Data Analytics Knowledge Required of CPA:  

A Normative Position

Deniz Appelbaum  
Assistant Professor  
Montclair State University  
Department of Accounting and Finance  
Office: 973-655-7689  
apprlbaumd@montclair.edu

D. Scott Showalter  
Professor of Practice  
NC State University  
Poole College of Management  
Department of Accounting  
Office: 919.513.0526  
scott_showalter@ncsu.edu

Ting Sun  
Assistant Professor  
The College of New Jersey  
Department of Accounting & Information Systems  
Cell: 732-310-6063  
sunt@tcnj.edu

Miklos A. Vasarhelyi  
KPMG Distinguished Professor of AIS  
Director of the CARLab  
Rutgers, the State University of New Jersey  
Cell: 201-454-4377  
miklosv@andromeda.rutgers.edu

Acknowledgements: The authors would like to thank Hanchi Gu of XXX University for his assistance in the preparation of this manuscript.
Data Analytics Knowledge Required of a CPA: A Normative Position

Abstract: Due to the nature of the current data driven business environment, many accounting firms are starting to re-align their processes to incorporate technology and Audit Data Analytics (ADA), as the traditional audit procedures are proving to be inefficient and ineffective and not meeting market expectations. This paper provides commentary on why data analytics knowledge should be required of the profession, in addition to that of traditional accounting. We discuss the current information-centric business environment, the availability of Big Data, and the existing data analytics efforts made by businesses, with its subsequent impact. Regarding the complementarity of available data analytics tools and data analytics knowledge, it proposes a guideline for the content and levels of ADA knowledge and skills that should be required of auditors in different positions. Finally, suggestions are provided to facilitate the adoption of ADA and provide solutions to challenges in CPA exam, audit standards, and education. In this data-centric business environment, acquiring the knowledge and skills of data analysis should be a current professional priority.

Keywords: Audit Data Analytics, Knowledge, Skill, Accounting Education, CPA Exam, Data Analytic Techniques
I. INTRODUCTION

Audit data analytics (ADA) is “the science and art of discovering and analyzing patterns, identifying anomalies, and extracting other useful information in data underlying or related to the subject matter of an audit through analysis, modeling, and visualization for the purpose of planning or performing the audit” (AICPA, 2017). In today’s real-time technology driven digital business environment, what level of ADA knowledge should generally be required of modern CPAs and CPA candidates? In a survey study (Crawford et al, 2011) practitioners rated data analytics knowledge as the most desirable skill for graduating student hires. Accounting firms recognize that leveraging ADA to serve their clients requires a broad paradigm change from the traditional training approach – both in employee skills and available training methodologies. To reflect this growing demand for the use of ADA, one would expect that future and current CPAs would exhibit a mastery of ADA knowledge and skills. Additionally, the AACSB has released new requirements for accounting department accreditation that demand technology skills of graduates¹. However, students complete accounting programs with few analytical skills and are unprepared for the emphasis by firms to use more ADA in engagements (PwC 2015).

To be proactive in this environment, CPAs should build a solid foundation with a knowledge of ADA and statistics beyond traditional ratio analysis and sampling (PwC 2015). They should have more than a cursory understanding of data analysis and how it is embedded in their audit tools. Firms have traditionally educated their staff about the use of certain standardized commercial audit software packages and have integrated certain features of these applications in their engagements, but these tools do not include modern analytics such as text mining, sentiment analysis and other data mining solutions. Consequently, although these applications can perform tests on 100% of the data instead of sampling a small portion, the applications are primarily being used to automate traditional manual tests.

While internal auditors are required by International Standards for the Professional Practice of Internal Auditing to “have sufficient knowledge of key information technology risks and controls and available technology-based audit techniques”, external auditors have no similar requirement

¹ https://www.aacsb.edu/accreditation/standards/accounting
to conduct data analysis work (Wang and Cuthbertson, 2015). For example, the AICPA requires that the engagement team “perform overall analytical procedures” and “confirm balances” and “perform other tests of details, balances, and journal entries” (AICPA, 2014). However, “today, many audit processes are essentially unchanged from those performed decades ago, even though newer technology may be used to perform them more efficiently” (AICPA, 2014).

This normative paper provides commentary on why ADA is needed by discussing the current information-centric business environment, Big Data analytics, and the existing data analytics efforts made by business as well as their influence on the auditing profession. Then the complementarity of available data analytics tools and data analytics knowledge and skills is examined. The paper next delves into the realm of ADA techniques and proposes a guideline for the content and levels of ADA knowledge and skills that should be required of auditors in different auditor roles. Then follows a discussion of the challenges of the CPA exam, the audit standards, and education. The paper then concludes with recommendations for future research.

II. THE BUSINESS ENVIRONMENT

The information-centric economy

The exponential rate at which information technology is evolving is so fast that it has fundamentally changed every aspect of the way business operates (Olson, 2016). Today’s business has become information centric. The development and deployment of global IT connections, data collection technologies (such as RFID and Barcodes), blockchains and cryptocurrencies (Deloitte, 2015: EY, 2018), eXtensible Business Reporting Language (XBRL), electronic data interchange (EDI) formats, cloud computing, and business databases, facilitates business data creation, collection and utilization (Chen, Chiang, and Storey, 2012). With such continuing innovations, business processes are constantly evolving. Such a digital business environment creates a demand for employees and other related parties to be able to adopt data analytics in a flexible or even creative manner (PwC, 2015). Business leaders are generally

---

2 On June 20, 2019, the AICPA issued an exposure draft for a Proposed Statement on Auditing Standards, Audit Evidence, to update auditing standards to recognize sources of information and technologies not available then the standards were last updated.
concerned that employees, including accountants, possess the essential analytical skills to extract value from the data, and these needs are magnified with the expansion of Big Data.

