Data Science and External Audit

Ahmad Faisal Hayek

1 Aspects of Using Big Data Analytics in External Audit

1. Assess fraud risk.
2. Audit evidence.

1.1 *Assess Fraud Risk*

BDA can help assess fraud risks and perform specific audit procedures to address those risks. For example, in identifying specific journal entries with unique risk profile. This identification results in the application of further audit procedures and analysis to support auditors' assessment of fraud risks.

Some benefits of using data analysis in fraud detection are (Rajvanshi, 2016):

- Identify low incidence of events: Analytics can be used to find fraud cases that are not very obvious and are low incidence, and then use predictive analytics on these fraud cases to analyze them further.
- Enterprise-wide solution: Analytics takes an enterprise-wide global perspective, which helps in detecting fraud by associating related information in a company.
- Data integration: A very useful way of detecting fraud in a company is getting as much data as possible. Analytics makes it possible by gathering data from different sources along with some outside data that may have some prediction capability.

A. F. Hayek (✉)
Department of Accounting, Faculty of Business, Higher Collogues of Technologies, Abu Dhabi, United Arab Emirates
e-mail: ahayek@hct.ac.ae

• Utilizing unstructured data: Sometimes the unstructured data is not properly stored and it contains the most valuable information related to fraud. Analytics help store this data and bring it to some value.

Artificial Intelligence and machine learning techniques for fraud detection can be divided into Supervised and Unsupervised learning. These methods predict what sections of customers or users are more likely to commit fraud by giving out a score or rule (Balios et al., 2020).

Supervised Learning It is a type of machine learning algorithm which is used to detect fraud cases that follow a pattern already been identified by the system. In this method, a sample from the entire data is taken and the fraudulent cases are separated from the rest. By training the model to identify the fraudulent cases, these models are then used to identify new fraud cases by applying them to new data.

Unsupervised Learning It is used to detect any new type of fraud that may occur which was never seen before. This, unlike supervised learning does not use labeled records.

1.2   Audit Evidence

As per International Standards on Auditing, standard 500 "Audit evidence," the auditor should obtain sufficient appropriate audit evidence to be able to draw reasonable conclusions on which to base the audit opinion.

Appropriate means reliable and relevant, while sufficient is related to the volume and variety, and Big Data can contribute to this aspect. When traditional data are not adequately "reliable" or "relevant," then more evidence from Big Data can be useful. Furthermore, Big Data can be more reliable and relevant than traditional sources, even though "noise" may impact on their reliability.

As far as reliability is concerned, some types of Big Data can contribute to evaluating the reliability of traditional audit information. Besides, Big Data from external sources can provide crucial nonfinancial evidence which can be used to assess financial accounts. For example, when a product receives adverse comments on Social Media, while the Sales registered in the financial statements have increased, then this could be a sign for further investigation.

Unlike traditional audit practices, the technology-enabled audit comes with a higher quality of audit evidence, which is derived from many new sources, including big data, exogenous data, the ability to analytically link different processes, database-to-database confirmation, and continuous monitoring alerts.

The amount of resources used could be reduced if Big Data Analytics has the potential to replace existing labor-intensive parts of the financial statement audit. For example, testing the value of the inventory from retailer is generally done by obtaining evidence examination which is labor intensive. Brown Liburd and

Vasarhelyi stated that using the information of radio frequency identification (RDFI) chips for validating inventory could make the process more labor efficient.

Big Data supported analytical tools do not just examine the entire population of available data, but also incorporate other unstructured data to establish a relationship between every data examined and draw useful insights from them.

For example, data analytic tools can analyze entire financial data at different dimensions like date, time, purpose, transaction types, transaction value, business type, customer type, geography, standards-based, and so on.

The analytic procedure is a significant part of the audit process, which includes analyzing data to find out any plausible relationship between financial and nonfinancial data. At the same time, BDA is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their operations, and make better, fact-based decisions (Jesus, 2018).

Analytical procedures, according to Auditing Standards 2305 (PCAOB, AS 2305 2016), are an important part of the audit process and mainly consist of an analysis of financial information made by a study of believable or plausible relationships among both financial and nonfinancial data. These analytical procedures could be as basic as scanning (viewing the data for abnormal events or items for further examination) to more complex approaches (not clarified by the standards, except that the approach should enable the auditor to appropriately develop an expectation and subsequently examine these expectations to the reported results).

Big Data Analytics (BDA) that is utilized by client management and their accountants has been defined as "the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their operations, and make better, fact-based decisions."

