tục cơ bản không cung cấp bằng chứng kiểm toán cho thấy các kiểm soát liên quan hoạt động có hiệu quả.

Một phép phân tích được thiết kế và thực hiện phù hợp có thể xác định được một số ít sai sót có thể xảy ra, tuy nhiên, trong nhiều trường hợp, số lượng sai sót có thể xảy ra sau khi kiểm tra chi tiết lại rất lớn. Trong những trường hợp như vậy kiểm toán viên cần đánh giá xem liệu phép phân tích đã được lên kế hoạch và thực hiện một cách thích hợp hay chưa, nếu chưa, phải thiết kế và thực hiện lại. Kiểm toán viên cũng có thể quyết định áp dụng cách phân nhóm và sàng lọc để giải quyết số lượng lớn các sai sót có thể được xác định.

Xem xét một ví dụ đơn giản khi kiểm toán viên thiết kế và thực hiện một phép phân tích để kiểm tra chi tiết đối chiếu tiền mặt nhận được với hóa đơn. Đây là một trong số những thủ tục được sử dụng để cung cấp bằng chứng liên quan đến cơ sở dẫn liệu chính xác của doanh thu. Áp dụng kỹ thuật phân tích k-means clustering để phân cụm, kết quả phân tích xác định một số lượng lớn các sai sót có thể xảy ra, vì đã phát hiện một lượng lớn nghiệp vụ có số tiền và hóa đơn không khớp nhau. Đánh giá của kiểm toán viên về kết quả phân tích này đã chỉ ra rằng nhiều sai sót thuộc trong một cụm liên quan đến tiền mặt được ghi có vào tài khoản tạm thời (suspense account). Kiểm toán viên đã hỏi ban quản lý về cách họ dự định thực hiện để ghi đúng lại tài khoản này. Kiểm toán viên cũng sử dụng kỹ thuật phân tích clustering để phân cụm các mục trong tài khoản theo nguồn đăng và các thuộc tính thích hợp khác. Thông tin này sau đó đã được sử dụng để đánh giá việc xử lý các mục trong tài khoản tạm thời. Kết quả của phép phân tích cũng xác định ra các cụm sai sót tiềm năng khác mà kiểm toán viên đánh giá là không quan trọng. Kiểm toán viên xem xét các khía cạnh định tính về bản

chất của các hạng mục trong mỗi cụm và kết luận rằng không cần thực hiện thêm thủ tục nào liên quan đến các khoản mục trong các cụm này.

4.3. Minh họa trường hợp công ty thuộc lĩnh vực sản xuất

Đây là một minh họa về việc sử dụng phân tích dữ liệu lớn khi thực hiện quy trình kiểm tra ba bước: hóa đơn - phiếu nhập kho - giá niêm yết. Báo cáo tài chính đang được kiểm toán trong ví dụ này là của một nhà sản xuất đồ nội thất. Có khoảng 60.000 giao dịch bán hàng mỗi năm. Doanh nghiệp ghi nhận doanh thu khi chuyển giao hàng hóa cho khách hàng trên cơ sở FOB (khi quyền kiểm soát được chuyển cho khách hàng). Phép phân tích này là một thử nghiệm chi tiết để cung cấp bằng chứng về cơ sở dẫn liệu phát sinh và chính xác của các giao dịch bán hàng.

Công ty có hai loại sản phẩm chính: đồ nội thất gia đình được bán cho các nhà bán lẻ và đồ nội thất thương mại được bán cho các công ty để sử dụng riêng. Có 7 nhóm sản phẩm (ví dụ: ghế sofa, ghế gỗ, bàn gỗ, bàn sofa,...) và khoảng 80 loại sản phẩm cụ thể trong các nhóm sản phẩm. Các loại sản phẩm này mô tả đặc trưng cụ thể của một nhóm sản phẩm (ví dụ: ghế sofa, ghế gỗ). Có hàng ngàn biến thể sản phẩm dựa trên sự kết hợp giữa vải, màu sắc và vật liệu. Mỗi kiểu cụ thể đều có mã số hàng hóa riêng (SKU – Stock Keeping Unit) và mỗi mã số có một mức giá riêng.