**Big Data Analytics**

Characterized by its 4Vs - volume, velocity, variety, and veracity (IBM, 2012; Laney, 2001), Big Data are generated and flow across a variety of different aspects of business. “With the advent of UPC scanners, data is expanded in scope by collecting details, for example, of items being acquired at the cash register” (Vasarhelyi, Kogan, and Tuttle, 2015). Radio Frequency Identification (RFID) allows a business to identify inventory and products and to track them throughout the supply chain (Davis, 2014). During this process of tracking, inventory and product-related data are created, stored and transmitted Big Data also originates or is collected from external social media providers like Facebook or Twitter. Unlike structured financial or transactional data which are understood with standard data analysis tools, most Big Data are unstructured or at least semi-structured data (such as text, video or audio). It needs to be organized, characterized, tagged, and categorized to generate metadata to form useful information for decision making. In other words, with the access to Big Data and with an all-encompassing deployment of automated data generation and recording, businesses have the opportunity to analyze data, thereby transforming it to usable, viable information to make knowledgeable decisions.

Furthermore, the information Big Data provides as a product of analytical activities may be valuable evidence for auditing (Yoon, Hoogduin, and Zhang, 2015). An enhanced method of auditing can exhibit the following characteristics:

- continuous auditing and continuous monitoring
- frequent usage of data analytical technology and tools
- testing the population, rather than subset or samples
- highly automated audit procedures and processes
- analytical and information technology competence
- more emphasis on critical thinking and complex problem-solving abilities
- utilizing technologies of blockchain, and smart contracts (Rosario, 2019)
- using machine learning, and other AI techniques in modules during the audit (Sun and Sales, 2018; Sun, 2018a)
applying robotic process automation as appropriate (Zang, C., 2019; Zhang, Rosario, and Cohen, 2019)

Analytics efforts made by business
Realizing the value of big data, companies are investing in a variety of data analytical techniques to examine existing business information and gain insights to better support the company and serve their customers. A survey (IDG, 2016) of 1060 participants from a variety of entities, finds that 53% of companies are implementing or planning to implement data-driven projects with the goal of improving customer relationships, making business more data focused, or changing ways operations are organized. Moreover, 44% of organizations expect to increase their data-driven initiative budget in the next twelve to eighteen months.

III. ADA TOOLS

Technology tools have become increasingly important in audit engagements (Wang and Cuthbertson, 2015). Recognizing this demand, vendors are providing ADA tools and Computer Assisted Auditing Techniques (CAATs) that use modern development interfaces, API calls, and web services for auditors to effectively process and analyze data and to conduct basic audit tasks.

The use of those tools by firms, as evidenced by their publications summarizing their audit data analytics efforts, is changing the required knowledge and skills set of auditors. For example, in the past, auditors were typically required to know statistical sampling and professional accounting and auditing. Nowadays, with the availability of data and the advancement of analytical tools, such as SAS, SPSS, Python, and R, and CAAT’s (i.e., ACL, Arbutus, and IDEA), auditors may need to understand data science and information technology in addition to acquiring accounting knowledge. However, unlike data scientists and other specialists in information systems and computer science, auditors do not need to have a deep and comprehensive mastery of these technologies to effectively leverage them. That is, auditors should possess sufficient skills and knowledge to effectively and efficiently conduct audit analytics with those tools. Table 1 provides a brief description of the complementarity between knowledge and tools:
Because many data analytical tools apply quality-tested machine learning algorithms that are constantly updated to reflect the latest statistical methodologies and automatically select and rank most suitable algorithms for the input data, auditors do not need to have a deep understanding of those algorithms and statistical methodologies. This is similar to how most drivers do not need to know how a car works to drive a car.

Tools like H2O\textsuperscript{3}, SAS and SPSS provide easy-to-use Graphical User Interface (GUI) giving users the power of data analysis without having to write code; all they need to do is click or drag. For sentiment analysis, auditors should not be expected to understand sophisticated linguistic theory or manually create dictionaries because the text mining module of those tools, such as SAS Contextual Analysis, have been embedded with sophisticated linguistic rules and analytical modeling tools. With natural language processing capabilities powered by AI, IBM Natural Language Understanding\textsuperscript{4} has advanced text analytics features to extract concepts, sentiment, emotion, entities, relationships, keywords, semantic roles and more. Data visualization tools are more efficient, as they can generate task-specific graphics for more effective interpretation and communication of results; building analytical-style graphs, maps and charts with any style of output to deliver analytic results where they can be used the most.

To maximize the benefits of these tools, auditors should understand the usage of databases and data warehouses, and accounting information systems. Additionally, the ability to prepare, preprocess, and analyze data, and interpret result of analysis is critical. It would also be helpful to acquire a general knowledge of emerging technologies such as robotic process automation, intelligent process automation, machine learning, blockchain, and dynamic visualization.(Dai and Vasarhelyi, 2017; Freakonomics, 2008; Adadi and Berrata, 2018; Issa, Sun, and Vasarhelyi, 2016). These are discussed in the next session.