The Auditing standards define the task for analytical procedures in each of the three audit phases (risk assessment/planning phase, substantive testing, and review phase), but are noncommittal about which techniques auditors should undertake to achieve these objectives. Hence, whether an auditor employs more complex analytics such as belief Functions or "traditional analytical procedure" techniques such as ratio analysis would seem to depend on the auditor's own knowledge and less so on the standards. It has also been proposed that any adoption by the external audit profession of either advanced analytics or big data would be due to market or business forces exogenous to the firms (Raphael, 2017).

2 Audit Data Analytics

Audit Data Analytics (ADA): is the science and art of using analysis, modeling and visualization to discover and analyze patterns, anomalies, and other information for the purpose of planning and performing an audit.

Audit Data Analytics (ADAs) help auditors discover and analyze patterns, identify anomalies and extract other useful information from audit data through analysis, modeling, and visualization. Auditors can use ADAs to perform a variety of procedures to gather audit evidence, to help with the extraction of data and facilitate the use of audit data analytics, and as a tool to help illustrate where audit data analytics can be used in a typical audit program.

3   Application of ADA

- Identifying and analyzing anomalies in the data.
- Identifying and analyzing patterns in the data including outliers.
- Building statistical (e.g., regression) or other models that explain the data in relation to other factors and identify significant fluctuations from the model.
- Synthesizing pieces of information from disparate analyses and data sources into wholes that are greater than the sum of their parts for purposes of the overall evaluation.

ADA mode can be exploratory or confirmatory:

Exploratory mode	
When	Planning
Question	What is going on here? Does the data suggest something might have gone wrong? Where do the risks appear to be? What assertions should we focus on?
Approach style	Bottom-up, inductive, few starting assumptions, assertion-free
Methods	Graphical visualizations used to discover patterns in and understand the data—Possibly several to get different viewpoints
Results	Identified risks, areas of focus, potential models for confirmatory stage
ADA examples	– Cluster analysis – Test and data mining – Scatterplots matrices – Line chart – Spread chart – Needle graphs – Small multiples of graphics – Heart maps – Treemaps – Relationship maps

Confirmatory mode	
When	Performance
Question	Does the data conform with and thus confirm my model for what ought to be?
Approach style	Top-down, deductive, model-driven, starts with the development of model based on assertions to be tested
Methods	Comparison of actual data to model taking into account materiality, desired assurance, and assertions being tested; more mathematical than graphical
Results	Identified anomalies, unexpected patterns, outliers, and other significant deviations
ADA examples	– Analytical procedures (regression analysis, ration analysis, reasonableness test) – Recalculations – Journal entry testing – Traditional file interrogation

4   Audit Data Analytics Techniques

- Ratio Analysis: use to calculate ratios using financial and nonfinancial data to gain a high-level understanding of entity operations.
- Sorting: use the software to sort financial and nonfinancial data by categories to identify outliers.
- Trend Analysis: use software to evaluate changes and trends in financial and nonfinancial data over time.
- Matching: use software to electronically match items from various sources on a predetermine characteristics to identify errors and unexpected differences.
- Comparison: use year-over-year comparisons in accounts balances and other nonfinancial data.
- Forecasting: use software to extrapolate past patterns onto future period.
- Predictive Analysis: use software to apply predicated patterns to existing data. Enables auditors to identify situations that have deviated from expectations.
- Cluster: use software to group data into natural clusters. Enables auditors to identify outliers or observations with unique risk characteristics.
- Regression Analysis: use software to understand existing relationships between various characteristics. Enables external auditors to make expectations as to current period balances.
- Process Mining: use software to identify deviations from expected process flow and patterns.

5 Audit Data Analytics Framework

Framework is defined by three Business Analytics dimensions; Domain, Orientation, and Techniques:

(a) Domain: The domain is the environment where audit teams apply analytics like client' enterprise and management.

- Pre-engagement Activities
- Planning
- Compliance Testing
- Substantive Testing and Review
- Opinion Formulation and Reporting
- Continuous Activities.

(b) Orientation: The orientation refers to the vision of analytics descriptive, predictive, or prescriptive.

(c) Technique: The technique is the analytical approach or method (Appelbaum et al., 2015).

