A QUALITATIVE STUDY ON PREDICTIVE MODELS IN ACCOUNTING FRAUD DETECTION

by

Anthony Cecil

Dissertation
Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Business Administration

Liberty University, School of Business
October 2021
Abstract

Companies lose an estimated 5% of revenue each year due to occupational fraud. This level of fraud can significantly disrupt the capital markets and cause companies to go bankrupt.

Unless organizations, the government, and the accounting profession develop a systematic approach for accounting fraud detection, investors and employees will continue to lose money.

This study explored subject matter experts’ perceptions of building and deploying artificial intelligence and predictive models to detect accounting fraud. This case study consisted of interviews with 10 participants with expertise in predictive modeling, auditing, and investigating, as well as a systematic literature review of research and technical documentation relating to artificial intelligence, predictive modeling, and accounting fraud detection. Six themes materialized from this research, including data, data mining techniques, model input and output, human agency, approach, and explainable artificial intelligence. This study attempted to expand the scope of prior research to identify the best machine learning algorithms to detect accounting fraud consistently and accurately. The results of this research revealed that labeled data is much more important to building accurate models than any machine learning algorithm or any other aspect of the model building process. However, quality labeled data is in short supply. To overcome this challenge, the research suggests using weak supervision and active learning to generate labeled data. Overall, the research shows that predictive models powered by machine learning algorithms using financial, nonfinancial, and textual data provide auditors and fraud investigators with the best chance of detecting accounting fraud.

Keywords: accounting fraud detection, data analysis, predictive models, machine learning, and artificial intelligence.
Dedication

I dedicate this dissertation to my family. To my wife, Jennifer, thank you for the walks and the opportunity to talk about my research, which always helped me to grasp the complex concepts better. Also, thank you for always being there for the kids when I could not. To my children, Nick and Olivia, thank you for understanding the significant commitment needed to complete a doctorate program and for going to your mom first for everything. Finally, I dedicate this dissertation to my mom, Dot, who always encouraged me to learn, grow, and explore, and showed me what it looks like to be curious about the world.
Acknowledgements

I want to thank Dr. Felicia Olagbemi, my dissertation chair, who always encouraged me and was always excited to talk about my dissertation. Each of our conversations provided the inspiration I needed to move forward. You handled each setback I encountered with calmness and gave me the confidence to see my dissertation through to the end. To Dr. Jamie Stowe, my committee member, thank you for all your help with my dissertation. Your insightful feedback was always helpful and appreciated. Finally, to Dr. Edward Moore, the DBA Director, thank you for your guidance as I worked my way through the doctoral program.
Table of Contents

Abstract ........................................................................................................................................... ii

Dedication ...................................................................................................................................... iii

Acknowledgements ........................................................................................................................ iv

List of Tables ................................................................................................................................... ix

List of Figures .................................................................................................................................. x

Section 1: Foundation of the Study................................................................................................. 1

  Background of the Problem ................................................................................................ 2
  Problem Statement .............................................................................................................. 3
  Purpose Statement ............................................................................................................... 4
  Nature of the Study ............................................................................................................. 5
    Discussion of Method ............................................................................................. 5
    Discussion of Design .............................................................................................. 5
    Summary of the Nature of the Study ...................................................................... 6

Research Questions ..................................................................................................................... 7

Conceptual Framework ............................................................................................................... 7

  Discussion of Diverse Data .................................................................................. 8
  Discussion of Data Analyses .................................................................................. 9
  Discussion of Fraud Warning Signals ................................................................ 9
  Discussion of Accounting Fraud Detection ....................................................... 9
  Discussion of Professional Judgment and Skepticism ....................................... 10
  Discussion of Relationships Between Concepts ............................................. 10
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of the Conceptual Framework</td>
<td>11</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>11</td>
</tr>
<tr>
<td>Assumptions, Limitations, Delimitations</td>
<td>12</td>
</tr>
<tr>
<td>Assumptions</td>
<td>12</td>
</tr>
<tr>
<td>Limitations</td>
<td>12</td>
</tr>
<tr>
<td>Overall</td>
<td>14</td>
</tr>
<tr>
<td>Delimitations</td>
<td>14</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Reduction of Gaps</td>
<td>15</td>
</tr>
<tr>
<td>Implications for Biblical Integration</td>
<td>15</td>
</tr>
<tr>
<td>Relationship to Field of Study</td>
<td>16</td>
</tr>
<tr>
<td>Summary of the Significance of the Study</td>
<td>17</td>
</tr>
<tr>
<td>A Review of the Professional and Academic Literature</td>
<td>17</td>
</tr>
<tr>
<td>The Evolution of Business Structures and the Need for Auditing</td>
<td>18</td>
</tr>
<tr>
<td>Accounting Fraud</td>
<td>20</td>
</tr>
<tr>
<td>Revenue Recognition Standards</td>
<td>21</td>
</tr>
<tr>
<td>Regulations and Auditing Standards</td>
<td>26</td>
</tr>
<tr>
<td>Considering Fraud in Financial Statements</td>
<td>28</td>
</tr>
<tr>
<td>Data Analytics</td>
<td>33</td>
</tr>
<tr>
<td>Advanced Data Analytics</td>
<td>36</td>
</tr>
<tr>
<td>Predictive Models</td>
<td>38</td>
</tr>
<tr>
<td>Potential Themes and Perceptions</td>
<td>50</td>
</tr>
<tr>
<td>Summary of the Literature Review</td>
<td>51</td>
</tr>
</tbody>
</table>
Transition and Summary of Section 1 ................................................................. 52

Section 2: The Project ......................................................................................... 52

Purpose Statement .............................................................................................. 52

Role of the Researcher ......................................................................................... 54

Participants ........................................................................................................... 55

Research Method and Design ............................................................................ 56

Population and Sampling .................................................................................... 58

Data Collection ..................................................................................................... 61

    Academic ......................................................................................................... 62

    Accounting, Auditing, and Investigative Professional Organizations ............ 64

    Audit Firms ..................................................................................................... 65

    Technology Firms .......................................................................................... 65

Data Analysis ........................................................................................................ 65

    Content Evaluation ......................................................................................... 65

    Synthesis ......................................................................................................... 67

Reliability and Validity ......................................................................................... 67

    Triangulation ................................................................................................ 68

    Member Checking ......................................................................................... 68

    Disconfirming Evidence ................................................................................ 69

    Reflexivity ...................................................................................................... 69

Transition and Summary of Section 2 ............................................................... 69

Section 3: Application to Professional Practice and Implications for Change ........ 70

Overview of the Study ........................................................................................ 70
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Demographics</td>
<td>73</td>
</tr>
<tr>
<td>Systematic Review Analysis</td>
<td>76</td>
</tr>
<tr>
<td>Academia</td>
<td>76</td>
</tr>
<tr>
<td>Professional Associations</td>
<td>78</td>
</tr>
<tr>
<td>Auditing and Investigating Firms</td>
<td>79</td>
</tr>
<tr>
<td>Data Technology Companies</td>
<td>80</td>
</tr>
<tr>
<td>Presentation of the Findings</td>
<td>80</td>
</tr>
<tr>
<td>Theme 1: Data</td>
<td>81</td>
</tr>
<tr>
<td>Theme 2: Data Mining Techniques</td>
<td>91</td>
</tr>
<tr>
<td>Theme 3: The Model</td>
<td>102</td>
</tr>
<tr>
<td>Theme 4: Human Agency</td>
<td>111</td>
</tr>
<tr>
<td>Theme 5: Approach</td>
<td>116</td>
</tr>
<tr>
<td>Theme 6: Explainable AI</td>
<td>125</td>
</tr>
<tr>
<td>Summary of the Findings</td>
<td>128</td>
</tr>
<tr>
<td>Application to Professional Practice</td>
<td>135</td>
</tr>
<tr>
<td>Recommendations for Action</td>
<td>139</td>
</tr>
<tr>
<td>Recommendations for Further Study</td>
<td>144</td>
</tr>
<tr>
<td>Reflections</td>
<td>147</td>
</tr>
<tr>
<td>Biblical Perspective</td>
<td>149</td>
</tr>
<tr>
<td>Summary and Study Conclusions</td>
<td>151</td>
</tr>
<tr>
<td>References</td>
<td>154</td>
</tr>
<tr>
<td>Appendix A: Participant Interview Questions</td>
<td>173</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Accuracy Measures for Supervised Learning Classification Algorithms.......................47
Table 2. Summary of Participant Demographics........................................................................74
Table 3. Initial Search Terms and Phrases Used for Systematic Review Analysis......................77
Table 4. Search Strings Used to Focus Systematic Review Analysis........................................78
List of Figures

Figure 1. Conceptual Framework ................................................................................................... 8
Figure 2. Research Question One and Applicable Themes ........................................................ 130
Figure 3. Research Question One and Key Data Themes ........................................................... 131
Figure 4. Research Question One and Applicable Key Technology Themes ......................... 132
Figure 5. Research Question One and Key Individual Themes ............................................... 133
Figure 6. Research Question One and Key Organization Themes ......................................... 134
Figure 7. Research Question Two and Applicable Themes .................................................... 135
Section 1: Foundation of the Study

Companies lose an estimated 5% of revenue each year due to occupational fraud, according to a 2019 survey conducted by the Association of Certified Fraud Examiners (ACFE) (ACFE, 2020). Although the United States Congress (Congress) has enacted laws over the years relating to fraud deterrence, fraud continues to destroy employee retirement savings and investments (Montesdeoca et al., 2019). Unless organizations, the government, and the accounting profession develop a systematic approach for accounting fraud detection, investors and employees will continue to lose money. Previous research on accounting fraud detection has focused on financial analysis using ratios (Gray & Debreceny, 2014). Other studies have focused on textual and nonfinancial analyses (Purda & Skillcorn, 2015). Fewer studies have focused on evaluating and providing guidance concerning the most useful data analysis techniques for accounting fraud detection.

Researchers studying accounting fraud detection have been increasingly interested in advanced data analytics, such as natural language processing (NLP) and machine learning (Ruan et al., 2019; Sadgali et al., 2019). Quite a bit of research exists regarding the various data analysis techniques. However, researchers have directed only minimal attention toward which data analysis techniques are the most useful for accounting fraud detection or how best to overcome the lack of labeled data to adequately build predictive models. Existing studies have used primarily quantitative research methods to evaluate various data analysis techniques but do not provide clear guidance on the best techniques to use (Omidi et al., 2019). The plethora of detection techniques only confuses readers, leading to inaction by practitioners. By examining predictive models through a qualitative approach, researchers can better understand and identify the most useful predictive models for accounting fraud detection. With this better understanding,
lawmakers, regulators, the accounting profession, and academia can develop laws, regulations, policies, and training to provide auditors and fraud investigators with the most effective tools and skills for accounting fraud detection.

**Background of the Problem**

For capital markets to operate effectively and efficiently, investors must trust the financial information provided by companies (Clikeman, 2020; Lins et al., 2017). Without true and accurate company financials, investors will not invest their hard-earned money in the capital markets (Ng et al., 2016). As evidenced by the Enron and WorldCom scandals in the early 2000s, accounting fraud drastically impacts companies and destroys employees’ lives by leaving them without retirement savings (Lin et al., 2015). As a result of these instances of accounting fraud and to mitigate recurrences, the United States enacted legislation such as the Sarbanes-Oxley Act of 2002, and the Auditing Standards Board (ASB) of the American Institute of Certified Public Accountants (AICPA) issued Statement on Auditing Standards (SAS) No. 99, *Consideration of Fraud in a Financial Statement Audit*, in 2002.

Even with the increased legislation, accounting fraud continues to occur and erode investor confidence (Lin et al., 2015). Legislation and corporate policies alone cannot address this problem. Societies must incorporate other measures to combat accounting fraud. The literature provides many predictive models for accounting fraud detection but does not evaluate the most useful predictive models. Further, the literature does not adequately address the lack of labeled data, which is essential to develop high-quality machine learning models. The vast number of predictive models only serves to confuse organizations concerning how best to proactively detect accounting fraud (Dutta et al., 2017; Gray & Debreceny, 2014; Salijeni et al., 2019).
Problem Statement

The general problem to be addressed is the gap between predictive models and accounting fraud detection that results in the failure to identify fraudulent behavior in a timely manner (ACFE, 2020; Gepp et al., 2018; Purda & Skillicorn, 2015). The ACFE (2020) has estimated that fraud costs organizations 5% of revenues each year, for a projected potential total fraud loss of up to $4.5 trillion worldwide. This level of fraud can significantly disrupt the capital markets and cause companies to go bankrupt if the business community does not promptly address the gap in accounting fraud detection (Gepp et al., 2018; Purda & Skillicorn, 2015). A 2020 ACFE study showed that fraud can go undetected for approximately 14 months, but by performing proactive advanced data analytics, organizations will be in a better position to detect accounting fraud earlier and reduce losses.

Further, the ACFE conducted a survey in 2019 regarding the use of anti-fraud technology by companies and found that only 30% of the companies use advanced data analytics such as predictive modeling. As more types of data and technology become available, the marketplace will be inundated with a range of predictive models. Without specific guidance on the most effective predictive models, organizations may become overwhelmed and avoid adoption of predictive models altogether or select an ineffective model. The specific problem to be addressed is the lack of practical predictive models incorporating financial, nonfinancial, and textual analyses for use by auditors and fraud investigators for timely accounting fraud detection (Gepp et al., 2018; Purda & Skillicorn, 2015; Salijeni et al., 2019).
Purpose Statement

The purpose of this qualitative case study was to explore current predictive models (financial, nonfinancial, and textual analysis techniques) that contribute to timely accounting fraud detection (Dutta et al., 2017; Gray & Debreceny, 2014; Salijeni, et al., 2019). Additionally, the study shed light on the best data and data analytic techniques to use so that auditors and fraud investigators can have the best chance to detect accounting fraud proactively. Auditors have had limited success in detecting accounting fraud using only financial analytical procedures (J. Tang & Karim, 2018). The literature has indicated that further research is needed to identify additional nonfinancial measures such as employee head counts and retail and warehouse space as well as textual information included in the management discussions from the annual reports of public companies (Gray & Debreceny, 2014).

In previous studies, researchers have used various predictive models to detect fraud (Lin et al., 2015). This study’s researcher attempted to identify the most useful predictive models for timely accounting fraud detection. The researcher interviewed subject matter experts from the auditing, accounting, technology, and academic communities. Further, the researcher evaluated previous studies to derive a predictive model that enhances accounting fraud detection for auditors and fraud investigators. This study provides organizations and the accounting profession with a foundation for using financial, nonfinancial, and textual analyses for timely accounting fraud detection (Dutta et al., 2017; Gray & Debreceny, 2014; Salijeni et al., 2019). Finally, the results of this research can be used by policymakers, the accounting profession, and corporations to develop better analytical procedures to increase the chances of timely accounting fraud detection by fraud investigators and auditors.
**Nature of the Study**

This research project used a qualitative case study research approach. This method and design provided the best opportunity to understand current predictive models used for accounting fraud detection (Purda & Skillicorn, 2015).

**Discussion of Method**

According to Stake (2010), qualitative studies are used to understand a concept instead of improve a concept. The researcher must first understand what is currently happening, especially if the concept is in the early stages of development. Since this study does not seek to quantify data, test relationships among variables, or develop new theories or hypotheses, a quantitative research design did not fit this study. A qualitative method provided the best chance to understand better the current predictive models and put forth the most practical combination of models for accounting fraud detection (Creswell, 2014; Stake, 2010).