IV. ADA KNOWLEDGE AND SKILLS REQUIRED OF AUDITORS

\textsuperscript{3} H2O is a fully open source machine learning platform. H2O supports the most widely used statistical & machine learning algorithms. This tool is used by over 18,000 organizations globally. https://www.h2o.ai/products/h2o/

\textsuperscript{4} https://www.ibm.com/watson/services/natural-language-understanding/
To address these changes, auditing practices have become increasingly multi-disciplinary. Accounting firms are investing in the latest analytics technologies for their advisory and tax engagements, which eventually will require auditors to acquire a more extensive knowledge base and to obtain enhanced and diversified skills sets (Chaffey, Van Peursem, and Low, 2011). One firm’s advisory client is another firm’s assurance client, and the advances realized by the advisory service may impact the audit engagement. Likewise, most major accounting firms have recently announced significant investment in data analytics for their audit practices (PCAOB 2017; KPMG, 2017). A proposed structure of the required knowledge and skills is described in Figure 1. This knowledge and skill set consists of three layers. Auditors should be required to obtain strong knowledge of the first two layers as these help auditors establish the foundation to understand and analyze data. Depending on the specific analytical task (based on audit type and audit function), auditors may also be required to apply at least one of the techniques described in the third layer and it is highly recommended that, to improve analysis capabilities, they understand some advanced data analytical techniques.

(Insert Figure 1 about here)

The following section further explores the structure of knowledge/skills required of auditors.

**Comprehensive understanding of data**

For most public companies, especially multinational ones, the IT department is responsible for the management of their ERP systems, legacy systems, and databases. Traditionally the auditor requests the IT department for the data which queries the database, generates the dataset, and then delivers this data. Progressively companies are creating their own audit data warehouses and/or allowing auditors to do direct extractions even from live systems. Because auditors and IT personnel approach requests for information from different perspectives these two approaches may lead to different results. In general data analytics efforts need substantive effort in data cleaning and preparation. Several tools are available for this purpose such as Alteryx and Tableau Prep. As auditors exert significant effort and perform time-consuming procedures to

---

5 http://www.alteryx.com  
6 http://www.tableau.com/prep
To optimize the efficiency and effectiveness of such audit analytics applications the AICPA’s Assurance Services Executive Committee’s Emerging Assurance Technologies Task Force has proposed and developed the Audit Data Standards (ADS\(^7\)). Providing a standard data format for the files and fields needed for financial statement auditing, the ADSs facilitate automation and enhancements in business information analysis. Efforts in Robotic Process Automation (Rosario, Moffitt, and Vasarhelyi, 2018) had further emphasized the need for common data formats in order to allow software such as UIPath\(^8\) and AutomationAnywhere\(^9\) to work on an overlay basis and replace extensive amounts of auditor labor. Unfortunately, the ADSs are non-authoritative. As a result, auditors working for firms or working with clients that do not voluntarily follow the ADS still need to obtain a comprehensive understanding of the data. To achieve this level of understanding, they should possess knowledge regarding of the firm’s business and processes, its ERP system, the basics of databases, and Big Data. Table 2 provides a summary of skill category, knowledge areas, and example of skills for a comprehensive understanding of data.

A description of each data skill objective follows:

\textbf{(1) Firm’s business and the characteristics}  
AU-C Section 315 (AICPA, 2012a) states that auditors should obtain a sufficient understanding of the entity and its environment to assess the risk of material misstatement. Additionally, specific to audit analytics, auditors should focus on data since it is generated by and reflects the operations of various business units and cycles. Without a general understanding of the business environment and industry, the auditor’s knowledge on how data is produced and collected, and the nature and characteristics of the data is limited.

---

\(^7\) For copy of Audit Data Standards see http://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/pages/auditdatastandardworkinggroup.aspx.
\(^8\) https://www.uipath.com/
\(^9\) https://www.automationanywhere.com/
(2) **Nature of the firm’s ERP system**

Auditors should understand the firm’s ERP system from which the data is obtained. While sharing certain common features, ERP systems in different companies vary from one another. It therefore is necessary to understand the nature and structure of the ERP system implemented by the company. Many large corporations employ several different information systems because they have merged with and/or acquired other companies with varied systems (Bae and Ashcroft, 2004).

(3) **Basics of a database system**

ERP systems enable one single, relational database to store an entity’s data, although large companies may have many instances of an ERP or many different ERPs interacting with legacy systems. Such databases provide complete sets of data available to auditors for a holistic view of the entity’s performance. A database management system (DBMS) is a “software designed to assist in maintaining and utilizing large collections of data” (Ramakrishnan and Gehrke, 2000). It is important that auditors understand the terminology of databases, their concepts, the major database users, and the basics of database management systems.

Auditors should be able to identify the design of the DBMS architecture. Specifically, understand whether the architecture of a DBMS is centralized, decentralized or hierarchical. Understanding the architecture of a DBMS helps auditors develop an overall understanding of how various identifiable components (i.e., data, software modules, metadata, and interfaces) are structurally related. Data models define how the logical structure of a database is organized in a DBMS and are equally important to auditors. Having essential knowledge of data modeling assists auditors to understand how data fields connect to each other and how data is processed and stored inside the system. Data modeling establishes the context of the recording of the actual transactions that occurred within the company and should be referenced by the auditor during the audit.

While DBMS is the overall system that manages a database, a data warehouse can be treated as such a database, with special facilities for analysis and reporting to provide insights that enable decision making. Traditionally, data warehouses mainly analyze structured, transactional data within relational databases built with the SQL programming language. Nowadays, the emergence of Big Data and NoSQL (or nonrelational) software platforms allows data
warehouses to extract unstructured data from various data sources. Data from social media, mobile, GPS, and smart devices is collected via sensors, transformed or structured for querying and analysis purposes, and loaded into the targeted data warehouses. In this way, the new data can be used to obtain more business insight from warehouse applications.

As a result, auditors should understand this process of Extract, Transform and Load (ETL)\(^{10}\) in data warehouses. Furthermore, auditors are expected to have the basic knowledge of NoSQL databases and, especially, its capability to support dynamic schema design, which offers increased flexibility, scalability, and customization, as compared to relational databases, for handling large amounts of unstructured data.