- Qualitative or Quantitative
- Deterministic or statistical
- Based on unstructured, semi-structured, or structured data

In addition to the above frameworks, ADA mode can be Expectation, Structural and Multivariate techniques:

- Expectation techniques: An empirical relationship among several accounting numbers or some other important quantitative measures of business operations and is inferred from the archive of historical records.
- Structural techniques look for various structural properties in the historical records. A very popular recent example is process mining.
- Multivariate techniques: The primary objective of multivariate techniques is to develop relationships between or among variables/features under study.

6 Performing Audit Data Analytics (ADAs)

6.1 Plan the ADA

The auditors consider the following when planning the ADA:

- Determine the financial statement items, transactions, accounts, or disclosures, and related assertions and the nature, timing, and extent of the population to which the ADA will be applied.

- Determine whether ADA is to be used in performing a risk assessment procedure, a test of controls, a substantive analytical procedure, a test of details, or in procedures to help form an overall conclusion from the audit.
- Select the techniques, tools, graphics, and tables to be used. Ration Analysis, sorting, trend analysis, matching, comparison, forecasting, predict analysis, cluster analysis, Regression analysis, and process mining.

6.2   Access and Prepare the Data for Purposes of the ADA

This step concentrates on obtaining the data from the entity's ERP or another data source and preparing the data for analysis.

6.3   Consider the Relevance and Reliability of the Data

Data's relevance (the extent to which it relates to the purpose of the ADA) and reliability (the extent to which the data is accurate, complete, and precise) is affected by the data's nature, source, format, timing, extent, and level of aggregation.

6.4   Perform the ADA

How the ADA will be performed depends both on the technique used (for example, regression analysis or trend analysis) and on the purpose of the ADA (for example, preliminary analytical procedure vs. substantive test).

Evaluate the results and conclude on whether the purpose and specific objectives of performing the ADA have been achieved.

7   BDA's Tools and Techniques

The more forward looking the task and the more varied and voluminous the data (big data), the more likely the analysis will be prescriptive or at the very least, predictive. Advanced or more complex BDA may be defined as "Any solution that supports the identification of meaningful patterns and correlations among variables in complex, structured and unstructured, historical, and potential future data sets for the purposes of predicting future events and assessing the attractiveness of various courses of action.

Advanced analytics typically incorporate such functionality as data mining, descriptive modeling, econometrics, forecasting, operations research, optimization, predictive modeling, simulation, statistics, and text analysis".

1. Artificial intelligence
2. Data analytics
3. Machine learning application
4. Data visualization
5. Data mining

7.1 Artificial Intelligence

The original goal with the creation of AI was to make computers more capable of independent thinking. AI uses machines that can interpret and learn from external data. Artificial intelligence is a "computing system that exhibits some form of human intelligence, which covers several interlinked technologies, including data mining, machine learning, speech recognition, image recognition, and sentiment analysis."

Integrating AI in each step of auditing process will remove the repetitive tasks common in the process and make analyzing large volumes of data to have an in-depth understanding of the business operation easier for auditors. Making it easier to concentrate on activities that will bring utmost value to the clients. As assessing the risk of material misstatement is a crucial part of auditing. Auditors are expected to carry out tests on the transactions to make certain that there are no misstatements, for if financial impacts are not accurately recorded, financial statements are bound to be materially misstated. If unauthorized transactions and/or other irregularities are not detected in time, it may be challenging for auditors to capture such later. AI-based tools in auditing make detecting such high-risk transactions easy. This which manual auditing may sometimes not capture fully as a result of sample population testing, unlike the AI technology that allows for full population testing.

Patterns of Artificial Intelligence
Hyper Personalization This pattern uses machine learning to develop a profile unique to each individual. This profile will adapt over time and is used to provide personalized content unique to each user instead of grouping users into categories.
Autonomous System An autonomous system can complete a task, goal, and interact with its surroundings with minimal or no human involvement. This autonomous system can have both hardware and software components, but the overall goal is to minimize human labor. Autonomous systems are used in cars, airplanes, boats, and more which provide information with minimal human involvement.
Predictive Analytics and Decision Support This process involves using cognitive approaches and machine learning to determine patterns that can help predict future outcomes. Predictive analytics is used in projection methods such as forecasting to help humans make better decisions.
Conversation and Human Interaction The objective of conversation and human interaction is to enable machines to interact with humans the way that humans interact with each other. The ability of machines to communicate with humans includes voice assistants, chatbots, and the generation of text, images, and audio.