Chỉ có sáu khách hàng cao cấp có khối lượng đặt hàng mua nhiều mới được hưởng giảm giá. Mỗi đại lý bán hàng được ủy quyền giảm giá tùy ý cho các khách hàng cao cấp được chấp thuận trước lên tới 15% so với giá quy định trong bảng giá. Nếu giảm trên 15% thì yêu cầu thêm sự phê duyệt của phó tổng giám đốc.

❖ Lập kế hoạch cho phép phân tích

• Xác định khoản mục và cơ sở dẫn liệu liên quan

Các cơ sở dẫn liệu cần quan tâm là phát sinh và chính xác của doanh thu. Các loại sai sót có thể xảy ra bao gồm sai lệch doanh thu do sự

khác biệt giữa các dữ liệu được sử dụng trong việc tạo hóa đơn bán hàng. Những dữ liệu này bao gồm loại sản phẩm, số lượng, giá cả và mức giảm giá, dữ liệu liên quan đến loại sản phẩm khách hàng đặt hàng và hạn mức giảm giá.

Phép phân tích này là một trong một số thủ tục được sử dụng để thu thập bằng chứng liên quan đến cơ sở dẫn liệu phát sinh và chính xác của doanh thu. Điều này bao gồm bằng chứng thu được từ các thử nghiệm kiểm soát có liên quan và bằng chứng từ các thử nghiệm cơ bản khác, chẳng hạn như xác nhận số dư khoản phải thu với khách hàng (bao gồm chi tiết hóa đơn).

• Xác định tổng thể dữ liệu cần kiểm tra

Kiểm toán viên đã quyết định rằng thủ tục kiểm tra chi tiết này sẽ tập trung vào việc so sánh sản phẩm và giá bán sản phẩm mà khách hàng đã đặt với hóa đơn liên quan và chứng từ vận chuyển do doanh nghiệp phát hành. Đơn đặt hàng của khách hàng là bằng chứng từ nguồn bên ngoài. Tuy nhiên, dữ liệu trong đơn đặt hàng của khách hàng đã được chuyển đến các tài liệu đặt hàng nội bộ để định dạng phù hợp với cơ sở dữ liệu của công ty.

Kiểm toán viên thu thập dữ liệu liên quan từ cơ sở dữ liệu của công ty. Dữ liệu được sử dụng bao gồm:

- Thông tin tài khoản khách hàng
- Thông tin đơn đặt hàng
- Thông tin sản phẩm được đặt
- Số lượng đặt hàng
- Giá bán của sản phẩm
- Thông tin đơn vận chuyển
- Thông tin sản phẩm được vận chuyển
- Số lượng sản phẩm vận chuyển
- Đơn vị vận chuyển
- Đơn giá vận chuyển
- Thông tin hóa đơn
- Thông tin sản phẩm trên hóa đơn
- Số tiền trên hóa đơn
- Tỷ lệ chiết khấu
- Ngày ghi sổ

Kiểm toán viên cũng sử dụng thêm các cơ sở dữ liệu khác của doanh nghiệp như các tập tin liên quan đến quá trình hợp đồng (liên kết các đơn đặt hàng của khách hàng với các đơn đặt hàng được chuẩn bị nội bộ) và bảng giá của từng loại sản phẩm.