(4) Metadata of Big Data

Big Data is a mix of large volumes of structured, semi-structured, and unstructured data that conventional analytics has difficulty processing. The data should dynamically be categorized, characterized, and tagged, all of which generates additional voluminous amounts of metadata. Metadata is the data (information) about data. For example, metadata about a call made with a cellular phone will include phone numbers, GPS locations, and length of the call but not the call text. The need for Big Data further expands the skills needed to be mastered by auditors as they will need to understand various types of metadata.

For video data, auditors should understand the format and content of metadata stored within videos. Such metadata supplies itemized descriptions for a video file. The descriptions provide auditors with sufficient information about the video, which make searches of video content much easier (Telestream, 2008). The metadata of videos include the author, keywords, file size, media format, people, the place, closed captioning, subtitling information, and so on. Video data could also be used by auditors to identify impairment of plant, property and equipment in remote locations or assess damages from a natural disaster where physical presence by the auditor may be difficult or not possible.

Furthermore, auditors may want to test the entire population instead of sampling (No et al, 2019). Sampling 50 – 60 records from a Big Dataset provides little audit evidence. However,

\(^{10}\) Extract, Transform, and Load (ETL) refers to the overall design process of data warehousing. To summarize, the data may be extracted from many different types of data, then these multivariate datasets are transformed into data structures that can be stored and analyzed, and which are then loaded into the data warehouse for storage.
100% testing may present another challenge: the transactions flagged as exceptions from these tests may be voluminous in number compared to a small subset of sampling. The auditor then faces the task of reducing this dataset further by prioritizing the results by degree of exception – i.e. “exceptional exceptions for further analysis” (Issa, 2013).

**Solid foundation of statistics**

Statistics is “the study of the collection, analysis, interpretation, presentation, and organization of data” (Dodge, 2006). Auditors do not need a degree in statistics to become more efficient when dealing with data. But having a mastery of basic statistical concepts, principles and theories can help them understand, accurately interpret and present data, as well as more effectively use the various available software tools. Table 3 describes required statistical knowledge and skills.

(Insert Table 3 about here)

A description of each statistical skill category follows:

1. **Descriptive statistics**
   
The first step of audit analytics is describing data. Descriptive statistics utilizes numbers, tables, charts, and graphs to inform auditors about the raw scores presented in the data. Descriptive statistics are most often used to examine central tendency (location), dispersion (variability), skew (symmetry), and Kurtosis (peaked-ness) of data. When the objective of audit analytics is to generally describe and discuss a significant amount of a data set, qualitative and/or quantitative, descriptive statistics are highly recommended. Moreover, an understanding of descriptive statistics is necessary for the effective use of hypothesis testing, correlation, regression analysis, and other normative and cause-and-effect statistical techniques (Accountability Modules, 2012b).

2. **Statistical Inference**
   
   Currently used statistics consists of two different categories: Classic statistics and Bayesian statistics. Classical statistical inference is helpful when one tries to further investigate the relationships among data and to answer cause-and-effect questions, because classical inferential statistics, based on information obtained from the current data samples, allows auditors to draw conclusions and/or make predictions about the properties of a population (Accountability
For instance, what characteristics (i.e., annual income, delinquency history, and age) of the customer are related to credit card delinquency. Auditors should be trained to understand the basic form of statistical inference-hypothesis testing. Hypothesis testing can be employed to test an assumption about population parameters based on samples from such populations. For example, auditors can use hypothesis testing to assess the probability that a management assertion about a specific population or condition is correct.

While classic statistics usually claims that truth is fixed and observations are random, Bayesian statistics generates probability statements about possible states of the truth and therefore could be more relevant to financial statement audits. It represents statistical estimation as the conditional distribution of parameters and unobserved data. Bayesian inference is based on “Bayes Theorem”, which offers clues regarding the revision of probability statements using sample data. Bayesian inference provides a useful framework for auditors’ judgement making process in the light of new audit evidence, particularly if this evidence appears complex or contrary to the auditor’s expectations.

To support their decision making, auditors continually seek to arrive at more logical, effective, and less subjective judgments. Therefore, auditors should acquire the knowledge of how to apply Bayesian inference in the audit pre-engagement judgment process. For example, a model can be built based on a prior probability distribution (distribution of the model that explains the relation between auditor’s judgment and the independent variables). The dependent variable would be the auditor’s judgment of client acceptance. The independent variables would include a company’s characteristics, opinions from previous auditors and related parties, and the expected relationship between auditor and the company (Erdogan and Uludag, 2014).

**Proficient in advanced data analytical techniques**

When auditors explore data with analytical techniques, they first should identify data trends, patterns, and correlations. Additionally, data analysis can be categorized as both exploratory and confirmatory (Tukey, 1980; Liu, 2014). While Confirmatory Data Analysis (CDA) aims to test whether the data fit pre-hypothesized measurement model, Exploratory Data Analysis (EDA) explores data to generate a hypothesis. CDA may seem to be more familiar to auditors, with its objective of confirming expectation models. However, EDA techniques, especially those
advanced approaches (e.g., advanced data visualization, machine learning, and process mining), are gaining momentum in today’s real-time economy.

In contrast to CDA, EDA emphasizes pattern recognition and hypothesis generation when processing raw current data, which provides clues for inspiring ideas (Tukey, 1980). Facing the high 4Vs of Big Data, auditors cannot merely rely on traditional EDA approaches such as descriptive statistics, basic data visualization techniques (e.g., pie chart, column chart, bar chart, and stem-and-leaf plot.), and data transformation techniques. Rather, auditors need to acquire related knowledge and skills of more advanced DA techniques (ADA).