Anomaly and Pattern Detection Machine learning is used to find patterns within the data. Anomaly and pattern detection are used to determine connections between the information and can determine if the data fits into a pattern or if it is an outlier. This pattern is primarily used to decide which data is similar to other information and which data is different.

Recognition The recognition pattern uses machine learning to specifically identify desired information within unstructured data. Unstructured data is data not easily identifiable such as audio and video. The primary objective of the recognition pattern uses machine learning to identify and understand desired things from unstructured content.

Goal-Driven Systems This uses g machine learning to give people the ability to determine the best solution to a problem An example of a goal-driven system is in a business that needs to find the optimal way to achieve a goal. Using this pattern will allow the business to have the best solutions to possible problems.

Types of Artificial Intelligence
1. Assisted AI: a means of automating simple processes and tasks by harnessing the combined power of Big Data, cloud, and data science to aid in decision-making. Assisted AI is to support humans in making decisions. Assisted AI has the benefit of being used to complete basic tasks, thus freeing up the user to perform more complex tasks.
2. Augmented AI: allows organizations and people to do things they could not otherwise do by supporting human decisions, not by simulating independent intelligence. Augmented AI is more advanced than Assisted AI because Augmented AI can make some decisions on its own but is not completely independent of the user. Overall, Augmented AI suggests new solutions rather than simply identifying patterns and applying predetermined solutions.
3. Autonomous AI: The most advanced form of AI is Autonomous AI, "in which processes are automated to generate the intelligence that allows machines, bots and systems to act on their own, independent of human intervention". Autonomous AI is the most sophisticated type and has the capability to operate without any user interference. Autonomous AI is able to "adapt to their environments and perform tasks that would have been previously unsafe or impossible for a human to do (e.g., the use of drones to perform inventory inspections autonomously of assets in remote locations auditors do not have access to)."

7.2 Data Analytics (Descriptive, Diagnostic, Predictive, and Prescriptive)

Predictive analytics is a subset of data analytics. Predictive analytics can be viewed as helping the accountant or auditor in understanding the future and provides foresight by identifying patterns in historical data. One of the most common

applications of predictive analytics in the field of accounting is the computation of a credit score to indicate the likelihood of timely future credit payments. This predictive analytics tool can be used to predict an accounts receivable balance at a certain date and to estimate a collection period for each customer.

7.3   Machine Learning Application

Machine learning is a subset of artificial intelligence that automates analytical model building. Machine learning uses these models to perform data analysis in order to understand patterns and make predictions. The machines are programmed to use an iterative approach to learn from the analyzed data, making the learning automated and continuous; as the machine is exposed to increasing amounts of data, robust patterns are recognized, and the feedback is used to alter actions.

Machine learning is a key subset of artificial intelligence (AI), which originated with the idea that machines could be taught to learn in ways similar to how humans learn. Common examples of machine learning can be found in e-mail spam filters and credit monitoring software, as well as the news feed and targeted advertising functions of technology companies such as Facebook and Google.

In machine learning applications, the expectation is that the algorithm will learn from the data provided, in a manner that is similar to how a human being learns from data. A classic application of machine learning tools is pattern recognition.

Facial recognition machine learning software has been developed such that a machine learning algorithm can look at pictures of men and women and be able to identify those features that are male driven from those that are female driven (Alles et al., 2002).

The machines are programmed to use an iterative approach to learn from the analyzed data, making the learning an automated and continuous process.

The goal of machine learning is to write an algorithm that can be trained using test data to look for specific patterns. For example, if a machine learning algorithm that could look at pictures of animals and identify those that contain cats is desired, one starts by identifying the general characteristics of a cat (four legged, furry animal with a tail) and providing it with a sample set of pictures of animals. Initially, the algorithm would be able to identify animals that do not contain the common characteristics such as snakes (no fur, no legs), birds (no fur, not four legged), and fish (no fur, no legs). But it would need to learn that there are other characteristics (distinctive sounds, claws, body shape) to differentiate it from other four-legged furry animals with tails (Verma & Mani, 2015).

Machine learning and traditional statistical analysis are similar in many aspects. The statistical analysis is based on probability theory and probability distributions, while the machine learning is designed to find the optimal combination of mathematical equations that best predicts an outcome. Thus, machine learning is well suited for a broad range of problems that involve classification, linear regression, and cluster analysis.

In 2011, Johan Perols from the University of San Diego compared six of the most popular statistical and machine learning models being used to detect fraud and determined an overlap in six of the 42 predictors that were consistently chosen by the programs. These predictors included auditor turnover, total discretionary accruals, unexpected jumps in employee productivity, and others that some auditors may have not noticed.