**Discussion of Design**

Of the qualitative research designs, the case study approach allowed this researcher to evaluate existing predictive models and understand the financial, nonfinancial, and textual components used in the models. This study neither addressed individual lives lived nor attempted to develop a theory grounded solely on participant views. Therefore, the narrative, phenomenology, grounded theory, and ethnography qualitative designs did not fit this study (Creswell & Poth, 2018). On the other hand, since the study sought to develop an in-depth analysis of multiple cases (multiple predictive models), the qualitative case study approach provided the best opportunity for this researcher to view the big picture and understand the complex interactions of existing predictive models using financial, nonfinancial, and textual analyses for accounting fraud detection (Creswell, 2016).
Since this study was geared towards auditors and fraud investigators, the researcher considered the profiles of these practitioners. Therefore, the researcher evaluated existing predictive models based on the following:

- Models must not take weeks or months to execute because auditors have tight time constraints and budgets. Further, fraud investigators are under pressure to complete their investigations in a timely manner.
- Models must be able to address financial, nonfinancial, and textual data analysis.
- Models must have a high accounting-fraud detection rate.
- Model concepts and techniques must be understood by auditors and fraud investigators.
  In other words, practitioners using the models should not be required to be data scientists or statisticians.
- Models must not generate too many false positives, which require significant time to resolve.

By evaluating existing predictive models using qualitative measurements, this study’s researcher can assist in bringing accounting fraud detection theory into practice for auditors and fraud investigators.

Summary of the Nature of the Study.

In the last 10 years, the literature has shown advances in technology and available digital data to develop predictive models for accounting fraud detection. These predictive models are still in the early stages of development and need to be understood better. This qualitative case study sought to understand the current predictive models to identify the most effective ones for timely accounting fraud detection.
Research Questions

Given the volume, velocity, and variety of data generated by companies and the marketplace, internal and external auditors as well as fraud investigators and regulators are overwhelmed and do not have the right tools to adequately perform their jobs (Janvrin & Weidenmier Watson, 2017). Without dependable and consistent predictive models to detect accounting fraud, these professionals are at a distinct disadvantage. By asking some innovative questions, researchers have opened the door to using predictive models to detect accounting fraud better, but more exploration of this topic is needed to identify the most useful predictive models for timely accounting fraud detection (Gray & Debreceny, 2014). This researcher sought to gather information to better understand the current predictive models used to detect accounting fraud. This study’s researcher attempted to answer the following research questions:

RQ1. Why are current predictive models used by auditors and fraud investigators not able to detect accounting fraud consistently and accurately?

RQ2. How can financial, nonfinancial, and textual analyses be used by auditors and fraud investigators to detect accounting fraud warning signals consistently and accurately?

Conceptual Framework

This researcher explored predictive models consisting of financial, nonfinancial, and textual analyses used for accounting fraud detection. The study’s researcher attempted to determine the best combination of these analyses to provide practitioners the necessary tools for timely accounting fraud detection. As shown in Figure 1, the concepts included diverse data, data analyses, fraud warning signals, and accounting fraud detection, along with professional judgment and skepticism.
Figure 1. Conceptual Framework

Discussion of Diverse Data

Diverse data represents structured and unstructured data. Structured data is the typical data found in spreadsheets and databases, formatted using columns and rows (ACFE, 2019). In other words, the formatting has structure. Examples of structured data include accounting records, sales transactions, and purchases. Unstructured data is typically found outside of spreadsheets and databases and includes data such as text documents, emails, and photos (ACFE, 2019; Y. Chen et al., 2019). In addition, a discussion of data must include quality labeled data. To successfully build a supervised classification model, quality labeled data is required. For accounting fraud detection, two labels usually exist: fraudulent and nonfraudulent. Without data
labeled as fraudulent or nonfraudulent, the chances of building a predictive model are low (Mohammadi et al., 2020; Sarkar et al., 2020).

Discussion of Data Analyses

Data analyses uses analytics software to identify patterns, trends, anomalies, and unexpected relationships using financial, nonfinancial, and textual data (ACFE, 2019). By using data analytics, auditors and fraud investigators gain insights from financial, nonfinancial, and textual data to identify fraud warning signals (Perols et al., 2017). One of the primary techniques in the current data analysis landscape includes machine learning, which consists of supervised learning, unsupervised learning, and semi-supervised learning. With supervised machine learning, labeled data is fed into models using algorithms to learn how to classify data based on the examples provided. In this way, the models are trained to classify data as fraudulent or nonfraudulent in regards to accounting fraud (Gepp et al., 2018; Margagliotti & Bollé, 2019).

Discussion of Fraud Warning Signals

The patterns, trends, anomalies, and unexpected relationships resulting from data analyses can highlight fraud warning signals. These red flags may be indicators of accounting fraud. Unfortunately, the results of a data analyses may include items that do not relate to fraud, which represent false positives because, even though they met the initial exception criteria, the transactions are not exceptions (Hájek & Henriques, 2017).

Discussion of Accounting Fraud Detection

Auditors and fraud investigators evaluate the resulting fraud warning signals or red flags using various techniques such as interviewing employees and management or reviewing source documents. After a thorough evaluation of the red flags, the practitioner decides on the propriety of the anomalies. Once wrongdoing is confirmed, the information is presented to corporate
management or, in some cases, to the Securities and Exchange Commission (SEC) for resolution (Leite et al., 2018).

**Discussion of Professional Judgment and Skepticism**

As auditors and fraud investigators work through the process of selecting data, performing data analyses, and evaluating red flags, they must use professional judgment and skepticism. Even with predictive models, accounting fraud detection requires human judgment (Moffitt et al., 2018). Auditors and fraud investigators must use professional judgment when selecting appropriate data to test or evaluating data analyses results. Professional skepticism relates to a questioning mind, which is critical when evaluating anomalies and interpreting fraud risk factors (Earley, 2015).

**Discussion of Relationships Between Concepts**

To accurately detect accounting fraud, the concepts presented must proceed through a logical progression of general to specific information. Practitioners can use vast amounts of diverse data to detect anomalies for further analysis. Performing financial data analysis using only structured financial data may not be enough to provide the necessary fraud warning signals to detect accounting fraud. Performing textual data analysis using only unstructured data may not be enough either. The literature has indicated that predictive models using a combination of diverse data as well as a variety of data analysis techniques has a better chance of detecting accounting fraud (Sun & Vasarhelyi, 2018). Further, with respect to data analysis using machine learning, the lack of labeled data can severely hamper the ability to build quality predictive models (Mohammadi et al., 2020; Sarkar et al., 2020).
Summary of the Conceptual Framework

Overall, the concepts presented are the most critical ones to accurately identify the most useful predictive models for timely accounting fraud detection. Without a proper understanding of diverse data and the available data analytic techniques, practitioners will have a difficult time detecting accounting fraud. Finally, if practitioners do not possess the appropriate professional judgement and skepticism required for integrating data analysis with fraud detection, they do not stand a chance of uncovering accounting fraud perpetrated by professionals who are well versed in deception and hiding accounting frauds.

Definition of Terms

Accounting fraud. SAS No. 99 defined fraud as “an intentional act that results in a material misstatement in financial statements that are the subject of an audit” (AICPA, 2002, p.6). This study used this definition to describe accounting fraud.

Anomalies. An anomaly is a deviation from a business rule or from what is expected and requires further investigation (Jans & Hosseinpour, 2019). For example, if an organization recorded 100% of its revenue on the last day of the fiscal year, then this would be considered an anomaly because organizations usually record revenue throughout the year.

Data analytics. Data analytics are used to gain insights from financial, nonfinancial, and textual data. By analyzing operational, financial, and other forms of digital data, auditors and fraud investigators can understand the data better. The process of analysis separates data into various classifications to isolate and identify anomalies (Wang & Cuthbertson, 2015).

Predictive models. A predictive model is used to examine vast amounts of digital data based on specific attributes and classifiers to identify anomalies. The examination is executed using data analytic techniques such as clustering or data segmentation. The predictive power of
the model depends on the combination of attributes, data, and data analytic techniques selected (Dutta et al., 2017; Gepp et al., 2018; Kim et al., 2016; Serrano-Cinca et al., 2019).

**Assumptions, Limitations, Delimitations**

A study’s assumptions, limitations, and delimitations play a significant role in the veracity of the study. Without understanding the assumptions, limitations, and delimitations of a study, the proper controls cannot be implemented to ensure the accuracy of a research project (Creswell & Poth, 2018). The next section discusses the assumptions, limitations, and delimitations below.

**Assumptions.** This researcher assumed that participants answered all interview questions truthfully. By providing participants with assurances of confidentiality and a venue in which the participants could talk privately, participants could feel free to provide honest answers. Signed informed consent forms helped assure participants that their confidentiality would be protected (Creswell & Poth, 2018). Also, the researcher assumed that literature from national and peer review journals was true and accurate (Creswell, 2016). Finally, this study’s researcher assumed that accounting fraud warning signals remain relatively the same regardless of the perpetrator’s behavior or methods used to commit and conceal the fraud. Therefore, this researcher did not attempt to predict human behavior but rather identify the red flags resulting from the fraudulent behavior. Clikeman (2020) put forth 16 accounting fraud cases that occurred over the last 100 years. For each case, the fraud warning signals were basically the same (Clikeman, 2020).

**Limitations.** To ensure the validity of a study, the researcher must be mindful of the study's limitations. This study included the following limitations:

1. Measurements used by the researcher to accurately evaluate useful predictive models might not have been well-defined or defined well enough to identify the most useful
predictive models for auditors and fraud investigators to consistently detect accounting fraud. Further, if the researcher used preconceived notions to develop the measurements, then the results of the study might be limited.

2. The research might not be adequately understood by the researcher, which impacts the inferences made for the study. Further, the inferences made in the study might not apply to the business environment, and therefore, the study might have limited benefits to auditors and fraud investigators for the purpose of accounting fraud detection.

3. The conclusions made based on a predictive model might not be solely due to the model but other factors, including professional judgment, professional experience, technology, available data, and the complexity of the fraud detected.

4. The findings in the study might not be generalizable in such a way as to be practical in a business environment. In other words, the most useful predictive models might not be practical for auditors or fraud investigators because the models require an advanced knowledge of statistics, modeling, and technology, which auditors and fraud investigators typically do not possess.

5. The study might not identify all relevant predictive models resulting in incorrect inferences impacting the study results.

6. This study reviewed existing predictive models that use data from nonfraudulent companies and fraudulent companies. With respect to nonfraudulent companies, it should be noted that although fraud has not been identified in the past, it does not mean that a company does not have fraud. Therefore, the results from previous studies might not be accurate if fraud existed in the nonfraudulent companies.
Overall, this researcher used multiple sources of data captured from various viewpoints including academia, accounting, auditing, and technology to mitigate the limitations discussed above. Exposure to this all-encompassing information provided a better understanding of predictive modeling to this researcher, which helped to overcome these limitations.

**Delimitations.** For many centuries, fraud, in its various forms, has caused misery for societies. The list of frauds includes academic fraud, consumer fraud, and financial fraud, as well as many other types of fraud. As individuals and organizations intentionally deceive others for personal gain, societies continue to suffer the consequences of fraud. The term fraud has many meanings, so it is essential that this study sets limits on the specific fraud addressed in the study.

The ACFE has classified occupational fraud into three categories: Corruption, Asset Misappropriation, and Financial Statement Fraud (ACFE, 2020). This study limited the types of fraud addressed to financial statement fraud. For this study, the terms accounting fraud and financial statement fraud have the same meaning. The accounting scandals in the early 2000s devastated the lives of many companies, employees, and investors. This researcher was interested in understanding accounting fraud and how best to detect it effectively. Preliminary research showed that predictive models may be useful for timely accounting fraud detection. Therefore, this study’s researcher considered predictive models useful for accounting fraud detection, not detection of other frauds such as asset misappropriation or corruption.

**Significance of the Study**

Since the accounting scandals of the early 2000s, regulators have implemented new laws and regulations to deter accounting fraud (Montesdeoca et al., 2019). However, accounting fraud still exists, and there are no signs that it is slowing down (Omidi et al., 2019). The accounting profession understands the severity of the problem but has yet to develop audit procedures that
can detect accounting fraud with any consistency or accuracy (Omidi et al., 2019). Organizations have implemented exception reporting processes to help detect fraud (ACFE, 2019). However, these fraud detection steps have been unsuccessful when it comes to timely accounting fraud detection. Overall, the accounting profession and organizations require better tools to be more successful in their efforts concerning timely accounting fraud detection. The research has shown that predictive modeling using diverse data and varied data analyses may be a better answer to the accounting fraud detection problem (Y. Chen et al., 2017). This study’s researcher attempted to identify a unified predictive model for timely accounting fraud detection.

**Reduction of Gaps**

The current literature has discussed a plethora of predictive models but has not recommended which models are most useful for accounting fraud detection (Gepp et al., 2018; Jans & Hosseinpour, 2019; Omidi et al., 2019; Salijeni et al., 2019). Further, the literature has not recommended the best ways to generate or collect quality labeled data to train predictive models. Today, there exists a gap between predictive models and predictive models that consistently and accurately detect accounting fraud. The literature has discussed models that use either financial, nonfinancial, or textual analysis but not all three types in one model (Gepp et al., 2018; Jans & Hosseinpour, 2019; Omidi et al., 2019; Salijeni et al., 2019). This researcher attempted to reduce the gap and provide clarity for using the most useful predictive models for accounting fraud detection.

**Implications for Biblical Integration**

Corporate executives without Christian faith may be more susceptible to corruption and greed. They may be more willing to make ethically questionable decisions. When executives love money and material worth more than God and their neighbor, they may be willing to record
accounting entries that are not accurate or tell lies to auditors to cover up accounting fraud. The Bible provides the resources and guidance to do the right thing for the good of people and not just for the good of oneself. As discussed in the Bible, Esther was willing to perish, if needed, to save Mordecai and the Jews (Esther 4:16, ESV). People “in the palace” sometimes have difficult decisions to make that may disadvantage themselves for the good of others, but if they have faith in God, they will make the right decision.

Executives are often the people in the palace and have an opportunity to do the right thing for many people, such as employees and shareholders. For example, when executives operate with integrity, they record actual revenue even when the revenue numbers do not meet Wall Street expectations. It would be very tempting for executives to book fictitious sales to avoid decreases in stock prices and ultimately decreases in executive bonuses. For executives with integrity, the decision is not hard. Companies can go bankrupt because executives perpetrate accounting fraud for short-term gains. Executives sometimes base their decisions on self-interest and not the interest of the employees, shareholders, or community. The desire for material wealth often causes executives to make bad business decisions or, even worse, perpetuate accounting fraud.

**Relationship to Field of Study**

Fraud auditing and forensic accounting play significant roles in the business community and marketplace. During 30 years working in the accounting, auditing, and fraud investigation fields, this researcher has noted that the accounting and auditing profession has been at a distinct disadvantage when it comes to combating fraud. Fraudsters are always ahead of the auditors. Lack of audit innovation is one of the key reasons why auditors have a difficult time proactively detecting fraud. For some time now, this researcher has desired to perform research and study
ways to improve the odds of detecting fraud promptly or even preventing fraud altogether. To adequately address this business problem, it is imperative that auditors use the latest technology. Using applied research and advanced business research methods, the researcher attempted to develop the best ways to help prevent, detect, and investigate fraud.