• Xác định phép phân tích phù hợp

Kiểm toán viên đã quyết định thực hiện kỹ thuật xử lý dữ liệu lớn cây quyết định (decision tree) để kiểm tra các chi tiết xem thông tin trên các tài liệu hỗ trợ có khớp với nhau hay không, và nếu không khớp thì tổng sai lệch có vượt quá mức trọng yếu hay không. Điều này đòi hỏi thực hiện hai quy trình kiểm tra ba bước. Đối với mỗi giao dịch bán hàng trong năm được kiểm toán, thực hiện hai so sánh sau:

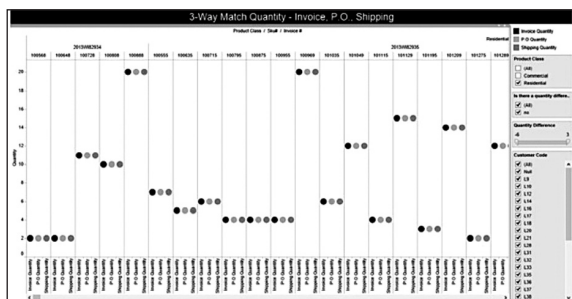
- Số lượng được bán theo hóa đơn bán hàng, chứng từ giao hàng và đơn đặt hàng nội bộ để xác định xem tất cả chúng có khớp không.

- Giá trên hóa đơn bán hàng, đơn đặt hàng hoặc tài liệu hỗ trợ tương tự và bảng giá chính của công ty để xác định xem tất cả chúng có khớp không.

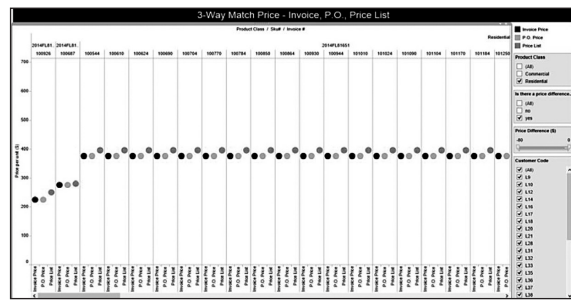
❖ Chuẩn bị dữ liệu

Kiểm toán viên sử dụng một phép phân tích cây quyết định (decision tree) khác để kiểm tra tính liên tục của các đơn đặt hàng, hóa đơn và chứng từ vận chuyển và để xử lý các số còn thiếu. Phần mềm kiểm toán cũng được sử dụng để xác định và xử lý các tập tin không có dữ liệu hoặc dữ liệu được định dạng không phù hợp trước khi thực hiện phân tích dữ liệu trong kiểm tra chi tiết.

❖ Thực hiện phân tích và đánh giá kết quả



Biểu đồ 3. Kiểm tra ba bước về số lượng



Biểu đồ 4. Kiểm tra ba bước về giá bán

Kiểm toán viên thực hiện hai quy trình kiểm tra ba bước. Đối với phép so sánh về số lượng thì không phát hiện được sự không phù hợp nào. Điều này được thể hiện trong biểu đồ 3. Số lượng sản phẩm được hiển thị trên trực tuyến, các dấu chấm màu khác nhau biểu hiện cho số lượng trên hóa đơn, số lượng trên đơn đặt hàng và số lượng được vận chuyển. Kết quả của phép so sánh cho thấy số lượng ở ba bên đều giống nhau.

Trong biểu đồ 4, trực tuyến thể hiện giá bán sản phẩm. Các dấu chấm tương tự biểu đồ 3 cũng được sử dụng để thể hiện giá bán sản phẩm trên hóa đơn, trên đơn đặt hàng và trên bảng giá. Ta thấy rằng các dấu chấm màu xanh dương (giá niêm yết trên bảng giá) nằm cao hơn các dấu chấm khác. Kiểm toán viên xử lý sự bất thường này bằng cách sử dụng một phép phân tích k-means clustering bổ sung để xác định xem sự không phù hợp này có phân tách theo các điểm đặc trưng nào hay không và có thể hiện một mô hình nào không, để giúp xác định các nguyên nhân có thể có. Ví dụ, kiểm toán viên dự đoán liệu sự chênh lệch này xảy ra vào một ngày cụ thể, trong một chuỗi ngày cụ thể hay liệu sự không phù hợp chỉ liên quan đến một số khách hàng hay không. Kiểm toán viên thu được bằng chứng từ phép phân tích bổ sung rằng sự không phù hợp này là do quá trình đối chiếu không được tính đến phần giảm giá. Sau đó, kiểm toán viên đã thiết kế lại phép phân tích để kiểm tra xem liệu có bất kỳ khách hàng nào, ngoài khách hàng cao cấp, có được giảm giá hay không và có bất kỳ khách hàng cao cấp nào có được giảm giá vượt quá 15% hay không. Sau khi thực hiện lại

phép phân tích thì không tìm ra được trường hợp nào như vậy.