**Requirement level of knowledge and skills for all ADA techniques**

It would be costly and unnecessary to require all auditors to obtain the same level of ADA knowledge and skills because they hold different positions with different responsibilities. Therefore, required knowledge and skills for all types of advanced ADA techniques are divided into three levels:

Basic level:

- understand basic concepts and theories
- use the basic function of the tools/applications
- interpret the results

Median level:

- understand key concepts and theories
- use key function of tools/applications
- interpret the results
- design and develop models

High level:

- have a full understand of all key concepts and theories
- select appropriate tools based on features of data and task
- design and develop statistical or machine learning models
- evaluate models using metrics
- interpret results and use the results to support audit decisions
- integrate the method or tool into audit plan/audit procedures

**Positions and responsibilities for auditors**

There generally are four types of positions in an accounting firm: Staff, Senior, Manager, and Partner. The AICPA\textsuperscript{11} briefly describes the nature and characteristics of the roles and responsibilities for auditors in different positions.

The Staff auditor performs the detail work of the financial audit under the supervision of a Senior. The Senior auditor directs audit field work under the general direction of a Manager. In turn, the Manager auditor supervises Seniors and Staff. A Partner faces a high level of responsibilities, including overall responsibility for the conduct of the audit.

The required knowledge/skills are based on role of the auditor. For certain techniques (process mining, text mining, and artificial intelligence) that require specialties, it is necessary to form a specialized team to exclusively conduct data analysis work in support of engagement teams. Other auditors only need to have a basic knowledge to facilitate the understanding of the final results of analysis. Table 4 displays the required advanced knowledge and skills and relates each to auditors in different positions.

(Insert Table 4 about here)

A description of each ADA skill category follows.

**Visualization techniques**

Simply examining or displaying numbers of a grid cannot fully reveal the underlying pattern of data and, in some cases, it may even mislead auditors to draw wrong conclusions about the data (Evelson and Yuhanna, 2012). Data visualization is a powerful tool to identify patterns and relationships among voluminous numbers of variables and determine the relative importance. As extensions of basic data visualization techniques, advanced data visualization techniques display data in more sophisticated ways, allowing more variables to be shown in one graph (Heer, Bostock, and Ogigevevsky, 2010), which ensures staff and senior auditors examine

\textsuperscript{11} http://www.aicpa.org/CAREER/CAREERPATHS/PUBLICACCOUNTING/Pages/default.aspx
and analyze data in a more instinctive way. Modern technologies have facilitated the use of more dynamic and interactive business graphics, such as heat maps, geographic maps, and real-time dashboards and charts that update automatically as the data changes (Liu, 2014).

**Classification/regression techniques**

Databases typically may have significant amounts of unidentified information which exhibits potential to contribute towards intelligent decision making (David and Balakrishnan, 2010). Classification/regression methods are frequently employed to develop models describing data classes and predict group membership for data instances. As one of the most frequently used classification/regression methods, decision tree techniques result in a flow chart-like tree structure, where each node represents a test of an attribute and each branch represents an outcome of the previous decision. This step-wise conditional logic is presented in Figure 2, which displays a tree illustrating the decision-making process for identifying possibly fraudulent transactions. This technology is often used to help staff and senior auditors perform predictive analytics, such as detecting management fraud. In contrast, it is difficult for traditional audit procedures like ratio analysis to provide predictive analysis (Porter & Cameron, 1987; Coderre, 1999; Kirkos, Spathis, and Manolopoulos 2007).

**Association analysis techniques**

Association analysis deals with discovering hidden, interesting associations and/or relationships among large datasets (Agrawal, Imielinski, and Swami, 1993). It specifically searches for frequent patterns, associations, correlations, or rules according to the occurrences of one data point based on the occurrence of other data points in the dataset. The uncovered relationships are represented in the form of association rules (Pang-Ning, Steinbach, and Kumar, 2006). With the ability of revealing patterns or associations among data, association analysis has been broadly used in the business field to help effectively make business decision (Moreno, Segrera, & López, 2005). In auditing, association analysis can be applied to items fields, such as purchase items or sold/returned items (Liu, 2014).

---

12 Deep Neural Networks technique belongs to the category of classification/regression techniques, but we discuss it separately later as it is an emerging approach that has not been applied in accounting and auditing as much as other traditional machine learning techniques.
**Text mining techniques**

Text mining could be regarded as a Big Data-oriented analytics approach, because one major type of Big Data is textual data. Text mining of emails, news reviews, or call transcripts provides more independent and abundant supporting evidence for audit decision making process. Gray and Debreceny (2011) utilize text mining to analyze Enron email database and explore the role of text mining in the support of continuous monitoring over business processes and continuous assurance on the integrity of the financial reporting process. Popular text mining software (OpenText, IBM SPSS Text Analytics, Lexalytics, IBM Natural Language Understanding) with capabilities of entity extraction, topic categorization, sentiment analysis and document summarization could further help auditors automate text analysis procedure and identify predictable variables for improved predictive models Sun (2018b).

**Process mining techniques**

In contrast to text mining, process mining focuses on procedures and work flows rather than text. It allows for the discovery of knowledge about process controls and improvement or deficiencies by analyzing “event logs”, i.e., a log recording the execution of activities in a business process (Van der Aalst, Reijers, and Song, 2005; Van der Aalst, 2011). The objective of process mining is twofold: (1) to solve problems by analyzing data recorded in the database and (2) to extract information from the dataset and transform it into an understandable structure for further usage.

In audit analytics, process mining of event logs provides a comprehensive mechanism in which all available data can be processed, in contrast with traditional substantive procedures that rely on a small sample of the population. Process mining techniques have been utilized to develop detection models to identify potential fraud (Yang and Hwang, 2006), to detect potential internal fraud risk (Jans, 2009), and to uncover anomalous transactions in analytical procedures (Jans, Alles, and Vasarhelyi, 2014).