**Summary of the Significance of the Study**

The results of this study provide a sophisticated and innovative approach to combating fraud in the business world. Further, this study adds to the accounting profession by providing policymakers, lawmakers, and researchers with information to establish more effective auditing procedures and techniques. This study provides auditors and fraud investigators with a predictive model using a uniform and consistent approach for accounting fraud detection. With a useful predictive model, auditors can approach accounting fraud detection in a consistent manner using technology, data analytics, and professional experience.

**A Review of the Professional and Academic Literature**

Throughout the centuries, accounting fraud has plagued society (Skalak et al., 2011). Although governments, organizations, and the accounting profession have endeavored to eliminate accounting fraud, it still thrives today. Laws, regulations, and accounting policies and practices have not curbed the spread of accounting fraud among organizations (DeZoort & Harrison, 2018). Without the perception of detection, corporate executives with bad intentions continue to manipulate accounting records for personal gain (Mangala & Kumari, 2017). However, with technology, statistics, mathematics, and professional expertise and judgment, the accounting profession can improve detection techniques to combat accounting fraud. The literature has shown that predictive models are the right course of action for timely accounting fraud detection (Omidi et al., 2019). Unfortunately, the current literature does not provide all the
answers or lay out a comprehensive plan for developing and executing the best predictive models for detecting accounting fraud in a timely manner. Research into this subject needs to continue in order to close the gap between accounting fraud perpetration and detection.

**The Evolution of Business Structures and the Need for Auditing**

Given the accounting scandals in the early 2000s involving Enron and WorldCom (Lin et al., 2015) and the collapse of Lehman Brothers in 2008 (Bănărescu, 2015), a casual observer may think that accounting fraud is a new phenomenon. However, as Skalak et al. (2011) pointed out, accounting fraud has existed for centuries. Ancient civilizations in 4000 B.C.E. began using recordkeeping to help organize government and business records. Without proper controls, trusted employees and government agents realized they could steal money by manipulating business records to conceal the stolen funds. Over time, governments instituted audits to account for the funds provided to public or private officials. After the 1600s, individuals were not the only ones audited; organizations began to be audited as well (Skalak et al., 2011).

Clikeman (2020) posited that the public accounting industry began, in large part, due to the English Companies Act of 1845, which required corporations to submit to annual audits. Corporations hired outside accounting firms to help perform the annual audits. Beyond conducting audits, public accounting firms also assisted organizations with fraud investigations. In the 1850s, Deloitte & Touche, a public accounting firm, helped untangle the accounting frauds perpetrated at the Great Northern Railway and the Great Eastern Steamship Company in England (Clikeman, 2020).

In the 18th century, prior to the Industrial Revolution, relatively unsophisticated sole proprietorships were the most typical form of business (Bryer, 2013). Operations were self-financed, and owners were usually part of the workforce. These types of operations included
planters, blacksmiths, and attorneys, who worked for themselves and did not use outside capital investments. Bryer (2013) and Richard (2015) explained that as business enterprises transitioned from sole proprietorships to a more sophisticated form of business such as a joint venture in the nineteenth century, the accounting profession in England and the United States was forced to change. Financing to fund businesses added to the complexity of the business form. As England and the United States moved towards the Industrial Revolution, business operations became more complex and wide-ranging; significant capital was required to fund the operations. Businesses formed joint ventures to fund shipping companies as well as railroads, which required a significant amount of money obtained through capital markets and debt (Bryer, 2013; Richard, 2015).

Richard (2015) and Dobija (2017) indicated that business organizations such as joint ventures paid dividends to shareholders from profits. With more complicated operations, companies required better methods to account for business transactions. Worker-manager businesses, such as sole proprietorships, required only simple bookkeeping to account for transactions and did not need to use complex accounting methods. Owners could determine profit by measuring the increase in total assets. However, the single-entry accounting method would not satisfy investors who provided resources to joint ventures. Because the investors were not involved in managing the business operations, they did not have access to the financial records. Early in the nineteenth century, corporate financial statements did not readily exist. Without better accounting standards, corporations could not provide accurate and reliable financial reports, which made investors more vulnerable to the misuse of their investments. The capitalist system required double-entry accounting to adequately account for business transactions (Dobija, 2017; Richard, 2015).
To determine whether businesses were properly safeguarding investments and paying dividends from profits, investors needed better financial reporting. Ultimately, the emergence and growth of business forms such as corporations and joint ventures stimulated the development of better accounting standards and procedures, which allowed investors the opportunity to make informed investment decisions (Dobija, 2017; Richard, 2015). Schroeder et al. (2017) posited that with the improved accounting standards, businesses needed trained accountants to execute the new accounting standards as well as audit the financial statements.

**Accounting Fraud**

As companies grew, so did the risk of fraud. The ACFE separates occupational fraud into three categories: asset misappropriation, corruption, and financial statement fraud (ACFE, 2020). The ACFE (2020) found that financial statement fraud was the least common of the occupational frauds but resulted in the most significant loss to companies, with approximately $954,000 per occurrence. Clikeman (2020) identified 16 financial frauds that occurred in U.S. companies over the last 100 years that had the most significant impact on the accounting profession:

1. Kreuger & Toll
2. McKesson & Robbins
3. National Student Marketing Corporation
4. Equity Funding Corporation of America
5. ESM Government Securities, Inc.
6. Lincoln Savings & Loan
7. ZZZZ Best
8. Crazy Eddie, Inc.
9. Waste Management
10. Sunbeam Corporation
11. Enron
12. WorldCom
13. Taylor, Bean & Whitaker
14. Lehman Brothers
15. Parmalat
16. Olympus Corporation

Clikeman (2020) found that these financial frauds forced the accounting profession to rethink its process for conducting audits and its responsibility to detect accounting fraud. Investors believed that auditors were responsible for detecting accounting fraud. However, with each fraud occurrence, investors and regulators realized that there was a gap between what users of financial statements expected and what auditors believed was their responsibility for detecting accounting fraud. As accounting fraud increased, so did the public expectation for auditors to detect accounting fraud. Eventually, with the issuance of SAS No. 99 in 2002, the accounting profession assumed more responsibility for detecting fraud in financial statements (Clikeman, 2020).

**Revenue Recognition Standards**

Within accounting fraud, Throckmorton et al. (2015) found that revenue fraud occurred most often. Further, Hájek and Henriques (2017) found that financial fraud usually involved manipulating the recording of revenue by recognizing fictitious sales or prematurely recognizing revenue. Therefore, it is essential to understand the evolution of the revenue recognition standards given the high fraud risk relating to revenue recognition.
Companies sometimes find it challenging to determine when to record revenue and how much to record because of the complexities of the revenue reporting standards (Hepp, 2018). The complexity of revenue accounting may distract auditors from detecting fictitious revenue or revenue prematurely recorded (Appelbaum et al., 2017). Standard setters have issued hundreds of standards to clear up the confusion surrounding revenue recognition. Unfortunately, the vast number of standards has confused the issue more. Also, many of the standards are inconsistent with each other. This confusion makes it more difficult for auditors and fraud investigators to detect real anomalies relating to revenue recognition (Schroeder et al., 2017).

To make matters worse, Mazza and Azzali (2015) found that the most severe deficiency of internal controls relates to revenue recognition. In other words, in addition to the confusing revenue recognition standards, companies do not always have strong internal controls to ensure properly recorded revenue. This situation creates an opportunity for accounting fraud and its concealment, making it critical for auditors to have the appropriate tools to detect accounting fraud in a timely manner.

Over the last 30 years, standard setters have issued numerous significant revenue recognition standards. In December 1984, the Financial Accounting Standards Board (FASB) issued Statement of Financial Accounting Concepts No. 5 (SFAC No. 5), Recognition and Measurement in Financial Statements of Business Enterprises. This statement laid out two essential conditions required before companies could recognize revenue. To be recognized, the revenue had to be realized or realizable and earned (FASB, 1984).

Further, with the popularity of the internet in the 1990s, technological companies began popping up across the United States. Most of the companies did not have real businesses and recorded revenue that mostly did not exist. This growing issue concerned the accounting
profession as well as regulators. In October 1997, the AICPA issued Statement of Position 97-2 (SOP 97-2), *Software Revenue Recognition*. This statement guided companies on when to recognize revenue for services and goods relating to computer software, such as licensing, selling, and marketing (AICPA, 1997).

In the 1990s, the SEC noted a significant number of revenue recognition issues. As a result, the SEC issued SEC Staff Accounting Bulletin No. 101 (SAB 101), *Revenue Recognition in Financial Statements*, in December 1999. The bulletin again addressed revenue recognition and mentioned the conditions to be met before revenue could be recognized, which had been provided earlier in SFAC No. 5. The SEC staff added that before revenue was realized or realizable and earned, the following four conditions had to be met:

- Persuasive evidence of an arrangement must exist,
- Delivery must have occurred or services must have been rendered,
- The seller’s price to the buyer must be fixed or determinable, and
- Collectability must be reasonably assured (SEC, 1999).

Even with the statements and bulletins from the AICPA and SEC staff, revenue recognition issues continued into 2000. The FASB identified potential issues companies might experience when recording revenue for multiple products over multiple periods. To address this issue, the FASB issued Emerging Issues Task Force No. 00-21 (EITF 00-21), *Revenue Arrangements with Multiple Deliverables*. This issue attempted to address how companies were to determine the unit of accounting relating to arrangements with multiple deliverables (FASB, 2000).

After the passage of the Sarbanes-Oxley Act of 2002, the government pushed the FASB to codify the accounting standards. By the 2000s, the vast number of accounting standards from
several standard setters had overwhelmed the accounting profession and business community. In 2009, the FASB codified the accounting standards. The FASB made the Accounting Standards Codification (ASC) the primary source of generally accepted accounting principles, effective July 1, 2009, for interim and annual periods ending after September 15, 2009 (Schroeder et al., 2017). The FASB believed that the FASB ASC would reduce the amount of time required to research accounting issues as well as limit the potential for noncompliance because of the difficulty in identifying all applicable standards (Schroeder et al., 2017). The FASB ASC incorporated previously issued accounting standards, including revenue standards such as SOP 97-02, SAB No. 101, and EITF 00-21.

The FASB incorporated SOP 97-02 into the FASB ASC under Industry Topic 985, “Software,” Subtopic 605, “Revenue Recognition,” Section 25, “Recognition.” This section of the FASB ASC addressed three basic principles to determine revenue recognition relating to software, which include (1) software requiring significant production, modification, or customization; (2) software not requiring significant production, modification, or customization; and (3) multiple element arrangements (FASB, 2009b). The FASB incorporated SAB No. 101 into the FASB ASC under Topic 605, “Revenue Recognition,” Subtopic 10, “Overall,” Section 25, “Recognition.” This section states that revenue recognition involves two factors relating to revenue: (1) realized and realizable and (2) earned. These two factors remain consistent with SFAC No. 5 (FASB, 2009a). The FASB incorporated EITF 00-21 into the FASB ASC under Topic 605, “Revenue Recognition,” Subtopic 25, “Multiple-Element Arrangements,” and Section 25, “Recognition.” This section addresses multiple deliverables by determining units of accounting and measurement and allocation of the arrangement consideration (FASB, 2009a).
To adequately maintain the FASB ASC, the FASB established a system to update the FASB ASC by issuing the FASB Accounting Standard Updates (ASU) (Schroeder et al., 2017). In October 2009, the FASB issued FASB ASU No. 2009-13, *Revenue Recognition*, an amendment to FASB ASC, Topic 605. This update continued to address the accounting relating to arrangements with multiple deliverables. The update guided companies on how to separate arrangements with multiple deliverables into discernable accounting units using vendor-specific objective evidence or third-party evidence (FASB, 2009a). The FASB continued to address revenue recognition issues by issuing FASB ASU No. 2014-09, *Revenue from Contracts with Customers*, in May 2014. The FASB ASU No. 2014-09 created a new topic, Topic 606. This update intended to revise the revenue recognition standards as well as address differences between the US Generally Accepted Accounting Principles (GAAP) and the International Financial Reporting Standards (IFRS).

The FASB developed five steps to recognizing revenue, the core principle of the update, as follows:

1. Identify the contract(s) with a customer.
2. Identify the performance obligations in the contract.
3. Determine the transaction price.
4. Allocate the transaction price to the performance obligations in the contract.
5. Recognize revenue when (or as) the entity satisfies a performance obligation (FASB, 2014).

Also, with the issuance of FASB ASU No. 2014-09, this update supersedes the revenue recognition requirements in Topic 605, *Revenue Recognition*. Further, in April 2016, the FASB issued FASB ASU No. 2016-10, *Revenue from Contracts with Customers – Identifying
Performance Obligations and Licensing. After the issuance of Topic 606, the FASB and the International Accounting Standards Board (IASB) formed the Joint Transition Resource Group to monitor the implementation of Topic 606. Implementation issues caused the FASB to update Topic 606 by revising the standards to help reduce the cost and complexity of implementing Topic 606 (FASB, 2016). The FASB deferred the effective date of the new revenue recognition standard by one year, beginning after December 15, 2017, for public organizations and December 15, 2018, for nonpublic organizations (FASB, n.d.).

Even with the codification of accounting standards, the complexity of revenue recognition still exists, and thus, one can understand why it may be difficult for auditors and fraud investigators to detect accounting fraud, especially when it involves revenue. Overall, even if auditors use advanced technology to identify revenue anomalies, the gray areas within the revenue recognition standards likely cause a high number of false positives because of the difficulty to account for all of the variables involved (Baader & Krcmar, 2018).

Regulations and Auditing Standards

In the 1930s, Congress, exasperated by the accounting scandals of the time, created an agency for securities oversight and rulemaking to protect investors against fraudulent companies (Clikeman, 2020). Congress enacted the Securities Act of 1933 (1933 Act), which regulated the initial offering of stocks and bonds to the public. Further, Congress enacted the Securities Exchange Act of 1934 (1934 Act), which established the SEC to administer and enforce securities laws. The 1934 Act also required public companies to be audited annually (Bird & Park, 2018; Klein, 2017). Not only did the 1933 and 1934 Acts impact public companies, but they also significantly impacted public accounting firms. These laws held accounting firms accountable for negligence, at least more so than previous laws had held them accountable. With
the new legislation, investors could sue auditors and only had to prove that the investor had suffered damages because of a material misstatement in the financial statements (Clikeman, 2020). Thus, the 1933 and 1934 Acts made it easier for third parties such as investors and creditors to sue and recover damages resulting from auditor negligence (Clikeman, 2020). With the new laws, public auditing firms were more susceptible to liability if they failed to detect accounting fraud. In other words, the SEC and the public had a higher expectation for auditors to detect accounting fraud (Nieschweitz et al., 2000).

In 1988, the ASB issued SAS No. 53, *The Auditor’s Responsibility to Detect and Report Errors and Irregularities*. With this auditing standard, the accounting profession attempted to close the “expectation gap” by taking on more responsibility for identifying irregularities in the financial statements. SAS No. 53 required auditors to exercise professional skepticism and to report material irregularities to senior management, the board of directors, or the board's audit committee (Clikeman, 2020). However, the accounting profession still did not believe it was responsible for detecting accounting fraud. This changed in 1997, when SAS No. 53 was replaced with SAS No. 82, *Consideration of Fraud in a Financial Statement Audit*. With SAS No. 82, the accounting profession appeared to be willing to assume more responsibility for detecting accounting fraud. SAS No. 82 provided detailed guidance for auditors to better assess the risk of fraud (Lin et al., 2015).