Even if they did notice these irregularities in certain factors within the firm, being able to put these facts into a machine learning software and compare the firm's behavior to other companies helps determine the possible presence of fraud faster and more efficiently than if the auditor were to cross-reference all of these facts themselves.

The predictive reliability of machine learning applications is dependent on the quality of the historical data that has been fed to the machine. New and unforeseen events may create invalid results if they are left unidentified or inappropriately weighted. As a result, human biases can influence the use of machine learning.

Such biases can affect which data sets are chosen for training the AI application, the methods chosen for the process, and the interpretation of the output. Finally, although machine learning technology has great potential, its models are still currently limited by many factors, including data storage and retrieval, processing power, algorithmic modeling assumptions, and human errors and judgment.

Jon Raphael, chief innovation officer at Deloitte, expects machine learning to significantly change the way audits are performed, as it enables auditors to largely "avoid the tradeoff between speed and quality."

Rather than relying primarily on representative sampling techniques, machine learning algorithms can provide firms with opportunities to review an entire population for anomalies. When audit teams can work on the entire data population, they can perform their tests in a more directed and intentional manner. In addition, machine learning algorithms can "learn" from auditors' conclusions on specific items and apply the same logic to other items with similar characteristics.

One of the most tedious and time-consuming parts of an audit is the time it takes to review and extract key terms from contracts. With artificial intelligence, this process has become automated, with systems being taught how to review the same documents and then identify and extract key terms. To solve this problem and speed up the document review process, Deloitte US developed an automated document review platform that natural language processing (NLP) to read electronic documents and machine learning to identify relevant information and flag key terms within the documents.

Machine learning technology for auditing is a very promising area (Dickey, 2019). Several of the Big 4 audit firms have machine learning systems under development, and smaller audit firms are beginning to benefit from improving viability of this technology. It is expected that auditing standards will adapt to take into account the use of machine learning in the audit process. Regulators and standard setters will also need to consider how they can incorporate the impact of this technology in their regulatory and decision-making process. Likewise, educational programs will continue to evolve in this new paradigm. We foresee that more

accounting programs with data analytics and machine learning specializations will become the norm rather than the exception.

It is possible that many routine accounting processes will be handled by machine learning algorithms or robotics automation processing (RPA) tools in the near future. For example, it is possible that machine learning algorithms can receive an invoice, match it to a purchase order, determine the expense account to charge and the amount to be paid, and place it in a pool of payments for a human employee to review the documents and release them for payment to the respective vendors.

In the same way, in auditing a client, a well-designed machine learning algorithm could make it easier to detect potential fraudulent transactions in a company's financial statements by training the machine learning algorithm to successfully identify transactions that have characteristics associated with fraudulent activities from bona fide transactions. The evolution of machine learning is thus expected to have a dramatic impact on business, and it is expected that the accounting profession will need to adapt so as to better understand how to utilize such technologies in modifying their ways of working when auditing financial statements of their audit clients.

One example is Deloitte's use of Argus, a machine learning tool that "learns" from every human interaction and leverages advanced machine learning techniques and natural language processing to automatically identify and extract key accounting information from any type of electronic documents such as leases, derivatives contracts, and sales contracts. Argus is programmed with algorithms that allow it to identify key contract terms, as well as trends and outliers. It is highly possible for a well-designed machine to not just read a lease contract, identify key terms, determine whether it is a capital or operating lease, but also to interpret nonstandard leases with significant judgments (e.g., those with unusual asset retirement obligations). This would allow auditors to review and assess larger samples—even up to 100% of the documents, spend more time on judgemental areas and provide greater insights to audit clients, thus improving both the speed and quality of the audit process.

Another example of machine learning technology currently used by PricewaterhouseCoopers is Halo. Halo analyzes journal entries and can identify potentially problematic areas, such as entries with keywords of a questionable nature, entries from unauthorized sources, or an unusually high number of journal entry postings just under authorized limits. Similar to Argus, Halo allows auditors to test 100% of the journal entries and focusing only on the outliers with the highest risk, both the speed and quality of the testing procedures are significantly improved.

In March 2016, Deloitte announced a partnership with Kira Systems to help "free workers from the tedium of reviewing contracts and other documents" (Kira Systems). Kira's advances in machine learning allowed Deloitte professionals to use the technology to simplify complex documents, allowing for quicker analysis.