Kiểm toán viên kết luận rằng quy trình lập hóa đơn của khách hàng về số lượng và giá cả là phù hợp với đơn đặt hàng của khách hàng, cung cấp bằng chứng về cơ sở dữ liệu phát sinh và chính xác của các giao dịch bán hàng trong năm.

5. KẾT LUẬN

Dữ liệu lớn vẫn là một chủ đề mới đối với các nhà nghiên cứu và nhà thực hành. Tuy nhiên, nó đã đem đến nhiều sự thay đổi ở nhiều thị trường. Phân tích dữ liệu lớn là một hành trình dài nhiều năm, nhiều giai đoạn. Điều quan trọng là phải có một tầm nhìn chiến lược, định hướng phù hợp để đáp ứng tốt với các yếu tố tiêu cực. Tang và Kostic (2017) cho rằng nghiên cứu trong một số cơ sở dữ liệu điện tử cho thấy tình trạng hiện tại của các nghiên cứu liên quan đến dữ liệu lớn và việc triển khai các quy trình kiểm toán theo dữ liệu lớn vẫn còn ở giai đoạn sơ khai.

Các học giả vẫn đang nỗ lực để đạt được sự thống nhất trong việc xây dựng một khung lý thuyết cho phép thực hiện và ứng dụng thành công dữ liệu lớn. Trong khi đó, các báo cáo của các công ty kiểm toán thuộc Big 4 lại tập trung nhiều hơn vào tiềm năng gây gián đoạn của các công nghệ này đối với ngành kiểm toán và các ngành công nghiệp khác. Một điểm chung giữa những người hành nghề và các nhà nghiên cứu học thuật là họ đều nhận thức được về sự hiện diện của dữ liệu lớn và họ phản ứng với những phát triển mới bằng cách đầu tư nhiều nguồn lực

hơn vào việc nghiên cứu chúng. Một dấu hiệu rõ ràng của xu hướng này là sự tăng vọt gần đây về số lượng ấn phẩm trong lĩnh vực này trong một khoảng thời gian rất ngắn, cụ thể là hai năm qua cho đến nay. Tuy nhiên, cả những người hành nghề và học giả vẫn còn nằm ở một khoảng cách khá xa so với đích đến là đưa ra những hướng dẫn cụ thể, rõ ràng trong việc áp dụng dữ liệu lớn vào ngành nghề của mình. Thực tế, các khách hàng lớn và có nhiều ứng dụng công nghệ tiên tiến như các quỹ đầu tư, các ngân hàng là những người đầu tiên sử dụng các công cụ phân tích dữ liệu lớn tiên tiến, và do đó, đây sẽ là động lực cho các công ty kiểm toán áp dụng các công nghệ này vào hoạt động kinh doanh của mình.

TÀI LIỆU THAM KHẢO

1. Mohanty, S., Jagadeesh, M., & Srivatsa, H. Big data imperatives: Enterprise 'Big Data' warehouse, 'BI' implementations and analytics, Apress, 2013.
2. American Institute of Certified Public Accountants (AICPA). Guide to audit data analytics, American Institute of Certified Public Accountants, 2017.
3. Global Technology Audit Guide (GTAG). Understanding and Auditing Big Data, 2017.
4. Tang, X., & Kostic, N. The future of audit: Examining the opportunities and challenges stemming from the use of Big Data Analytics and Blockchain technology in audit practice, Working paper, 2017.