In 2002, the ASB replaced SAS No. 82 with SAS No. 99, primarily due to the Enron and WorldCom accounting scandals. SAS No. 99 required auditors to brainstorm regarding the potential for fraud. The brainstorming session would take place at the beginning of an audit during the audit planning stage. Also, auditors were required to interview senior management and other management and nonmanagement employees regarding the existence of fraud.
Auditors were also required to focus on other areas, including revenue recognition and management override of accounting controls (J. Tang & Karim, 2018).

In early 2000, over 80 years after Congress passed the 1933 and 1934 securities legislation, Congress acted again on behalf of investors because of accounting frauds. This time, the frauds involved Enron, Tyco, and WorldCom. Even though the accounting profession issued an auditing standard to assume more responsibility for detecting accounting fraud after the early-2000s frauds, it was too little too late. Congress and the public had had enough. Congress stepped in and began to regulate the accounting profession by passing the Sarbanes-Oxley Act of 2002 (SOX). Like in the 1930s, Congress again created an agency to protect investors from fraudulent companies. SOX established the Public Company Accounting Oversight Board (PCAOB) (Clikeman, 2020).

Gunz and Thorne (2018) posited that the accounting scandals in early 2000 caused a crisis in the accounting profession. Before these scandals, the accounting profession was a self-regulating body. However, with the new legislation, the accounting profession was no longer allowed to self-regulate. The PCAOB began to regulate the accounting profession by setting ethical standards and inspecting accounting firms to ensure that the firms were operating ethically (Gunz & Thorne, 2018).

**Considering Fraud in Financial Statements**

The ASB issued SAS No. 99 in 2002 to provide auditors with a roadmap to help detect fraud in financial statements, including the overstatement of revenue. SAS No. 99, which superseded SAS No. 82, set forth crucial tools and techniques for auditors to use as they perform financial statement audits, including brainstorming, professional skepticism and judgment, analytical procedures, and the fraud triangle theory (J. Tang & Karim, 2018).
Unlike the predecessors SAS No. 53 and SAS No. 82, SAS No. 99 requires auditors to perform brainstorming as a team to consider what fraud would look like within the financial statements. For example, how could company management perpetrate fraud, how could management conceal the fraud, and which financial statement accounts could the fraud impact (AICPA, 2002)? DeZoort and Harrison (2018) posited that auditors would have a better chance of detecting financial statement fraud if they knew what types of frauds could occur given the client's business operations. J. Tang and Karim (2018) found that brainstorming performed at the beginning of an audit offered an opportunity to generate new fraud detection ideas among the audit team, allowing the auditors a better chance of detecting accounting fraud.

Accounting firms face significant pressure from the public and regulators to detect accounting fraud. Failing to detect accounting fraud that materially misstates a client’s financial statements could potentially cause the accounting firm to go bankrupt, which ultimately happened to Arthur Andersen (Petherbridge & Messier, 2016). Accounting firms can reduce the risk of failure by incorporating data analytics into their audits.

To use data analytics successfully, Jans and Hosseinpour (2019) argued that auditors must do more than execute the commands to run the data analytics software. The auditors must understand the business and what anomalies look like, given the client's business operations. Because data analytics requires an understanding of statistics and complex algorithms, businesses may believe that they need data scientists to perform the audits, but that would be a mistake, according to Richins et al. (2017). Given the professional experience and education of auditors, they have an advantage over data scientists. Auditors have a better understanding of the business world, which significantly enhances their ability to test transactions and use a problem-solving approach compared to data scientists (Richins et al., 2017). However, Moll and
Yigitbasioglu (2019) found that while auditors have the necessary skills to use computer-assisted auditing tools, they do not possess the more advanced data analytics skills needed to analyze nonfinancial data.

Earley (2015) indicated that as the accounting profession relies more on data analytics, auditors will need to use more professional judgment as they evaluate the results of the analytics. Without proper judgment, auditors may become overwhelmed with analyzing and evaluating the vast number of potential anomalies. Further, Lin et al. (2015) agreed that auditors need to make judgments based on their work experience and professional knowledge, especially given the complexity of the accounting standards and technology used by businesses.

Beyond auditors' professional judgment, Nolder and Kadous (2018) found that auditors must possess the ability to exercise professional skepticism to properly perform audits. By using a questioning mind, auditors have a better chance of recognizing anomalies, leading to the detection of accounting fraud. Audit deficiencies often occur because auditors do not use professional skepticism. Therefore, audit quality requires a high level of professional skepticism by auditors (Nolder & Kadous, 2018). J. Tang and Karim (2018) found that auditors did not consistently detect accounting fraud because fraud detection relies on auditors’ judgment. In other words, auditors do not have the same ability to judge fraud risks or factors indicating fraud. Auditors with a high level of professional skepticism will be more likely to ask additional questions or perform additional testing leading to the detection of fraud while auditors with a lower level of professional skepticism may not ask additional questions or perform additional testing, overlooking potential fraud (J. Tang & Karim, 2018).

In the 1970s, the accounting profession realized that it needed to do a better job of analyzing information to improve audit quality. Data analysis consists of disaggregating
information into manageable segments to better understand the data collected in order to make meaningful decisions and conclusions (Ramlukan, 2015). Soileau et al. (2015) posited that fraud cases in the 1970s, including the Equity Funding scandal in 1973, increased the focus on auditor analytics and led to the creation of the National Commission on Fraudulent Financial Reporting (NCFFR). The Equity Funding scandal highlighted the importance of auditors more competently performing analytical procedures than they had been. In 1978, the ASB issued SAS No. 23, *Analytical Review Procedures*, and in 1988, the ASB issued SAS No. 56, *Analytical Procedures*. By performing analytical procedures, auditors can quickly identify areas to audit, which can help them be more efficient with their time and concentrate on high-risk areas (Soileau et al., 2015).

As part of the analytical procedures, SAS No. 99 recommends using ratio analysis (AICPA, 2002). Omidi et al. (2019) indicated the benefit of using financial ratios relating to receivables, bad debt, and inventory. Also, Appelbaum et al. (2018) found that ratios and trend analyses help auditors develop expectations relating to account balances. By comparing account balances trending in the wrong direction to expected account balance trends, auditors can more easily identify anomalies. If auditors do not develop expectations or have poorly developed expectations, they may miss anomalies indicative of fraud (Appelbaum et al., 2018). However, not all researchers believe that ratios are useful in predicting fraud. Trigueiros (2019) argued that ratios have limited predictive value for multivariate models.

Although SAS No. 99 requires auditors to use analytical procedures to aid the audit planning process, the standard does not specify how auditors should perform the analytical procedures. Also, the standard does not require practitioners to use a mechanized approach to perform the required analytical procedures (Appelbaum et al., 2018). Appelbaum et al. (2017) indicated that audit evidence is changing because of the evolving and complex information
technology environment. Appelbaum et al. (2017) posited that auditors began using analytical procedures in the 1970s to analyze financial data, which resulted in lower audit costs compared to detailed testing. Janvrin and Weidnmier Watson (2017) indicated that business data has evolved from paper-based systems to legacy computer-based systems to enterprise resource-planning systems. Appelbaum et al. (2017) found that although the transition from paper to digital can facilitate cost-efficient analytical reviews, the electronic information can be manipulated much easier than paper-based evidence.

SAS No. 99 classifies fraud risk factors using three conditions: (1) incentive or pressure, (2) opportunity, and (3) attitude or rationalization (AICPA, 2002). Donald Cressey, an American criminologist, first discovered the importance of viewing fraud risk factors in terms of these three conditions. This theory became known as the fraud triangle theory (Lin et al., 2015). Incentive or pressure relates to the pressure or incentive that motivates a perpetrator to commit fraud. In this situation, the fraud perpetrator may experience pressure resulting, for example, from drug addiction or gambling problems, creating a need for money to support the addiction. Opportunity relates to the opportunity for the perpetrator to commit fraud, such as the lack of management supervision or weak internal controls. Without proper internal controls, the perpetrator does not fear detection. Finally, a perpetrator develops a rationalization to justify committing fraud (Huang et al., 2017).

Although SAS No. 99 does not refer to the three conditions as the fraud triangle theory, the standard uses the theory to categorize fraud risk factors and helps guide auditors as they assess the risk of fraud in the financial statements (Boyle et al., 2015). Huang et al. (2017) found that pressure or incentive had the highest weight of the three conditions followed by opportunity and then by attitude or rationalization. Huang et al. (2017) noted that the top five risk factors
included: (1) "Poor performance," (2) "The need for external financing," (3) "Financial distress," 
(4) “Insufficient board oversight," and (5) “Competition or market saturation” (p. 1354). Huang 
et al.’s research indicates that companies with poor performance or ones requiring external 
financing have more incentive to commit fraud to obtain financing. Using data analytics, auditors 
can profile financial data to determine whether companies are susceptible to fraud and to identify 
high-risk areas based on fraud risk factors. Also, Omidi et al. (2019) posited that company 
executives have additional incentives for exaggerating the performance of companies because 
executives want to get promoted or receive higher pay or a bonus. Auditors must take advantage 
of all available tools and techniques to help detect accounting fraud in a timely fashion. SAS No. 
99 provides the foundation on which auditors can build a sophisticated data analysis approach to 
detect fraud in financial statements (Lin et al., 2015).

Data Analytics

Analytical procedures provide auditors the ability to sift through voluminous amounts of 
aggregated financial data in search of anomalies or unexpected relationships. Christy et al. 
(2015) described an anomaly as an outlier that is inconsistent with the data being reviewed. 
Without the appropriate technology, auditors will have difficulty isolating anomalies. Also, 
performing analytical procedures manually takes time and may result in human errors (Brands 
& Holtzblatt, 2015). In the past, auditors performing manual data analytics used a significant 
portion of the audit budget to manually add columns of numbers, determine differences among 
accounts by period, develop trend lines, and calculate ratios (MacManus, 2017). Because 
auditors spent a good portion of their time on manual tasks, they did not have time to adequately 
analyze the results of the analytical reviews, which negatively impacted audit quality (Bănărescu, 
2015; West & Bhattacharya, 2016a).
At first, data analytics techniques were limited, primarily due to limited technology (Schneider et al., 2015). For example, Brown-Liburd and Vasarhelyi (2015) noted that data collected at the start of the computer age existed in the form of punched cards, paper tape, and magnetic tape, which made it difficult for auditors to analyze data without the proper technology. With the increased access to electronic data and advanced technology, auditors began performing analytical procedures more efficiently, allowing more time for analysis and investigation. Anomalies and trends not noticed in the past by auditors can now be identified more easily with technology (West & Bhattacharya, 2016a).

Salijeni et al. (2019) indicated that since the 1960s, audit firms have used technology to perform data analytics. Practitioners have referred to computerized data analytic tools as Computer Assisted Audit Techniques (CAATS), which include statistical sampling capabilities (Salijeni et al., 2019). To assist auditors with data analytics, technology companies, in the 1980s, began to develop specialized software such as Interactive Data Extraction and Analysis (IDEA) data analysis software and Galvanize, formerly known as Audit Command Language (ACL) (Salijeni et al., 2019). As auditors began using technology to perform audits, they used narrow, rules-based definitions to identify exceptions. These exceptions related to violations of company and accounting policies, which were more like compliance or operational audits as opposed to financial audits. Auditors have used exception reporting to highlight issues noted during an audit. The exception reporting represented one of the first steps towards a more advanced analytical process (F. Tang et al., 2017). To advance to the next level of analytics, auditors must shift their focus from exceptions to anomalies in samples, according to Earley (2015).

With data warehousing, companies began to store large amounts of electronic business records. To mine this data, auditors required a way to manage and efficiently review the vast
amount of electronic data. To help understand and interpret the large volume of information, auditors used statistical and quantitative analysis (Janvrin & Weidenmier Watson, 2017). J. Tang and Karim (2018) found that accounting practitioners use data mining techniques to find patterns in financial data such as journal entries. Further, auditors use data mining techniques such as classification, clustering, and regression analysis to help reduce an entire data population to a subset of data to better analyze the data. In other words, data mining techniques can help eliminate expected transactions to concentrate on unexpected transactions and anomalies (Choi et al., 2018).

In the past, auditors used sampling techniques such as simple random sampling, stratified sampling, and discovery sampling to detect fraud (Abdallah et al., 2016; Ekin et al., 2018). However, Dilla and Raschke (2015) argued that because of the nonrandom and deliberate nature of fraud committed by executives, traditional audit methods using statistical sampling would not necessarily detect fraud. Cao et al. (2015) posited that one of the significant benefits of using data analytics is the ability for auditors to test 100% of a population and not just a sample. The reliability of the sample depends on how accurately the sample represents the total population.

By reviewing a complete population of transactions and not just a sample of the population, auditors have a better chance of detecting fraud. Wang and Cuthbertson (2015) posited that larger sample sizes or even entire populations should be used to identify meaningful trends and patterns. However, they noted that there is a risk that the auditor will not be able to follow up on all the false positives or false negatives generated by using an entire population compared to a sample. Baader and Krcmar (2018) indicated that the task of analyzing false positives uses valuable time auditors could use on investigating true anomalies. Cao et al. (2015) found that false positives can be reduced by better identifying what represents true anomalies.
**Advanced Data Analytics**

From its introduction in the 1960s, data analytics has continued to evolve in terms of data and technology. Data usually reviewed by auditors in the past has evolved from financial data to diverse data. Gepp et al. (2018) indicated that practitioners often refer to diverse data as "Big Data." Earley (2015) posited that big data refers to an accumulation of all types of data, including structured, unstructured, financial, nonfinancial, and textual. Brown-Liburd and Vasarhelyi (2015) indicated that the detection of fraud was a top benefit of analyzing big data. However, Jin et al. (2015) pointed out that past statistical tools should not be relied upon to analyze big data. They believed that diverse data requires new analytical approaches to adequately analyze the data.

Appelbaum et al. (2017) separated diverse data into two groups: structured and unstructured. Structured data consists of the typical data found in financial records, such as purchase orders, sales orders, shipping records, accounts payables and receivables, inventory, and timesheets. Unstructured data includes emails, audio files, internet click streams, social media, news media, sensor records, and videos (Appelbaum et al., 2017). Currently, social media consists of social platforms, including Facebook, Twitter, and Google (Salijeni et al., 2019). Moll and Yigitbasioglu (2019) added to the types of diverse data available, including radio frequency identification (RFID) chips or tags that help verify inventory as well as product discussion forums that help auditors to understand the operations of their clients better. Also, Y. Chen et al. (2019) indicated that diverse data from external sources include:

- financial news,
- stock trading volume,
- debt structure indicators, and
The variety of data generated by companies can help auditors identify unexpected patterns and trends, which ultimately can lead to accounting fraud detection (Moll & Yigitbasioglu, 2019). Although diverse data provides many benefits, it can present numerous challenges to auditors as well. Not all auditors possess the required skill set to analyze diverse data adequately. Y. Chen et al. (2019) posited that the auditing profession is behind when it comes to using diverse data for audits. Also, beyond data variety, the sheer volume of data can overwhelm auditors. Auditors may find it challenging to review the vast amount of diverse data to detect accounting fraud (Appelbaum et al., 2017). Overall, regardless of the types and volume of data available, auditors still need to ensure the data is verifiable and reliable (Appelbaum et al., 2017).