**Neural Networks techniques**

As an emerging machine learning technique (Sun, 2018b) for classification and regression problems, Deep Neural Network (DNN) is an information processing paradigm inspired by the way biological nervous systems (such as the brain) process information (Birdi, Aurora, & Arora, 2013). DNN are particularly beneficial in applications where the underlying process is complex –
i.e., forecasting consumer demand and stock market, evaluating credit risk, predicting the probability of marketing mail, and detecting fraudulent transactions for insurance companies. Previous research proposes that traditional artificial neural networks can be applied to conduct many relevant audit analytical tasks such as predicting bankruptcy (Wilson and Shared, 1994), assessing financial position of the client (Etheridge, Sriram, and Hsu, 2000), conducting audit planning (Kosivaara, 2000) and in estimating manipulation risks in financial statements (Lin, Hwang, and Backer, 2003). In March of 2016, KPMG announced plans to apply IBM’s Watson cognitive computing technology (IBM, 2016), an AI technology which uses DNN (Hinton, Osindero, and Teh, 2006; EY, 2017) to analyze massive amounts of unstructured data as part of its advisory services offerings. Thanks to its multiple layers and numerous units within a layer, DNN is typically powerful for processing and analyzing natural language and unstructured text data. Supplied with a large amount of training data (e.g., conversations from social media, emails, and call transcripts), A DNN mimics human intuition and automatically learns the inner features and develops a new way of presentation for the data without human invention. An increasing number of deep learning applications in accounting (e.g. Sun, 2018a) utilize these techniques.

Computer-Assisted Audit Techniques (CAATs)

The emergence of Computer-Assisted Audit Techniques (CAATs) (for instance, ACL, IDEA) further facilitates the effectiveness and efficiency of audit analytics. With powerful visualization and dashboard creation, these software packages are capable of standardizing and streamlining key audit analytics functions, allowing auditors to integrate standard ADA capabilities into comprehensive audit tasks like risk assessment, audit management, exception identification, and issue tracking. Auditors could take advantage of the tools and integrate audit analytics throughout the entire audit process. For example, Figure 3 summarizes the processes and expectations for leveraging ADA in the engagement as suggested by KPMG (2012). While the processes of steps 1, 2, 3, 4, 7, 9, 10, and 11 are part of the current audit process, in ADA, these steps require a greater depth of analytics knowledge than is typical for most current auditors or accountants. Table 4 also summarizes ADA techniques required for auditors at different positions.
V. CHALLENGES TO IMPLEMENTATION

The ambiguity of DA requirements from the perspective of the CPA exam (AICPA, 2015) and the audit standards presents major challenges for the effective application of ADA analytics. The AICPA announced the blueprints for 2017 CPA exams after polling practitioners, academics, and businesses as to the current skills required of newly licensed CPAs in a practice analysis (AICPA, 2016). In the section of Business Environment and Concepts, it requires the candidate to “recognize the role of big data/data analytics and statistics in supporting business decisions” (AICPA, 2016). However, both the blueprints and the exposure draft merely identify the specific tasks that should be accomplished with the skill of analysis without specifying the exact types and the level of data analysis candidates are expected to perform as well as the nature of the assertion being tested. This ambiguity reflects a similar level of vagueness found in the professional audit standards.

While internal auditors are required by International Standards for the Professional Practice of Internal Auditing to “have sufficient knowledge of key information technology risks and controls and available technology-based audit techniques”, external auditors have no similar requirement to conduct data analysis work (Wang and Cuthbertson, 2015). According to AU-C Section 520 about Analytical Procedures (AICPA, 2012b), to conduct substantive analytical procedures the auditor should (1) determine the suitability of a certain substantive procedure, given the account; (2) evaluate the reliability of the data from which these ratios are developed; (3) develop an expectation of recorded amounts and ratios and whether these are accurate; and (4) finally, determine the amount of difference (if any) between the recorded amounts and the auditor’s expected values and if the difference is significant or not. The standards provide little specific guidance about ADA; rather they provide guidelines that leave much flexibility for auditors as long as the audit objectives are achieved.

Unfortunately, the curriculum design of most current accounting programs still focus on accounting, tax, and finance theory (Sledgianowski, Gomaa, and Tan 2017; Vasarhelyi, Teeter, and Krahel, 2010), ignoring those more practical fields such as applied statistics, ADA.
techniques, accounting information systems, and accounting applications on computers, all of
which prepare students with essential ADA competence. Topics related to Accounting
Information Systems (such as ERP, XBRL and E-commerce) are usually just introduced briefly
in accounting courses. Even among those universities that provide Accounting Information
Systems (AIS) courses that cover information technology (IT) topics (i.e., Enterprise Resource
Planning systems, Accounting Information Systems, E-commerce), the AIS courses struggle to
cover the technologies needed by accountants. This leads to the conclusion that perhaps the
materials, content, and level of the entire accounting curriculum should be expanded to cover
many of these emerging techniques. The recent AACSB A5 standard requires that accounting
departments show how analytics have been integrated in courses at all levels (see footnote 1).

More ADA and use of various audit software are needed across the entire curriculum, not just
focused in a few advanced courses. To possess advanced ADA competence, the mastery of more
types of software is necessary. In addition, case studies can help students effectively use these
tools with real-world data. The value of cases in accounting curricula is well documented in the
literature (Vasarhelyi, Teeter, and Krahel, 2010; Gelinas, Schwarzkopf, and Thibodeau, 2008;
McConnell and Sasse 1999). Nevertheless, it has not been sufficiently incorporated into the
design of relevant accounting information systems courses.