While originally designed for contracts, Deloitte auditors now use Kira to find "foregone revenues or reduce third party cost and risk" (Kira Systems), and Kira recently released additional platforms for Deloitte's tax and advisory practices.

Expected Innovations in Machine Learning

In May 2018, PricewaterhouseCoopers announced a joint venture with eBravia, a contract analytics software company, to develop machine learning algorithms for contract analysis ("PwC Announces Legal AI partnership with eBrevia for Doc Review," Artificial Lawyer, 2018, https://bit.ly/2APAKZr). Those algorithms could be used to review documents related to lease accounting and revenue recognition standards as well as other business activities, such as mergers and acquisitions, financings, and divestitures. Deloitte has advised retailers on how to enhance customer experience by using machine learning to target products and services based on past buying patterns.

While the major public accounting firms may have the financial resources to invest in machine learning, small public accounting firms have the agility to use pre-built machine learning algorithms to develop expertise through implementations at a smaller scale.

Many routine accounting processes will be handled by machine learning algorithms in the near future. Accounting processes such as expense reports, accounts payable, and risk assessment may be easily automated using machine learning. The jobs requiring the processing of documents have already started disappearing with the advent of document scanners, optical character recognition, and software to match source documents. As an example, machine learning algorithms can receive an invoice, match it to a purchase order, determine the expense account to charge, and place it in a pool of payments to release; a human worker can review the documents and release them for payment. While accounting jobs in businesses will change in the near future, the question of how the public accounting profession will evolve remains (Younis, 2020).

Given that companies will deploy machine learning in their operations to improve accuracy and reduce costs, the advisory services of public accounting firms could dramatically change. It is estimated that 80% of the time spent in advisories services is processing information about a company's operations. Much of this information processing could be handled by machine learning algorithms, meaning that most of the time billed to clients would focus on valued added services that analyze the information produced by machine learning.

The impact of machine learning will most likely be less pervasive in tax preparation services, due to the need for specialized advice and technical research in the context of complex corporate and individual planning issues.

Auditing is an area that will significantly change in the future. Many have predicted that the automation of analyzing a company's financial statements and source documents will result in smaller audit staffs. Auditing standards, however, require an auditor to understand the systems and processes related to the preparation of the financial statements—meaning that the technical expertise required of auditors to understand the machine learning algorithms used in a company's financial systems will be very different from what it is today. Auditors will need to understand the technologies involved and their interaction with internal controls to avoid material misstatements. Potential fraud in a company's financial statements could

become easier to identify by using a machine learning algorithm to identify trans-actions that have characteristics associated with fraudulent activities.

Global companies face large and increasingly complicated tax compliance requirements; allocating revenue and expenses to various taxing jurisdictions requires significant data processing and analysis. Machine learning can help tax professionals keep up with relevant tax law changes. Creating algorithms to extract relevant planning information from vast amounts of data is ideal for machine learning. It is hard to accomplish effective tax planning without the relevant and important facts; machine learning can make the fact gathering and analysis function much more efficient and effective. In addition, taxing authorities are exploring the use of machine learning to increase transparency and audit efficiency. The IRS has already begun to develop machine learning algorithms to identify patterns that are associated with tax evasion and fraud. Former IRS agent Michael Sullivan indicated that the public should be aware that the IRS has begun using a new audit method, the 'Machine Learning Tax Audit'. As tax laws continue to grow more complex and the IRS's processes for identifying a taxpayer for an audit become more sophisticated, machine learning may allow tax accountants to better predict deductions that will be disputed by the IRS and identify the regulations that allow for those deductions.

7.4 Data Visualization

Data Visualization is the process of selection, transformation, and presentation of various forms of data in a visual form that helps facilitate exploration and understanding.

The main goal of data visualization is to help users gain better insights, draw better conclusions and eventually generate hypotheses.

This is achieved by integrating the user's perceptual abilities into the data analysis process, and applying their flexibility, creativity, and general knowledge to the large data sets available in today's systems.

In general, there are five main stages to data visualization:

- The collection and storage of data.
- The preprocessing of data.
- The hardware used for display.
- The algorithms used to visualize the data.
- The human perceptual and cognitive system (the process of thinking).

In general, there are two categories of data visualization, each serving different purposes: explanation and exploration. Explanatory data visualization is appropriate when we know what the data is and has to say.