Traditionally, auditors have reviewed financial data to determine whether a company's financial statements fairly state its financial position. By using diverse data, auditors can go beyond just using financial data and include textual data. Throckmorton et al. (2015) found that using linguistic analysis to analyze texts of earnings calls and interviews of company executives could help identify deception. However, they did not find that linguistic characteristics had predictive power. Y. Chen et al. (2017) argued that practitioners could enhance their chances of accounting fraud detection by analyzing textual information. By taking a quantitative approach to analyzing textual data, auditors can enhance their chances of detecting accounting fraud as opposed to just using financial data (Y. Chen et al., 2017). Also, Y. Chen et al. indicated that text mining methods could be used to identify cues for deception detection, including the use of linguistic inquiry and word count (LIWC). However, the accounting profession must be cautious when attempting to use the latest technologies and analysis techniques such as LIWC. For
example, auditors will likely find it challenging to use LIWC as it requires the user to generate psychological words, which requires significant human judgment (Y. Chen et al., 2017).

Further, Y. Chen et al. (2017) indicated that NLP is another technique that can be used to analyze textual information for clues of accounting fraud. Finally, Gepp et al. (2018) posited that the textual content from financial statements could be analyzed to help detect accounting fraud. Montesdeoca et al. (2019) posited that by analyzing the language used in financial statements, auditors might be able to identify deception, uncertainty, or negative feelings. Omidi et al. (2019) argued that a high rate of negative sentiment may indicate that financial reports are fraudulent. To use these advanced technologies successfully, auditors must obtain the appropriate textual analysis training. For example, textual analysis requires the user to have an initial training period using reports with known fraud to develop a list of words to distinguish between fraud and non-fraud (Gepp et al., 2018). As stated earlier, the textual analysis may overwhelm auditors because of the steep learning curve, which may impact audit quality. The accounting profession must decide at what point the use of diverse data and enhanced technology no longer adds value to an audit.

**Predictive Models**

Not so long ago, statistical data analysis represented the most advanced analytics (Appelbaum et al., 2017). Currently, predictive models represent the next level of advanced data analytics. Omar et al. (2017) discussed how auditors can move beyond conventional methods of data analysis such as trend analysis, common-size financial statement analysis, and financial ratios to build predictive models.

However, before moving directly into the details for specific predictive models, this section provides a general overview of predictive models and how they operate. Accounting
fraud detection requires auditors and investigators to find anomalies, patterns, and trends in data. Anomalies represent extreme values existing outside the class or the norm that are not expected (Goldstein & Uchida, 2016; Leite et al., 2018). Sophisticated fraudsters make detecting anomalies extremely difficult because of the extensive concealment and deception techniques they use. In most cases, company management understands their businesses and financial operations much better than auditors do. If desired, company managers can hide fraudulent transactions deep in financial systems or manipulate fraudulent transactions to appear legitimate (Omidi et al., 2019). These deception techniques place an auditor at a distinct disadvantage when it comes to detecting accounting fraud (Zainudin & Hashim, 2016).

Predictive models can assist auditors in detecting anomalies in vast amounts of financial and textual data. Salijeni et al. (2019) found that predictive models improve professional judgment, which ultimately enhances an auditor's ability to detect accounting fraud (J. Tang & Karim, 2018). Predictive modeling attempts to target objects of importance using sophisticated computational methodologies such as algorithms and machine learning (Trigueiros, 2019). As discussed above, advanced data analytics uses statistical analyses that point the auditor in the right direction to detect accounting fraud. Nevertheless, this type of analysis can only help an auditor so much. By using complex algorithms, models learn from the data to better predict fraudulent transactions (Sadgali et al., 2019). However, without quality labeled data, models cannot learn, because the models do not have the appropriate examples from which to learn. In other words, labeled data is an essential ingredient to build classification models (Miller et al., 2020).

Algorithms provide the primary engines for the models’ operations. Advanced statistical techniques, theorems, and even biological neural networks provide the foundation for the
algorithms. An algorithm uses a set of rules, like a computer program, to map the input data to output (Lomborg & Kapsch, 2019). In some cases, the algorithm can generate its own rules, as is the case with the artificial neural network (Omar et al., 2017). Based on the algorithm used, the model places the output data in different groups using classifiers. When auditors correctly set up a model, the algorithm identifies transactions that are not normal and considered outliers. With respect to accounting fraud detection, algorithms assign transactions into two groups: fraudulent and nonfraudulent (Omidi et al., 2019). Benford's Law uses an algorithm to detect the potential manipulation of financial transactions. The law uses the logic that the digit 1 in the first position occurs more often than the rest of the digits—2 through 9—occur in the first position. Auditors use Benford's Law to test the authenticity of vendor payments or test employee compliance with a company's procurement card policy. Even though Benford's Law uses an algorithm, this technique does not represent predictive modeling since machine learning is not used (Tutino & Merlo, 2019).

Machine learning consists of supervised learning, unsupervised learning, and semi-supervised learning. The main difference among supervised, unsupervised, and semi-supervised learning relates to the labeling of training data sets. For supervised learning, the auditor uses historical data and labels the data set as fraudulent or nonfraudulent. Auditors know the labels beforehand. In this case, the algorithms can learn from the data by using examples provided by the auditor. The auditors label the data using auditors’ professional judgment. However, because supervised learning requires labels, collecting labeled data sets may be challenging, limiting the use of the supervised learning technique (Sun & Vasarhelyi, 2018). With unsupervised learning, the auditor does not label the data sets. Auditors do not know the labels beforehand. The unsupervised learning algorithm automatically analyzes the data for a normal data pattern and
then identifies data that does not match the normal pattern and is outside the expected pattern (Omidi et al., 2019). Therefore, the unsupervised approach does not rely on labeling the data beforehand. Finally, semi-supervised learning exists between supervised and unsupervised learning since some of the data is labeled while a much more significant portion of the data is unlabeled. In this situation, the goal is to train the classifier based on the labeled and unlabeled data (Abdallah et al., 2016).

Supervised learning consists of two classic types of algorithms: regression and classification. Regression algorithms consist of linear regression, simple regression, and logistic regression. Classification algorithms consist of artificial neural networks, decision trees, Naïve-Bayes, and support vector machine techniques (Abdallah et al., 2016). Unsupervised learning includes two commonly used algorithms: clustering algorithms such as K-means techniques and dimensionality reduction algorithms such as Principal Component Analysis (PCA) (Abdallah et al., 2016).

Predictive models require input variables, also known as predictors or features. The models use the input variables to predict fraud based on the examples of fraud provided to the model during the training period. A predictive model classifies new data as fraudulent or nonfraudulent based on the parameters established during training. The input variables provide the model with data points to measure in order to make predictions (Omidi et al., 2019). Models commonly use financial variables such as total revenue, total assets, net income, and other financial information from a company's financial statement (Hájek & Henriques, 2017; Rahimikia et al., 2017). By using financial ratios as financial variables, auditors can add more intelligence to the financial input variables. Omidi et al. (2019) found that the use of financial ratios appeared to be the most conventional approach for input variables. Omar et al. (2017)
noted that selecting financial variables representing the three elements of the fraud triangle provided auditors with the best financial variables to use. Solvency ratios such as debt to equity or total debt to total assets represented the pressure element. Asset turnover ratios and firm size represented opportunity, and profitability ratios represented rationalization (Omar et al., 2017).

However, Omidi et al. (2019) and Trigueiros (2019) argued that while financial ratios can be useful for financial analysis, they may not work for predictive models. Because the selection of financial ratios can be subjective, they may unduly influence the predictive model if the right financial ratios are not selected for the situation. Gepp et al. (2018) argued that by using a two-stage process, auditors can select the most pertinent input variables to use. The first stage of the selection process uses two decision tree algorithms: Classification and Regression Tree (CART) and Chi-Squared Automatic Interaction Detector (CHAID). Stage two uses the variables selected in stage one. By using this two-stage selection process, auditors improve their chances of selecting relevant variables to use in the model (Gepp et al., 2018).

Auditors should use a diverse selection of financial variables to ensure that all types of financial statement frauds have a chance of being detected. In general, accounting fraud consists of overstating or understating a company's revenue, assets, expenses, or liabilities. To address these accounting frauds, Hájek and Henriques (2017) argued that financial variables should include nine categories:

- firm size
- corporate reputation
- probability ratios
- activity ratios
- asset structure
- business situation
- liquidity ratios
- leverage ratios
- market value ratios

Beyond financial variables, the literature has addressed the value of using linguistic variables that come in the form of textual information contained in the Management Discussion and Analysis (MD&A) section of a company’s annual report. Hájek and Henriques (2017) found that the models using a combination of financial and linguistic variables performed better than models just using financial variables to predict accounting fraud. For their research, Hájek and Henriques (2017) used linguistic variables from a company's MD&A, including positive and negative word counts. Also, Hájek and Henriques (2017) used words denoting tone, uncertainty, and litigation as well as strong and weak modal words. Their study also considered the frequency of constraining words used in the MD&A. Overall, auditors must select variables that provide the most information and the least amount of noise (Hájek & Henriques, 2017).

Predictive models require training and testing. If auditors do not train and test their models appropriately, then the models will fail to accurately make predictions, even if the model uses the best input variables. To ensure that a model has enough sample data for adequate testing, auditors must randomly split the initial sample data set into training and testing data. The data split consists typically of 70% for training and 30% for testing (Lin et al., 2015). With machine learning algorithms and input variables selected, auditors load the training data into the model. Predictions during this initial training may be significantly off, requiring adjustments to the algorithms. Auditors can change the weights for certain factors to improve the model's predictions. Since the training data set for supervised learning consists of labeled data, the
auditor knows when the model makes accurate predictions. In other words, the auditor knows the answers beforehand because the data is labeled. However, this is not the case for unsupervised learning, where the data is not labeled beforehand.

Auditors must be careful when selecting data to train predictive models. If the training data has an imbalance of classes, this could result in the overfitting of the model, resulting in data misclassification. For accounting fraud detection models, the classes consist of fraudulent and nonfraudulent. Therefore, if the training data has an imbalance of nonfraudulent companies, then the model may classify too many companies as nonfraudulent, including some that may be fraudulent. Dutta et al. (2017) used the Synthetic Minority Oversampling Technique (SMOTE) to overcome data imbalance.

Also, auditors should be concerned with too many false positive (FP) and false negative (FN) predictions. False positives occur when the model classifies a company as fraudulent when it is nonfraudulent. False negatives occur when the model classifies a company as nonfraudulent when the company is fraudulent. In some cases, a significant false-positive rate could overwhelm an auditor and cause auditors to exceed their allotted audit budgets as they spend significant time resolving the false positives (Salijeni, 2019). Omidi et al. (2019) found that adding a third class labeled suspicious could help limit false positives. The suspicious class would require additional investigation.

To adjust predictive models during the training phase, auditors use sensitivity, specificity, and accuracy measurements. Sensitivity, also known as recall, measures positive predictive value using this formula: true positive rate (TPR) = (TP)/(TP + FP). Specificity, also known as precision, measures negative predictive value using this formula: true negative rate (TNR) = (TN)/(TN + FN). Accuracy, a validation parameter of precision, uses the formula (TP + TN)/(TP
If an auditor desires to identify all fraudulent companies, they should select the sensitivity measurement to adjust their models since this will likely result in capturing all fraudulent companies but will also include more false positives than the specificity measurement (Ekin et al., 2018; West & Bhattacharya, 2016b). On the other hand, if the auditor desires to limit false positives, then specificity should be selected because this measurement limits the number of false positives but will likely misclassify fraudulent companies as nonfraudulent. The selection of performance measures depends on what the auditor deems most important—identifying all fraudulent companies or limiting false positives.

The literature provides a plethora of learning algorithms from which to choose. Sadgali et al. (2019) put forth the following predictive techniques.

- logistic regression
- decision trees
- classification and regression tree
- C4.5 decision trees
- neural networks
- probabilistic neural network
- support vector machines
- Naïve Bayes
- Bayesian belief network
- Bayesian skewed logit model
- K-nearest neighbor
- bivariate probit model
Also, Sadgali et al. (2019) presented the following artificial and computational intelligence techniques:

- Genetic Algorithm
- Genetic Programming
- Scatter Search
- Hidden Markov Model
- Iterative Dichotomiser
- Artificial Immune System
- Artificial Immune Recognition System
- Artificial Neural Network
- Multilayer Perception Algorithm
- Parenclitic Network
- Multi-Layer Feed Forward Neural Network

Which input variables to select is just one of the many vital decisions auditors must make as they build their predictive models. Another critical decision relates to which learning algorithm to use. Omidi et al. (2019) argued that predictive models involve mainly supervised learning. As shown in Table 1, the predictive models using supervised learning and classification algorithms have an accuracy rate as high as 92.8%. With classification algorithms, only a finite number of classes can exist, which works well for fraud prediction models since only two or three classes are used—fraudulent, nonfraudulent, and suspicious (Dutta et al., 2017; Omidi et al., 2019; X. Tang et al., 2018). Therefore, this paper focused on data mining techniques using supervised learning and classification algorithms including:

- Artificial Neural Network
• Decision Tree
• Naïve Bayes
• Support Vector Machine

Table 1

*Accuracy Measures for Supervised Learning Classification Algorithms*

<table>
<thead>
<tr>
<th>Research</th>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin et al. (2015)</td>
<td>Regression Logistic</td>
<td>88.5%</td>
</tr>
<tr>
<td></td>
<td>Decision Trees</td>
<td>90.3%</td>
</tr>
<tr>
<td></td>
<td>Artificial Neural Network</td>
<td>92.8%</td>
</tr>
<tr>
<td>Kim et al. (2016)</td>
<td>Support Vector Machines</td>
<td>85.4%</td>
</tr>
<tr>
<td></td>
<td>Bayesian Networks</td>
<td>82.5%</td>
</tr>
<tr>
<td>Dutta et al. (2017)</td>
<td>Decision Tree</td>
<td>79.9%</td>
</tr>
<tr>
<td></td>
<td>Artificial Neural Network</td>
<td>77.7%</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
<td>60.1%</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machines</td>
<td>73.4%</td>
</tr>
<tr>
<td>Sadgali et al. (2019)</td>
<td>Decision Tree</td>
<td>73.6%</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
<td>67.3%</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machines</td>
<td>65.8%</td>
</tr>
</tbody>
</table>

The artificial neural network (ANN) uses common neural networks such as learning vector quantization, multi-layer feed-forward neural networks, and radial basis function networks. The brain's biological neuron network provides the framework for the artificial neural network. As artificial neurons fire or activate, they trigger other neurons in different layers to
activate (Dutta et al., 2017; Omar et al., 2017). Neurons activate depending on the weights in each neural network layer. The weights can be adjusted based on training the network using a set of examples. The neural network uses three layers: an input layer, a hidden layer, and an output layer. The iterative process of running training data sets through the model adjusts the network to improve predictions. The ANN models are adaptive and trained to make predictions based on examples (Omar et al., 2017). The most popular ANN models include multi-layered feed-forward neural networks (MFFNN) and probabilistic neural networks (PNN). Omidi et al. (2019) found that the MFFNN model yielded the most accurate classification results. However, because of their complex structures, Omidi et al. (2019) found that the artificial neural network was not the most popular predictive model to use. Also, as Omar et al. (2017) pointed out, the disadvantage of using the ANN predictive models relates to the extensive training that the networks require to recognize patterns.