VI. RECOMMENDATIONS

Many firms have integrated CAATs software into the engagement which tests 100% of the
transactions (KPMG, 2015). This base level of automation is creating new engagement
expectations for clients which are currently putting pressure on the entire audit profession to
evolve. As PwC noted regarding assurance services, “more companies are asking their
professional service providers to assist with compliance activities…and are also asking
professional services firms to take subsets of the data that they already collect and use them for
other purposes, like financial and risk analysis” (PwC, 2015). Much of the impetus for the
adoption of analytics seems to originate from audit clients as business has become more
information and data centric.

As these clients demand and expect more value from their audit experience, this demand has
compelled firms to innovate and expand beyond traditional practice. As result of firm efforts to
date, educators may need to enact changes in accounting programs to prepare students for use of ADA within the profession. PwC concluded there are three primary ways that universities can better prepare their students for the modern business climate (PwC, 2015). First, universities “should infuse analytical exercises into existing curriculum to help students develop data analytical proficiency on top of their core accounting skills.” Secondly, technical skills in accounting programs should be emphasized by adding an analytics course to the curriculum, or by replacing a current elective with an analytics requirement. Third, the accounting students should have broad softer skills such as leadership ability, business acumen, global acumen, and social skills (PwC, 2015).

Public Company Accounting Oversight Board (PCAOB) and AICPA regulators should clarify the degree that they support the use of more advanced analytics.13 The issue of CPA exam requirements and CPA certification requirements should be thoroughly discussed as they relate to accounting department offerings and requirements. Answers are needed for how these CPA requirements could be changed to encourage the study of analytics and Big Data by potential CPAs.

VII. CONCLUSION

This paper describes the business demand for ADA, discusses the existing ADA tools that lead to a change for the knowledge and skills required of auditors, proposes a framework of ADA knowledge and skills that future auditors would receive, points out the current challenges for the CPA exam, professional regulators, as well as the education, and finally provides suggestions. As one Big 4 senior partner stated, the modification of audit methodologies currently underway regarding the incorporation of ADA is the “largest sea change in the audit process since the passage of the ’33 and ’34 Securities Exchange Acts (KPMG, 2015).” These efforts should be a collaborative one between the regulators, the industry, and the educators as all three parties have an integral role to play in realizing the adoption of ADA in the modern audit engagement. This paper identifies the skills and knowledge necessary to prepare the next generation of auditors and identifies opportunities to modify the curriculum. Ellen Glazerman14 executive director of the PCAOB announced adding “Changes in the Use of Data and Technology in the Conduct of Audits” to its research agenda (PCAOB, Standard-setting Update, September 30, 2017).

13 PCAOB announced adding “Changes in the Use of Data and Technology in the Conduct of Audits” to its research agenda (PCAOB, Standard-setting Update, September 30, 2017).
14 Extracted from presentation at the AAA National Meeting, 2016.
EY Foundation states: “We are preparing students for jobs that do not exist, using technologies that have not been invented, to solve problems that have not been identified.” Shouldn’t accounting education get an upgrade?
VIII. REFERENCES


Rosario, A., Three Essays on Audit Innovation: Using Social Media Information and Disruptive Technologies to Enhance Audit Quality, Ph.D. Dissertation, Rutgers Business School, 2019, Newark, NJ.


TABLES AND FIGURES

<table>
<thead>
<tr>
<th>Tools</th>
<th>Examples</th>
<th>Knowledge/skills NOT required</th>
<th>Prerequisite Knowledge/skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data analytical tools</td>
<td>SAS, SPSS, R, Matlab, Stata, Python, H2O, Natural language understanding</td>
<td>extensive understanding of machine learning algorithms, deep understanding of statistical methodologies, complex programing, sophisticated linguistic theories</td>
<td>basic understanding of machine learning algorithms, Basic understanding of blockchain characteristics and utilization, pros and cons of a single algorithm, simple programming, preparing, preprocessing and analyzing data, interpretation of analysis results, generating data analytics reports</td>
</tr>
<tr>
<td>Data visualization tools</td>
<td>Tableau</td>
<td>manually create tables, flow charts, figures, and graphs</td>
<td></td>
</tr>
<tr>
<td>CAATs</td>
<td>ACL, IDEA, Arbutus</td>
<td>Sampling, Duplicate Tests, Gap Tests, Stratifications, Benfords Law, Descriptive Statistics</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Existing tools and DA knowledge and skills
Figure 1: The structure of knowledge/skills required of auditors
<table>
<thead>
<tr>
<th>Skill Category</th>
<th>Knowledge Areas</th>
<th>Examples of skills/knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm’s business and the characteristics</td>
<td>Business processes and cycles</td>
<td>Transactional data from Order-to-cash process</td>
</tr>
<tr>
<td>Nature of the firm’s ERP system</td>
<td>Nature and structure</td>
<td>modules, functions, extensions, customizations, outside partners, access privileges, procedures</td>
</tr>
<tr>
<td>Basics of a database system</td>
<td>Terminology Concepts</td>
<td>relation, attributes, tuple domain, null values, row ordering, column ordering, key, foreign key administrators, designers, end users</td>
</tr>
<tr>
<td></td>
<td>Users</td>
<td>architecture, data models, data schemas, data dictionary</td>
</tr>
<tr>
<td></td>
<td>Design</td>
<td>Data warehouse, NoSQL databases, ETL activities</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Metadata of Big Data</td>
<td>data type, data content, data processing</td>
<td>understanding of data fields or call detail records data collection, transformation, and structuring</td>
</tr>
</tbody>
</table>