In the accounting and auditing literature, prior research has examined the impor-tance of presentation format and its linkages to decision-making performance. They put focus on the comparison of different visual techniques and their impact on

decision-making. Furthermore, the growing number of studies examining presentation format provide an indication of its importance in decision-making.

In contrast, visual data exploration is appropriate when little is known about the data and the exploration goals are vague. Translating large data sets into a visual medium can help in identifying interesting trends and/or outliers. Exploratory data visualization facilitates the user exploring the data, helping them unearth their own insights. Depending on the user's context, it is a discovery process that could or could not potentially lead to the finding of many different insights. Ultimately though, it could help users obtain interesting information, and build hypothesis from large amount of data.

Users who utilize exploratory data visualization generally do not know what the data will show, and would usually analyze and look at the data from a couple of different angles, searching for relationships, connections, and insights that might be concealed in the data. In contrast, users who use explanatory data visualization can typically be called presenters as they are already experts in their own data. They have already explored and analyzed the data and highlighted the data points that support the core ideas they want to communicate.

Specifically, explanatory data visualization is part of a presentation phase, where we want to convey certain information in a visual form. With Big Data, companies need better ways, to not only explore data, but to synthesize meaning from it. Producing visuals that provide explanation and understanding can have significant effects in guiding users toward a conclusion, persuading them to take different actions, or inviting them to ask entirely new questions. Nevertheless, creating such visuals requires preplanning, setting clear objectives, and obtaining the right visual elements.

Data visualization tools are becoming increasingly popular because of the way these tools help users obtain better insights, draw conclusions, and handle large datasets. For example, auditors have begun to use visualizations as a tool to look at multiple accounts over multiple years to detect misstatements.

If an auditor is attempting to examine a company's accounts payable (AP) balances over the last 10 years compared to the industry average, a data visualization tool like PowerBI or Tableau can quickly produce a graph that compares two measures against one dimension. The measures are the quantitative data, which are the company's AP balances versus the industry averages. The dimension is a qualitative categorical variable. The difference between data visualization tools from a simple Excel graph is that this information ("sheet") can be easily formatted and combined with other important information ("other sheets") to create a dashboard where numerous sheets are compiled to provide an overall view that shows the auditor a cohesive audit examination of misstatement risk or anomalies in the company's AP balances. As real-time data is streamed to update the dashboard, auditors could also examine the most current transactions that affect AP balances; thus, enabling the auditor to perform continuous audits. With the real-time quality dashboard that provides real-time alerts, it enables collaboration among the audit team on a real-time continuous basis coupled with real-time supervisory review. Analytical procedures and tests of transactions can be done more continually, and the

auditor can investigate unusual fluctuations more promptly. The continuous review can also help to even out the workload of the audit team as the audit team members are kept abreast of the client's business environment and financial performance throughout the financial year.

7.5 Data Mining

Data Mining is a technique which has advanced classification and prediction capabilities and can contribute to fraud detection.

Sun et al. noted that BDA uses data mining to uncover knowledge from a data warehouse or a big dataset to support decision-making creating predictive models to forecast future opportunities and threads, and analyzing and optimizing business processes.

Hence, the Big Data Analytics offer the capability of capturing "sequential causational and correlational processes" on a real-time basis and may change the financial accounting dramatically and reporting that is legacy based, relied on structured data and successive layers of summary and aggregation and reports on a periodic basis.

8 How to Use ADAs in Risk Assessment Procedures?

One common preliminary analytical procedure that can be enhanced through ADAs is performing a year-over-year general ledger fluctuation analysis.

In Step 1 of the ADA process, auditors plan to evaluate all financial statement line items for risk, with the general purpose of assessing the risk of material misstatement for the audited entity's year-end balances. Auditors perform the analysis using year-over-year comparisons in Tableau. Using professional judgment, auditors decide that all accounts that have changed by more than $three million are deemed "notable" and merit additional investigation.

Step 2 of the process is to access and prepare the data. In this example, this process is relatively straightforward, because the data used in this ADA is the trial balance provided to the auditor by management, as well as audited trial balances from the past four years. Because these trial balances exist in standard spreadsheet format, loading them into Tableau™ is a straightforward process.