The decision tree (DT) algorithm uses a tree-like structure where the branches represent possible outcomes based on if-then steps. The choice for each decision is either yes or no, a binary choice. The decisions relative to the input variables create branching, which ultimately leads to the classification of data, producing a prediction. Algorithms that use the decision tree model include ID3, CART, and C5.0 (Dutta et al., 2017; West & Bhattacharya, 2016a).

The Naïve Bayes (NB) model uses the Bayes theorem of conditional probability, which relies on information relating to prior or past events. The NB algorithm is considered a probabilistic classifier and can be used to classify textual or numerical data. Dutta et al. (2017) found that researchers generally use the NB algorithm to detect accounting fraud, and the model produces a low number of false positives and false negatives.
The support vector machine (SVM) algorithm uses a linear model to map input vectors to specific classes of output (Kim et al., 2016). SVM uses statistical learning theory and structural risk minimization (S. Chen, 2016). Dutta et al. (2017) found that the SVM model was the most widely used classification technique.

To illustrate the predictive model building process, this paper used the study conducted by Lin et al. (2015), who used three types of predictive models: a logistic regression model, an artificial neural network, and a decision tree. Lin et al. (2015) used 32 fraud factors representing the three elements of the fraud triangle: pressure and incentive, opportunity, and attitude and rationalization. The fraud factors included information such as conventional financial ratios, sales growth rates over time, percentage of related party transactions, and the number of changes in CEO, CFO, and external auditors in the past three years. Lin et al. (2015) selected 129 fraud companies and 516 non-fraud companies. To avoid oversampling, the study selected four non-fraud companies for every fraud company, as fraud does not occur equally in all companies. Next, Lin et al. (2015) randomly divided the sample data set into a training data set and a testing data set. The training data set represented 70% of the fraud companies and 70% of the non-fraud companies. The study used the remaining 30% of the fraud and non-fraud companies to test the effectiveness of the predictive model (Lin et al., 2015).

For the Logistic Regression model, Lin et al. (2015) found that the overall classification accuracy totaled 83.7% for the training data set and 88.5% for the testing data set. For the Decision Tree model Lin et al. (2015) found that the overall classification accuracy totaled 90.4% for the training data set and 90.3% for the testing data set. Finally, for the Artificial Neural Network model, Lin et al. (2015) found that the overall classification accuracy totaled 91.2% for the training data set and 92.8% for the testing data set. Therefore, the Artificial Neural
Network performed the best with respect to overall classification accuracy. However, it should be noted that predictive models require massive amounts of training data to properly train learning algorithms. In the above example, only 129 fraud companies were used, which does not appear to be enough labeled data to adequately train a classification model (Kokina & Davenport, 2017).

The literature has shown that predictive models can effectively detect accounting fraud; however, the training required to build predictive models creates a significant burden on auditors as they push to meet client deadlines (Dutta et al., 2017; Omidi et al., 2019; J. Tang & Karim, 2018). Auditors must perform audits within budget given the razor-thin operating margins because of the extensive competition among audit firms. If models require thousands of iterative steps before they work correctly, then auditors will quickly exceed their budgets. Also, high false positive rates could significantly impact audit timing and budget. Because of the time-sensitive nature of financial statement audits or fraud investigations, auditors and investigators may not have the time to gather the right data set, train a model, and resolve any false positives. Finally, the complexity of predictive models may prove too much for auditors to handle and cause more cost than benefit, especially if auditors miss clear accounting fraud because they are too overwhelmed with building models.

**Potential Themes and Perceptions**

The literature has provided a window through which to view the evolution of accounting fraud and the various techniques for detecting it. The literature has made it clear that auditors and fraud investigators have no chance of detecting fraud in a timely manner without a robust data analytics approach (Montesdeoca et al., 2019; Salijeni et al., 2019). With advances in technology, auditors can review more than just samples; they can review 100% of data
populations. Auditors must do so by generating the least number of false positives. To do this, auditors must shift their focus from exceptions to anomalies. In addition, auditors and fraud investigators must acquire the appropriate skills to use the next level of technology. Finally, to detect anomalies, auditors and fraud investigators must move to the next level of data analysis and begin using predictive models and artificial intelligence. However, it is not as easy as gathering more data and analyzing the data using familiar techniques. Labeled data must be gathered to train predictive models using learning algorithms. Unfortunately, labeled data is not so easy to come by, which presents a significant roadblock to building machine learning models. The lack of labeled data is one of the major themes discussed throughout this paper.

**Summary of the Literature Review**

Accounting fraud continues to be a serious problem. Although the use of technology to detect accounting fraud has increased since 2012, auditors have not used advanced data analytics to its fullest extent. Auditors typically analyze financial data to determine the reasonableness of financial statements. However, audit quality could be improved if auditors analyze a combination of financial, nonfinancial, and textual data. To effectively detect accounting fraud, auditors must possess the appropriate skills to design and execute sophisticated approaches to data analysis using predictive models. Today, the accounting profession has access to tools and techniques to detect accounting fraud that are far superior to the manual analytical procedures introduced in the 1970s. However, without the knowledge and skills to use these new tools and techniques, auditors cannot reap the full benefits to combat accounting fraud. Ultimately, detecting accounting fraud remains a complex task for auditors and forensic accountants, but, as the literature makes clear, advanced data analytics provides these practitioners with the best chance of detecting accounting fraud in a timely fashion.
Transition and Summary of Section 1

This study includes three sections. Section 1 provided a background for the study by discussing the research problem and providing evidence justifying the problem. Also, Section 1 discussed the deficiencies in the literature by highlighting the limited research attention directed toward the issue of the lack of labeled data, as well as the dearth of attention given identifying the most useful data analysis techniques for accounting fraud detection. By addressing these deficiencies, lawmakers, regulators, and practitioners will benefit from this study. Section 2 provides the methods used in the study, including the data collected and the qualitative data analysis techniques used. Section 3 summarizes the research results and provides recommendations for action and further study.

Section 2: The Project

The researcher used a qualitative research approach to identify the most useful predictive models for auditors and fraud investigators to use for accounting fraud detection. Section 2 discusses the proposed research method and design to conduct participant interviews and sift through the vast predictive modeling research and technical papers. Furthermore, Section 2 addresses the sampling techniques used to select and gain access to the appropriate participants with specialized knowledge as well as analyze and evaluate predictive models from previous studies and technical papers. Finally, Section 2 discusses the reliability and validity of this qualitative study.

Purpose Statement

The purpose of this qualitative case study is to examine current predictive models using financial, nonfinancial, and textual factors to identify the best models for auditors and fraud investigators to use to detect accounting fraud (Dutta et al., 2017; Sadgali et al., 2019; Salijeni et
A 2020 global study conducted by the ACFE found that external auditors detected only 4% of the accounting frauds investigated in the study. Further, the ACFE found that tips led to the detection of 43% of the accounting fraud cases (ACFE, 2020). Based on these findings, external auditors have plenty of room to improve their techniques to proactively detect accounting fraud. By incorporating financial, nonfinancial, and textual factors into the right predictive models, auditors have a better chance of identifying accounting fraud in a timely manner (J. Tang & Karim, 2018). The literature has indicated that further research is needed to identify the best predictive models and factors to use to detect accounting fraud (Appelbaum et al., 2018; Gepp et al., 2018; Gray & Debreceny, 2014; West & Bhattacharya, 2016b).

Over the last five years, researchers have introduced various predictive models using financial, nonfinancial, and textual factors, which could help auditors detect accounting fraud (Dutta et al., 2017; Lin et al., 2015; Omidi et al., 2019; Sadgali et al., 2019). In this study, the researcher identified the types of predictive models that may be useful for timely accounting fraud detection. Moreover, the researcher identified opportunities to improve predictive models by recommending techniques to efficiently generate labeled data. This researcher interviewed subject matter experts and evaluated previous studies to identify predictive models that provide auditors and fraud investigators with the best chance of detecting accounting fraud. This study sought to provide organizations and the accounting profession with a foundation for using financial, nonfinancial, and textual analyses for timely accounting fraud detection (Dutta et al., 2017; Gray & Debreceny, 2014; Salijeni et al., 2019). Finally, based on the results of this research, standard setters and the accounting profession can develop better tools and techniques so that auditors have a better chance of detecting accounting fraud.
Role of the Researcher

This project used information gleaned from interviews with participants in academia, the accounting profession, audit firms, and technology companies. Also, the researcher used data from research studies and technical papers related to predictive models used for accounting fraud detection. Therefore, the role of the researcher for this study included interviewing participants and analyzing predictive models. Beyond interviewing and analyzing data, the role of this study’s researcher included observing and interpreting the data collected. In quantitative studies, researchers use measurements derived from the data they analyze in order to interpret the meaning of data collected. With qualitative studies, researchers do not always have a focal point such as measurements to signal that an interpretation is needed (Stake, 2010). Qualitative researchers must be reflexive as they interpret the meaning of the data collected (Creswell, 2014; Stake, 2010).

Researchers must consider quantitative aspects of predictive models along with qualitative aspects. Over the years, researchers have developed methods to test the accuracy of predictions generated by models (Ekin et al., 2018; West & Bhattacharya, 2016b). Besides these quantitative measurements, this researcher evaluated human factors relating to auditors and fraud investigators, such as time pressures, budget constraints, and technology challenges. These qualitative factors influence how well predictive models work in the field. By only evaluating the accuracy of model predictions, researchers could miss valuable insights contained in qualitative data (Stake, 2010).

To determine which predictive model works best for auditors and fraud investigators, the researcher disaggregated the selected models to understand better how they work. By taking the models apart, the researcher could study the components of each model. This helped avoid
oversimplifying or overcomplicating data interpretations. Further, the researcher did not compare the models but rather observed how the models worked in broad terms. By doing this, the researcher was able to identify the importance of labeled data to build quality predictive models.

Finally, education and professional experiences aided the researcher in interpreting the data. However, education and professional experiences can also present biases that influence the researcher’s interpretations (Creswell, 2014; Stake, 2010). The researcher for this study has extensive professional experience performing internal and external audits as well as performing fraud investigations. Also, the researcher has an undergraduate degree in accounting and a graduate degree in financial crimes and compliance management. Further, the researcher has a Certified Public Accountant (CPA) license and a Certified Fraud Examiner (CFE) certification. Because of the heavy concentration in auditing and investigating education and work experiences, the researcher had the potential to view the data collected from the perspective of an auditor and fraud investigator and not as a researcher. Therefore, this researcher sought to collect and interpret all data in a holistic manner. By developing protocols to properly conduct interviews and perform a systematic analysis of prior research, the researcher was able to identify and report research that addressed the research problem thoroughly and fairly.

Participants

The primary source of information for this study came from interviewing participants in a variety of fields including academia, accounting, auditing, and technology. To recruit participants with the proper expertise in auditing, investigating, and predictive modeling, this researcher contacted the ACFE and the Institute of Management Accountants (IMA) research departments and obtained referrals to experts with the appropriate expertise to participate in this study. With respect to academia, the researcher contacted universities and colleges with which
the researcher has professional contacts and obtained introductions to individuals with backgrounds in data science. Finally, the researcher contacted technology, auditing, and accounting firms with which the researcher has professional contacts and obtained introductions to individuals with backgrounds in predictive modeling and artificial intelligence. The researcher contacted and invited the individuals to participate in this study. All participants used for this study are clearly documented in Section 3.

Since this study involved participants, the researcher developed strong protocols to protect the human participants. This researcher issued informed consent forms to each participant as well as obtained approvals, when necessary, from organizations. Also, the researcher avoided deceiving participants and organizations and clearly communicated the purpose of the study and how the collected data would be used prior to any interaction with participants. It should be noted that this study did not involve vulnerable groups such as children (Yin, 2014). Finally, the study’s researcher also used secondary sources including research studies and technical papers prepared by academia, the accounting profession, and companies that provide predictive modeling software for accounting fraud detection. These secondary sources did not involve participants.

**Research Method and Design**

This study used a qualitative case study research approach. This method and design provided the best opportunity to understand current predictive models used for accounting fraud detection (Creswell, 2014; Purda & Skillicorn, 2015; Stake, 2010). Researchers use qualitative studies to understand a concept instead of improving a concept. The researcher must first understand what is currently happening, especially if the concept, such as predictive modeling and artificial intelligence, is in the early stages of development (Stake, 2010). Since this study
sought to develop an in-depth analysis of multiple cases (multiple predictive models), the qualitative case study approach provided the best opportunity for the study’s researcher to view the big picture and understand the complex interactions of existing predictive models using financial, nonfinancial, and textual analyses for accounting fraud detection (Appelbaum et al., 2018; Dutta et al., 2017; Keele, 2007).

Creswell (2016) noted that good qualitative research results from using multiple sources of data. For this study, the researcher used two sources of data: (1) primary source data consisting of data collected from interviews with participants who have a specialty, experience, and background in predictive modeling and (2) secondary sources consisting of data collected from a systematic review of research and technical documentation. Yin (2014) stated that interviews provide important evidence to support case study findings. Also, the literature has shown that conducting interviews with subject matter experts is a primary source of data to gain insights into the use of technology to solve real-world issues. Boyle et al. (2015), Curtis et al. (2016), Desai et al. (2017), Hegazy et al. (2017), Salijeni et al. (2019), and F. Tang et al. (2017) interviewed participants who specialized in data analytics, auditing, and technology to gain insights into improving audits and forensic investigations by using data analytics and technology. Further, Moll and Yigitbasioglu (2019) suggested that it may be useful to conduct interviews with accounting professionals to understand how accountants can better use technology and incorporate big data analytics and artificial intelligence into their work processes. By interviewing participants who have firsthand knowledge of the latest predictive modeling technology, this researcher obtained the necessary insights to address the research questions posed in this study.
Also, this researcher performed a systematic analysis of the academic and professional literature as well as comprehensive documentary sources to identify useful predictive models that auditors and fraud investigators can use for accounting fraud detection. Amani and Fadlall (2017), Appelbaum et al. (2018), Gepp et al. (2018), Montesdeoca et al. (2019), and Sadgali et al. (2019) used similar methodologies to analyze literature for their research. Their studies employed comprehensive and systematic analyses of literature to identify the latest data analytic techniques used by the accounting profession. The research methodology allowed this researcher to efficiently view an array of predictive models, providing the best opportunity to close the gap between theory and practice. In other words, by synthesizing the latest predictive modeling research, the researcher identified the most practical combination of models for accounting fraud detection.

This study’s researcher used the guidelines provided by Keele (2007, as cited by Appelbaum et al., 2018) for performing a systematic analysis of the literature. Keele stated, in general, that a systematic analysis is conducted to:

- summarize existing research,
- identify any gaps in the research, and
- suggest areas for new research activities.

The researcher lays out the analysis protocol in the next sections: Population and Sampling, and then Data Collection and Data Analysis.