Table 2: Auditor DA knowledge and skills - a comprehensive understanding of data
<table>
<thead>
<tr>
<th>Skill category</th>
<th>Knowledge area</th>
<th>Examples of skills/knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics</td>
<td>Variables</td>
<td>Nominal variables, Ordinal variables, Interval variables</td>
</tr>
<tr>
<td></td>
<td>Measure of frequency</td>
<td>Frequency distribution</td>
</tr>
<tr>
<td></td>
<td>Measures of central tendency</td>
<td>Mean, Median, Mode</td>
</tr>
<tr>
<td></td>
<td>Measures of dispersion</td>
<td>Variance, Standard Deviation, Coefficient of Variance, Range, Percentiles, Quartiles</td>
</tr>
<tr>
<td></td>
<td>Measures of symmetry</td>
<td>Skewness, Kurtosis</td>
</tr>
<tr>
<td></td>
<td>Measures of whether the data are heavy-tailed or light-tailed relative to a normal distribution</td>
<td></td>
</tr>
<tr>
<td>Statistical inference (including Bayesian Inference)</td>
<td>Experimental design</td>
<td>Statistical significance, Null hypothesis, Alternative hypothesis</td>
</tr>
<tr>
<td></td>
<td>Types of tests</td>
<td>Correlation, one-tail hypothesis test, Chi-square tests, analysis of variance (ANOVA), analysis of covariance (ACOVA)</td>
</tr>
<tr>
<td></td>
<td>Results</td>
<td>Test statistic, rejection region, degree of freedom, significance level, confidence level, confidence interval, Type I and II errors</td>
</tr>
</tbody>
</table>

Table 3: Auditor DA knowledge and skills- a solid foundation of statistics
<table>
<thead>
<tr>
<th>Skill/Technique Category</th>
<th>Knowledge Areas</th>
<th>Examples</th>
<th>Required levels of Skill Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Partner</td>
</tr>
<tr>
<td>Visualization</td>
<td>Application and interpretation of dynamic and interactive business graphics</td>
<td>Heat Maps, Real-Time Dashboards and Charts</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification/regression</td>
<td>Key concepts</td>
<td>Root Nodes, Leaf Nodes, Tree Depth, Branches, Decision Rules, Entropy</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Mainstream algorithms</td>
<td>Decision Tree: C&amp;RT, CHAID</td>
<td></td>
</tr>
<tr>
<td>Association</td>
<td>Key concepts</td>
<td>Itemset, Support Count, Support, Frequent Itemset, Confidence, Association Rules</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Major algorithms</td>
<td>A priori Algorithm</td>
<td></td>
</tr>
<tr>
<td>Text mining</td>
<td>Document collection</td>
<td>Static and Dynamic Collection</td>
<td>High level for specialized team</td>
</tr>
<tr>
<td></td>
<td>Document representation</td>
<td>Feature Based Representation and Relational Representation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Document features</td>
<td>Characters, Words, Terms, Concepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic text mining tasks</td>
<td>Document Classification Information Retrieval Organization of Documents Information Extraction</td>
<td></td>
</tr>
<tr>
<td>Process mining</td>
<td>Basic concepts, major analysis procedures</td>
<td>Process Discovery, Conformance Check, Performance Analysis, Decision Mining and</td>
<td>High level for membe rs in</td>
</tr>
<tr>
<td>Field</td>
<td>Knowledge and Skills</td>
<td>Specialized Team</td>
<td>High Level for Members in Specialized Team</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Neural networks / Deep learning</td>
<td>Basic concepts and structures, types of deep neural networks, hyperparameters training algorithms</td>
<td>Specialized team</td>
<td>High level for members in specialized team</td>
</tr>
<tr>
<td>Computer-Assisted Audit Techniques (CAATs)</td>
<td>Select proper AD tools effectively use the tools, evaluate tools, combine multiple tools for different audit procedures</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Blockchain and smart contracts</td>
<td>Understand applicability, areas of exposure, proper valuation of cyber-assets</td>
<td>High level for specialized team</td>
<td>Basic</td>
</tr>
<tr>
<td>Robotic process automation</td>
<td>Understand the features, conditionings, and approached or RPA</td>
<td>Basic</td>
<td>Basic</td>
</tr>
</tbody>
</table>

Table 4: Auditor DA knowledge and skills – being proficient in advanced DA techniques
Figure 2: This decision tree illustrates the step-wise conditional logic that is undertaken when analyzing employee transactions to identify legitimate purpose/non-legitimate purpose of purchases (adopted from Al-Awadhi et al, 2017).
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Define audit analytics objectives</td>
</tr>
<tr>
<td>2.</td>
<td>Determine relevant analytics functions for the objectives</td>
</tr>
<tr>
<td>3.</td>
<td>Determine the “exception”</td>
</tr>
<tr>
<td>4.</td>
<td>Identify relevant IT systems and assess availability and quality of data</td>
</tr>
<tr>
<td>5.</td>
<td>Acquire data (i.e., extract, transform, load process)</td>
</tr>
<tr>
<td>6.</td>
<td>Develop analytics (i.e., script, program)</td>
</tr>
<tr>
<td>7.</td>
<td>Run analytics and perform initial validation of results to identify data and/or logic flaws</td>
</tr>
<tr>
<td>8.</td>
<td>Confirm the results of the analytics support achieving the audit objective and revise, abandon or rerun analytics</td>
</tr>
<tr>
<td>9.</td>
<td>Validate results of analytics with business owners</td>
</tr>
<tr>
<td>10.</td>
<td>Research, follow up, and determine root cause of identified exceptions</td>
</tr>
<tr>
<td>11.</td>
<td>Report findings and recommendations to business owners and management</td>
</tr>
<tr>
<td>12.</td>
<td>Update analytics repository and enhance repeatability, as appropriate</td>
</tr>
</tbody>
</table>

Figure 3: Step-wise listing of the data analysis task, based on a report by KPMG (2012)