Step 3 considers the relevance and reliability of the data. The trial balances are clearly relevant as they relate directly to the test of year-over-year differences in account balances. Further, the historical audited data is reliable based on prior year audit results. Because the entity's control environment was effective during interim testing, and because this ADA is a preliminary analytical procedure and not a substantive test, no further tests were performed on the trial balance's data reliability other than agreeing on the beginning balance to the prior year's audited trial balance.
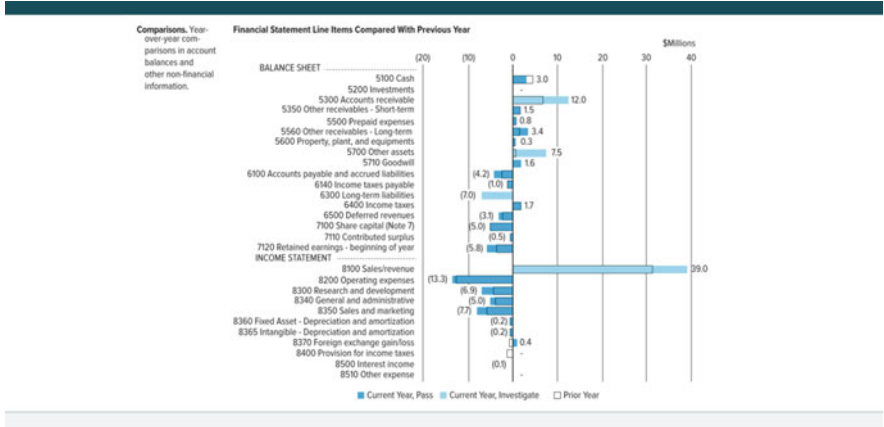
Fig. 1 Audit Data Analytics (ADAs)

Step 4 is to perform the ADA, shown in Fig. 1. Dark blue bars represent year-over-year fluctuations that are below the $three million threshold, while light blue bars are those that exceed the threshold. Accounts Receivable, Other Assets, Long Term Liabilities, and Revenue all changed by amounts exceeding the threshold. Based on previous inquiry, the significant change in the year-end Accounts Receivable (AR) balance was most unexpected. Therefore, in response to this initial ADA, auditors follow up with an additional ADA focusing on AR. Because of the audited entity's multinational customer base, the auditors use a trend analysis to examine AR balance, by currency, over the past 5 years.

9 How to Use Audit Data Analytics (ADA) in Substantive Analytical Procedures?

Audit data analytics used as a substantive analytical procedure often use more data, or different data (for example, more disaggregated), with different techniques than traditional substantive analytical procedures. However, when auditors use Audit Data Analytics (ADAs) as substantive analytical procedures, they must be careful to follow the explicit audit guidance related to substantive analytical procedures including careful documentation of how the auditor developed an expectation, defined a tolerable difference, performed the analytic, and investigated differences.

10 How to Use ADAs in Tests of Detail?

To this point, ADAs have been presented primarily as more advanced analytical procedures and visualizations. However, ADAs can also be used as a test of detail because of their ability to process and analyze large amounts of data. For repetitive

processes and calculations, ADAs can be used to test entire populations that have previously been tested using samples. For example, instead of testing a sample of contracts for revenue purposes, software can be used to have the computer "read" *all* contracts and search for predetermined key phrases that indicate side agreements or unusual terms that indicate a heightened risk of material misstatement.

References

Alles, M. G., Kogan, A., & Vasarhelyi, M. A. (2002). Feasibility and economics of continuous assurance. *Auditing: A Journal of Practice & Theory, 21*(1), 125–138.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. (2015). Moving towards continuous audit and big data with audit analytics: Implications for research and practice. *Rutgers Accounting Research Center (RARC) and Continuous Auditing & Reporting Laboratory (CAR-Lab)*.

Balios, D., Kotsilaras, P., Eriotis, N., & Vasiliou, D. (2020). Big data, data analytics and external auditing. *Journal of Modern Accounting and Auditing, 16*, 5.

Dickey, G. (2019). Machine learning in Auditing current and future applications. *CPA Journal, 2019*(June), 1–10.

Jesus, M. (2018). *How machine learning is disrupting the accounting industry*. BigML Blog. https://bit.ly/37hIvU5

Rajvanshi, A. (2016). *Predictive Analytics/machine learning for fraud detection*. Business Intelligence Engineer at Amazon.

Verma, R., & Mani, S. R. (2015). Using analytics for insurance fraud detection. *FINsights, Infosys, 10*, 1–23.

Younis, N. M. M. (2020). The impact of big data analytics on improving financial reporting quality. *International Journal of Economics, Business and Accounting Research (IJEBAR) Peer Reviewed – International Journal, 4*(3) E-ISSN: 2614-1280 P-ISSN 2622-4771.