**Population and Sampling**

For the primary source of data, the researcher selected participants to interview using purposeful sampling from four categories of professionals: (1) academic; (2) accounting, auditing, and investigative professional associations; (3) audit and investigative firms; and (4)
technology firms. By selecting participants from these four categories, the researcher obtained a balance of views relating to predictive models. While the literature does not recommend a specific number of participant interviews to conduct for qualitative research, it does provide various suggestions. Curtis et al. (2016), Desai et al. (2017), Salijeni et al. (2019), and F. Tang et al. (2017) interviewed five to 20 participants for their studies. This researcher conducted participant interviews until the saturation point was achieved, which was when the same or similar information was provided from the participants (Creswell, 2014). For this study, 10 participant interviews were performed. No additional participants were required since the saturation point was achieved by the 10th interview.

For the secondary source of data, the systematic analysis, the researcher collected data generated by the same four categories used for the selection of participant interviews: (1) academic; (2) accounting, auditing, and investigative professional associations; (3) audit and investigative firms; and (4) technology firms. The researcher identified and sampled documentation using the following four steps.

First, the research identified peer-reviewed academic journal articles using top electronic databases, including:

- ABI/INFORM Collection,
- Academic Search Ultimate,
- Association for Computing Machinery (ACM),
- Business Source Complete,
- Emerald Insight,
- IEEE Xplore,
- JSTOR,
• ProQuest Central,
• ScienceDirect, and
• Springer Link.

Second, the researcher identified information developed by accounting, auditing, and investigative professional associations including journal articles, technical papers, demonstrations, discussions, and conferences relating to predictive modeling. The accounting, auditing, and investigative professional organizations included:

• the American Institute of Certified Public Accountants (AICPA),
• the Association of Certified Fraud Examiners (ACFE), and
• the Institute of Internal Auditors (IIA).

Third, the researcher identified information developed by top audit firms including articles, technical papers, demonstrations, discussions, and conferences relating to in-house and third-party predictive models. The top audit firms included:

• Deloitte LLP,
• Ernst & Young (EY),
• Grant Thornton LLP,
• KPMG LLP,
• McGladrey, and
• PricewaterhouseCoopers (PwC) LLP.

Fourth, the researcher identified information developed by technology companies including articles, technical papers, demonstrations, discussions, and conferences relating to predictive modeling. The technology companies included:

• CaseWare,
Galvanize,

IBM, and

SAS.

Since innovative technology such as predictive modeling evolves quickly, this study’s researcher limited the population of data to include only information generated in the last 5 years. By reviewing data from multiple disciplines, the researcher triangulated predictive modeling information using various sources with different points of views. A detailed discussion of this study’s reliability and validity occurs later in the paper.

**Data Collection**

This section puts forth the data collection framework for the study. As discussed in the Population and Sampling section above, the researcher collected data from four categories: (1) academic; (2) accounting, auditing, and investigative professional organizations; (3) audit and investigative firms; and (4) technology firms. For the interviews, the researcher recruited participants by sending emails to professional contacts at universities and colleges, as well as professional contacts at auditing, investigative, and technology firms. Further, the researcher sent emails to professional organizations requesting referrals for individuals with the appropriate expertise to participate in this study. The emails contained the purpose of the study and how the data would be used.

Participant interviews were conducted using Zoom video conferencing services (Zoom) and lasted approximately 60–90 minutes. With the knowledge and approval of each participant, the researcher recorded the video and audio of each participant interview using Zoom. The transcripts were generated by Zoom. The researcher compared each transcript with the interview audio to ensure the accuracy of the transcript generated by Zoom. Prior to the interviews, the
researcher provided the participants with the interview questions so that they could prepare for the questions. Advance notice of the interview questions provided the best chance to obtain full and detailed answers. The questions were open-ended and used to guide the conversation (Yin, 2014). This type of interview approach provided a more relaxed conversation, leading to insightful answers. Once the interview recordings were transcribed, the researcher reviewed the transcriptions for obvious errors and followed up with the participants to ensure the data was captured accurately and as intended. Participate interview questions are included in Appendix A.

To collect data for the systematic analysis, the researcher used keywords and search strings to collect data from each of the four categories as follows:

**Academic**

The researcher selected journal articles from electronic databases using the following keywords:

- accounting fraud
- algorithm
- anomaly detection
- artificial intelligence
- audit automation
- audit software
- audit technology
- big data
- classification
- clustering
- computational intelligence
To identify the most relevant journal articles, the researcher used search strings. For example, to identify journal articles relating to the research questions, the researcher used a combination of search terms such as “predictive model” and “accounting fraud” to form a search string. Further, the researcher ensured that the search strings captured all relevant journal articles noted in the analysis of the literature. Beyond the initial search for journal articles using search
terms and search strings, this researcher reviewed cited reference lists of selected articles for additional relevant journal articles to collect.

This study’s researcher excluded journal articles based on the following criteria:

- in-progress research and dissertations
- articles not written in English
- articles not published in the last 5 years
- articles not published in peer-reviewed journals
- articles not focused on using predictive models or artificial intelligence to detect accounting fraud

The researcher used a broad interpretation for the initial phase of the data collection process. In other words, the researcher did not exclude articles based solely on the article’s title and abstract. If articles met the initial criteria for selection, then the researcher obtained the full studies for analysis. After a review of the article, the researcher decided to include or exclude the article based on the research questions for this study. To manage the journal articles collected, the researcher used Mendeley software, which organizes articles and provides an efficient article retrieval process.

**Accounting, Auditing, and Investigative Professional Organizations**

The researcher manually collected data from the AICPA, the ACFE, and the IIA websites. The researcher first searched websites using limited keyword searches such as “predictive model,” “artificial intelligence,” and “deep learning.” Then, the researcher searched websites using dropdown menus to identify relevant information relating to predictive modeling and artificial intelligence, including fraud reports, case studies, surveys, technical papers, and webinars.
Audit Firms

The researcher manually collected data from the websites of the top audit firms listed above by searching websites using limited keyword searches such as “predictive model,” “artificial intelligence,” and “deep learning.” Then, the researcher used dropdown menus to identify relevant information relating to predictive modeling and artificial intelligence, including fraud reports, case studies, surveys, technical papers, and webinars.

Technology Firms

Likewise, the researcher manually collected data from websites of the technology firms listed above as follows: the researcher searched websites using limited keyword searches such as “predictive model,” “artificial intelligence,” and “deep learning” and then searched websites using dropdown menus to identify relevant information relating to predictive modeling and artificial intelligence, including fraud reports, case studies, surveys, technical papers, and webinars.

Data Analysis

Data analysis consisted of analyzing the data collected by taking the data apart, classifying the data, and interpreting the data. This researcher used these general guidelines to conduct the data analysis. The primary focus of this data analysis addressed predictive modeling and artificial intelligence used by auditors and fraud investigators for accounting fraud detection. Once the data was collected, the researcher reviewed and evaluated the data and identified emerging themes.

Content Evaluation

The researcher reviewed the transcribed participant interviews and the notes taken by the researcher during the interviews. The researcher sorted and coded the contents of the data
collected. For the secondary data collected, the researcher used the Mendeley software to review and analyze all selected articles to address the research questions. The researcher reviewed the following information from the data collected:

- bibliographic details (title, authors, journal, publication date, volume, issue, pages, and URL)
- reporting focus (retrospective or prospective)
- type of artificial intelligence (supervised machine learning, unsupervised machine learning, or semi-supervised machine learning)
- classifiers (for example: association, classification, clustering, regression)
- algorithm (for example: decision tree, Naïve Bayes, neural networks, or support vector machine)
- data (financial, nonfinancial, or textual)
- scoring or ranking system (likelihood of fraud)
- model accuracy (false-positives, false-negatives, true-negatives, and true-positives)
- software (for example: IDEA, Galvanize, SQL, Python or software developed in-house)
- model complexity
- time required to perform data analysis
- modeling insights and recommendations
- additional notes
**Synthesis**

Once the study’s researcher evaluated and coded the content of the data collected, the researcher synthesized the results. The researcher organized the data into themes such as supervised, unsupervised, and semi-supervised machine learning. The researcher used several layers to analyze the data and took the data apart. In other words, by using layers such as the type of algorithms used, the type of data used, or the type of software used, the researcher organized the data collected to understand the data better. Once the data were separated into groups, the researcher reassembled the data to show new insights or concepts as well as potential gaps in current auditing and investigating practices as well as the literature (Yin, 2014). To help understand the data extracted, the researcher broke down the data and placed them into categories illustrated by themes, figures, charts, and tables (Creswell & Poth, 2018; Keele, 2007; Yin, 2014). This study’s researcher tabulated the frequency of the different events found within the data to develop trends and patterns relating to predictive models. By illustrating the data, the researcher presented the data in broad terms, allowing for a better chance to interpret and synthesize the data into meaningful concepts.

**Reliability and Validity**

The researcher established procedures and strategies to ensure the reliability and validity of the research findings. Establishing certain strategies provides the readers and the users of this study with confidence that the quality of evidence is good. To ensure reliability, the researcher thoroughly documented all the steps taken to perform the research. Section 2 lays out and discusses the key components of this study. Section 3 provides the procedures used to achieve the results of this research. As discussed above, this researcher recorded and transcribed participant interviews and reviewed the transcriptions for accuracy. Also, the researcher used the
Mendeley software to document the review and analysis of the secondary data collected. By documenting the steps taken, the analysis performed, and the results attained, the researcher implemented the necessary steps to ensure the reliability of the research results.

Further, the researcher used four strategies to ensure the validity of the study. The four strategies included (1) triangulation, (2) member checking, (3) disconfirming evidence, and (4) reflexivity.

**Triangulation**

Stake (2010) believed that the use of triangulation is a win-win situation. He stated that “if the additional checking confirms that we have seen it right, we win. If the additional checking does not confirm, it may mean that there are more meanings to unpack, another way of winning” (Stake, 2010, p. 124). This study’s researcher obtained data from more than one source. The researcher used participant interviews with industry experts as the primary source of data. Also, the researcher analyzed secondary documentation to help corroborate the primary source of data. Finally, as discussed in the Population and Sampling section, the researcher collected data from four key sources, including academia, professional organizations, audit firms, and technology firms. By collecting data from a variety of sources, the researcher improved the research and validated the themes developed.

**Member Checking**

Once the specific themes and descriptions were developed, the researcher followed up with the participants to ensure that the researcher had accurately captured the information provided by the participants. This follow-up process provided the participants with an opportunity to add comments and correct any inaccuracies in the data collected (Creswell, 2014).
**Disconfirming Evidence**

The researcher combed through all the data collected from primary and secondary sources for disconfirming evidence or exceptions to the themes. The researcher understood the importance of presenting all evidence even if the evidence contradicted the working themes and hypotheses. Further, the researcher realized that not all evidence would fit neatly into positive and negative categories. The researcher considered all evidence to provide a realistic assessment of the status of predictive modeling used by auditors and fraud investigators to detect accounting fraud.

**Reflexivity**

This study’s researcher played a key role in evaluating and interpreting the data collected. Therefore, this paper includes relevant information about the researcher’s professional experiences relating to data analytics, data mining, and predictive modeling. These experiences may have shaped the themes developed by the researcher. Thus, where appropriate, the paper includes relevant stories about the researcher’s experiences relating to the concepts researched.

**Transition and Summary of Section 2**

Section 2 discussed the research method and design for this study. Based on the literature review, the researcher identified the data to collect and the best way to analyze the data to address the research questions. Section 3 discusses the research procedures and strategies preformed. Further, Section 3 presents the findings of the research, as well as providing auditors and fraud investigators with the most practical combination of predictive models to use for accounting fraud detection. Finally, Section 3 provides recommendations for action and research topics for further study.
Section 3: Application to Professional Practice and Implications for Change

Section 3 provides a summary of the results for the research and analysis performed for this study. The researcher collected data by interviewing 10 participants with expertise in predictive modeling, auditing, and investigating, as well as conducting a systematic literature review of research and technical documentation relating to artificial intelligence, predictive modeling, and accounting fraud detection. In this section, the researcher provides an overview of the study, presentation of the findings, and application to professional practice. Also, the researcher provides recommendations for actions and future study, reflections, and study conclusions.

Overview of the Study

The purpose of this qualitative case study was to examine current predictive models using financial, nonfinancial, and textual factors to identify the best models for auditors and fraud investigators to use to detect accounting fraud (Dutta et al., 2017; Sadgali et al., 2019; Salijeni et al., 2019). This study produced data that illustrated the complexities of using artificial intelligence and predictive modeling to detect accounting fraud. By providing a better understanding of the current predictive modeling and artificial intelligence environment, as well as strategies for overcoming the pitfalls, the accounting, auditing, and investigative communities will greatly benefit from the results of this study. Accounting fraud still exists, and there are no signs that it is slowing down (Omidi et al., 2019). The accounting profession understands the severity of the problem but has yet to develop audit procedures that can detect accounting fraud with any consistency or accuracy (Omidi et al., 2019). Organizations have implemented exception reporting processes to help detect fraud (ACFE, 2019). However, these fraud detection steps have been unsuccessful in terms of timely accounting fraud detection. Overall, the
accounting profession and organizations require better tools and techniques to be more successful in their efforts concerning timely accounting fraud detection.

For the primary source of data, the researcher used purposeful sampling to select 10 participants to interview from three categories of professionals: (1) academic, (2) audit and investigative firms, and (3) data technology firms. The researcher attempted to select participants from a fourth category, accounting, auditing, and investigative professional associations, but was referred to experts outside of these professional organizations. Therefore, the participant interviews did not include experts from professional organizations. Initially, this researcher planned to conduct between 10 to 20 participant interviews. After interviewing six participants, the researcher began to reach the saturation point for most of the interview questions. After the 10th participant interview, the researcher determined that the saturation point had been achieved and decided that no additional interviews were required.

The researcher conducted the 10 interviews using Zoom video conferencing services. The study design concentrated on two principal research questions. The first research question focused on the challenges of using predictive models for accounting fraud detection. The second research question focused on overcoming the challenges noted to develop a more accurate method for accounting fraud detection. The 10 participants used for this study had expertise in auditing, fraud investigation, or predictive modeling. To identify qualified participants, this researcher contacted the ACFE and the Institute of Management Accountants (IMA) research departments and obtained referrals to experts with the appropriate expertise to participate in this study. Further, the researcher contacted universities and colleges with which the researcher has professional contacts and obtained introductions to individuals with backgrounds in data science. Finally, the researcher contacted technology, auditing, and accounting firms with which the
researcher has professional contacts and obtained introductions to individuals with backgrounds in predictive modeling and artificial intelligence.

The recruitment and consent letters used in this study were approved by the Liberty University Institutional Review Board before they were sent to participants. Upon receipt of consent by the participant, the researcher sent a follow-up email to schedule the interview. The researcher received all consent letters via email prior to each interview. At the start of each interview, the researcher thanked the individual for their willingness to participate in the interview to discuss their understanding of artificial intelligence and predictive modeling. The researcher also ensured participants of confidentiality in the study. The researcher recorded the video and audio of each participant interview using Zoom. The video and audio recordings were initially stored by Zoom in the Cloud. All participant interview recordings were subsequently downloaded to the researcher’s laptop and deleted from Zoom’s Cloud storage. In other words, no recordings of participant interviews exist outside of the researcher’s laptop, which is password protected. Each participant provided consent to be recorded via signed consent forms received.

Data triangulation occurred in this research study by incorporating participant interviews, member-checking, interview notes drafted by the researcher, and a review of research and technical papers generated by academia, professional organizations, auditing, and technology firms. Interview notes included any observations or notes taken during the interview outside of participant responses to interview questions. The transcripts were generated by Zoom. The researcher compared each transcript with the interview audio to ensure the accuracy of the initial transcript. Any inconsistencies between the transcript and the audio were corrected by the researcher. In addition, the researcher used member-checking, an external validity strategy that
institutes trustworthiness and credibility in the research conclusions (Creswell, 2014). The researcher emailed the updated text copy of the transcript to the interviewee for review and comment. Included in the body of the email was a request for each participant to review the interview transcript and notify the researcher within seven days of any inaccuracies. Once participants provided comments on their transcripts, the researcher read all transcripts and recorded possible themes in an Excel spreadsheet. Finally, the researcher sorted and coded the data manually, arranging the Excel spreadsheet by the possible themes.

All 10 participant interviewees answered 20 questions regarding:
- background, education, and training (four questions)
- predictive models (five questions)
- technology skills (two questions)
- limitations and challenges (five questions)
- future (four questions)

**Participant Demographics**

The experiences of the 10 participants comprise the single case study. Table 2 outlines the participant demographics based upon current position, years working in their current position, educational background, and prior experience. Participants in the study included a data science professor, a research director, product managers for technology firms, senior accountant and managing directors for accounting and investigative firms, a chief analytics officer, and a data analytics manager. Of the 10 participants, five had a master’s degree, one had a doctoral degree, and one had a law degree. To ensure confidentiality, all participants were assigned identifiers of P1, P2, P3, P4, P5, P6, P7, P8, P9, and P10.
### Table 2

**Summary of Participant Demographics**

<table>
<thead>
<tr>
<th>ID</th>
<th>Current Position</th>
<th>Time in Position</th>
<th>Educational Background</th>
<th>Work Experience and Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Current position – Product manager for a data analytics team</td>
<td>1 year</td>
<td>BS in Chemistry, a BS in Biochemistry, an MBA and a Master’s in Electrical Engineering</td>
<td>Master’s thesis was on predicting bankruptcy using artificial neural networks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>Current position – Assistant Professor of Data Science and Program Director of</td>
<td>3 years</td>
<td>BS in Geographic Information Systems, a Master’s in Geography, and a PhD in Geography</td>
<td>Training obtained via masters and graduate programs. Work experience consisted of healthcare</td>
</tr>
<tr>
<td></td>
<td>the Data Science program</td>
<td></td>
<td></td>
<td>analytics.</td>
</tr>
<tr>
<td>P3</td>
<td>Current position – Research Director for a national crime lab</td>
<td>3 years</td>
<td>PhD in Computer Science</td>
<td>PhD was focused on predictive analytics.</td>
</tr>
<tr>
<td>P4</td>
<td>Current position – Product Lead for Data Analytics for a technology firm.</td>
<td>2 years</td>
<td>Degree in Computer Engineering with a Specialty in Digital Networking</td>
<td>Performing business intelligence for financial companies. Training included statistics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>predictive modeling.</td>
</tr>
</tbody>
</table>
### Table 2, continued.

|   | Current position – Managing Director of a Disputes and Investigations practice focusing on forensic data analytics and compliance intelligence.  
|   | Time in position – 1 year  
|   | Educational background – BS in Business Administration, Certified Fraud Examiner  
|   | Work experience and training – Working with auditing and consulting firms using anti-fraud techniques including predictive modeling.  

|   | Current position – Senior Manager, Fraud Risk Mitigation. Prior to current position, participant lead a Fraud and Security Intelligence Division for a technology firm for 10 years.  
|   | Time in position – 1 year  
|   | Educational background – BS in Economic Crime Investigation and MS in Economic Crime Management  
|   | Work experience and training – Co-inventor of two patented fraud detection models.  

|   | Current position – Managing Director focusing on data analytics in the forensic litigation consulting department.  
|   | Time in position – 4 Years  
|   | Educational background – Computer Engineering degree and Law degree  
|   | Work experience and training – Work experience includes fraud investigations. Training includes Certified Ethical Hacker, Certified Fraud Examiner, and Encase certified.  

|   | Current position – Chief Analytics Officer  
|   | Time in position – 3 years  
|   | Educational background – Computer Science degree  
|   | Work experience and training – Performed advanced technology research projects for government agencies.  

|   | Current position – Data Analytics Manager for Special Investigations Unit  
|   | Time in position – 3 years  
|   | Educational background – BS in Fire Science and Technology; MS in Financial Crimes and Compliance Management  
|   | Work experience and training – Work experience includes fraud investigations. Certified Fraud Examiner  

|   | Current position – Chief Product Officer at technology firm.  
|   | Time in position – 1 Year  
|   | Educational background – MS in Computer Science with specialization in machine learning and natural language processing  
|   | Work experience and training – Working with predictive models using large data sets to make inferences of consumer and human behavior.  

Systematic Review Analysis

For the secondary source of data, the researcher performed a systematic review analysis, which consisted of collecting data generated from (1) academia; (2) accounting, auditing, and investigative professional associations; (3) audit and investigative firms; and (4) technology firms. The researcher identified and collected data for the systematic review analysis as follows.

Academia

The researcher identified peer-reviewed academic journal articles using top electronic databases including the following:

- ABI/INFORM Collection
- Academic Search Ultimate
- Association for Computing Machinery (ACM)
- Business Source Complete
- Emerald Insight
- IEEE Xplore
- JSTOR
- ProQuest Central
- ScienceDirect
- Springer Link

To focus the database searches, the researcher performed a series of steps to identify the most relevant journal articles to address the two research questions. The initial search retrieved peer-reviewed journal articles published within the last 5 years. Table 3 shows the list of search terms and phrases used to begin the database search process.
Table 3

*Initial Search Terms and Phrases Used for Systematic Review Analysis*

<table>
<thead>
<tr>
<th>algorithm</th>
<th>clustering</th>
<th>&quot;fraud detection&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;accounting fraud&quot;</td>
<td>&quot;computational intelligence&quot;</td>
<td>&quot;machine learning&quot;</td>
</tr>
<tr>
<td>&quot;anomaly detection&quot;</td>
<td>&quot;data analysis&quot;</td>
<td>&quot;natural language processing&quot;</td>
</tr>
<tr>
<td>&quot;artificial intelligence&quot;</td>
<td>&quot;data analytics&quot;</td>
<td>&quot;predictive model&quot;</td>
</tr>
<tr>
<td>&quot;audit automation&quot;</td>
<td>&quot;data mining&quot;</td>
<td>&quot;predictive process discovery&quot;</td>
</tr>
<tr>
<td>&quot;audit software&quot;</td>
<td>&quot;data modeling&quot;</td>
<td>&quot;robotic process automation&quot;</td>
</tr>
<tr>
<td>&quot;audit technology&quot;</td>
<td>&quot;deep learning&quot;</td>
<td>&quot;statistical methods&quot;</td>
</tr>
<tr>
<td>&quot;big data&quot;</td>
<td>&quot;financial accounting fraud&quot;</td>
<td>&quot;supervised learning&quot;</td>
</tr>
<tr>
<td>&quot;classification&quot;</td>
<td>&quot;financial statement fraud&quot;</td>
<td>&quot;unsupervised learning&quot;</td>
</tr>
</tbody>
</table>

However, these initial search terms yielded far too many results to review manually. Therefore, the researcher developed the search strings shown in Table 4 to help limit the search results to the most relevant journal articles.

To efficiently cull through the volume of the search results, the researcher reviewed the title and the brief description of the article and obtained the full text of the selected articles. This effort yielded 26 journal articles to include in the research. Next, the researcher selected the most relevant articles gathered during the literature review performed at the start of this research project, which totaled 49 articles. Finally, the researcher reviewed the references used for the articles selected to identify additional articles to gather. This resulted in an additional 37 articles that were added to the repository of journal articles for this research. In total, 112 academic journal articles were selected and reviewed.
Table 4

Search Strings Used to Focus Systematic Review Analysis

"accounting fraud" and "anomaly detection"
"accounting fraud" and "anomaly detection"
"accounting fraud" and "artificial intelligence"
"accounting fraud" and "big data"
"accounting fraud" and "classification"
"accounting fraud" and "computational intelligence"
"accounting fraud" and "clustering"
"accounting fraud" and "data analysis"
"accounting fraud" and "data analytics"
"accounting fraud" and "data mining"
"accounting fraud" and "data modeling"
"accounting fraud" and "deep learning"
"accounting fraud" and "detection"
"accounting fraud" and "machine learning"
"accounting fraud" and "natural language processing"
"accounting fraud" and "predictive model"
"accounting fraud" and "statistical methods"
"accounting fraud" and "supervised learning"
"accounting fraud" and "text mining"
"accounting fraud" and "unsupervised learning"
"artificial intelligence" and "financial statement fraud"
"artificial intelligence" and "financial accounting fraud"
"artificial intelligence" and "accounting fraud"
"machine learning" and "financial statement fraud"
"machine learning" and "financial accounting fraud"
"machine learning" and "accounting fraud"

Professional Associations

The researcher obtained information developed by accounting, auditing, and investigative professional associations, including journal articles, technical papers, and discussions relating to artificial intelligence and predictive modeling. The researcher searched the websites of the following accounting, auditing, and investigative professional associations:

- the American Institute of Certified Public Accountants (aicpa.org),
• the Association of Certified Fraud Examiners (acfe.com),
• the Institute of Internal Auditors (theiia.org), and
• the Institute of Management Accountants (imanet.org).

Once on the association’s website, the researcher identified relevant information using two key-phrase searches, “artificial intelligence” and “predictive model.” The researcher continued to collect documents until the information from the search results became repetitive. The researcher selected 37 documents to include in the body of research for this project.

**Auditing and Investigating Firms**

The researcher obtained information developed by top audit firms including articles, technical papers, and discussions relating to predictive modeling and artificial intelligence. The researcher searched the websites of the following audit firms:

• BDO (bdo.com),
• Crowe (crowe.com),
• Deloitte LLP (deliotte.com),
• Ernst & Young (ey.com),
• Grant Thornton LLP (grantthornton.com),
• KPMG LLP (kpmg.com), and
• PricewaterhouseCoopers LLP (pwc.com).

Once on the website of the audit firm, the researcher identified relevant information using two key-phrase searches, “artificial intelligence” and “predictive model.” The researcher continued to collect documents until the information from the search results became repetitive. The researcher selected 32 documents to include in the body of research for this project.
**Data Technology Companies**

The researcher obtained information developed by data technology companies including articles, technical papers, demonstrations, and discussions relating to predictive modeling and artificial intelligence. The researcher searched the websites of the following data technology firms:

- CaseWare (caseware.com),
- Galvanize (wegalvanize.com),
- IBM (ibm.com),
- MindBridge (mindbridge.com), and
- SAS (sas.com).

Once on the website of the data technology firm, the researcher identified relevant information using two key-phrase searches, “artificial intelligence” and “predictive model.” The researcher continued to collect documents until the information from the search results became repetitive. The researcher selected 42 documents to include in the body of research for this project. For this research project, the researcher selected and reviewed a total 223 academic journal articles, technical papers, and discussion blogs relating to the use of predictive models and artificial intelligence to detect accounting fraud.

**Presentation of the Findings**

This section discusses the findings of this research study. The investigation focused on two research questions. Research question one concentrated on understanding why predictive models are not able to detect accounting fraud consistently and accurately. Research question two concentrated on how to use financial, nonfinancial, and textual analyses to improve the detection of accounting fraud warning signals. The researcher organized the conclusions of the
research in the presentation of findings around the themes identified to address the two research questions. Also, the conclusions of this research amplified the available knowledge explored in prior sections.

After the interviews were completed and transcribed, the researcher manually reviewed the transcripts to code and discern themes using an Excel spreadsheet to record key points made by each participant. The researcher then compressed the key points into potential themes. Upon completion of the data analysis, the researcher identified six main themes resulting from the research: (1) Data, (2) Data Mining Techniques, (3) Model Input and Output, (4) Human Agency, (5) Approach, and (6) Explainable AI. Finally, the systematic review analysis validated and reinforced the themes developed. Each theme is analyzed below.

**Theme 1: Data**

Data is the most important element when it comes to developing successful predictive models. Without data, models cannot be built, and investigators cannot search for anomalies. However, it is much more complex than simply having data. The data must represent all classes or at least two classes—fraudulent and nonfraudulent—when it comes to fraud detection (Jofre & Gerlach, 2018). The data should be complete and have no missing values or information. Any missing data values must be addressed; otherwise, the predictive model may be biased, especially when the missing values relate to the variables being predicted (Trigueiros, 2019). Models require vast amounts of data to be trained (Baader & Krcmar, 2018; Bao et al., 2020; S. Chen, 2016; Omidi et al., 2019). However, acquiring the right data presents many challenges.

While exploring these challenges, the researcher attempted to learn why predictive models do not always make accurate predictions of accounting fraud detection (research question one). However, instead of finding why predictive models are not accurate, the researcher
discovered that a predictive model for accounting fraud detection does not exist in the form of an all-inclusive model that can classify companies or transactions as fraudulent or not fraudulent with the touch of a button. There just is not enough labeled data to build the perfect predictive model. Given the current environment, there seems to be little chance or opportunity to improve the accuracy of accounting fraud predictive models without fundamental changes in business relationships and government regulations.

**Volume, Representation, and Breadth.** In exploring research question one, which asked why current predictive models used by auditors and fraud investigators are not able to consistently and accurately detect accounting fraud, the researcher found that the lack of labeled data was one of the leading reasons. P6 noted that a sufficient volume of quality data was one of the key characteristics for building successful predictive models. P6 added that sufficient volume of quality data is difficult to obtain, especially when building supervised models. Further, P9 explained that not having enough data to feed models is a mistake that organizations are making when it comes to using predictive models. “They [organizations] don’t realize how much data and the quality of data that you need,” said P9.

On the surface, collecting enough data would not seem to be a big problem, given the enormous amount of financial data generated by companies each day. However, it is not enough to have the volume of data. The data must include examples of fraudulent and nonfraudulent transactions. Without having examples of fraudulent companies or transactions, a model cannot learn to distinguish between fraud and non-fraud. If data used to train a model only represented one class, good transactions, then the model would classify all transactions as good. Therefore, for accounting fraud detection, it is critical to train models using data that includes fraudulent transactions. However, Jofre and Gerlach (2018) explained that collecting data that includes
fraudulent transactions is difficult since fraudulent transactions are usually concealed and underreported. This makes labeling data very difficult to do. Baader and Krcmar (2018) stated:

A large volume of realistic data with both fraudulent and nonfraudulent cases is necessary to detect fraudulent activities. Due to security concerns, very few companies are willing to provide their data to scientists, especially when fraud is suspected in the data. (p. 6)

**Quality and Integrity.** The research has shown that data quality is another important characteristic of successful predictive models (Bănărescu, 2015; Dutta et al., 2017; Moepya et al., 2016; Shukla & Mattar, 2019). When asked about the importance of data quality in terms of building a successful predictive model, four participants summed it up this way: “garbage in, garbage out.” In other words, if the quality of the data going into a model is bad, then the output of the model will be bad as well. The participants agreed that the quality of the data fed into a model has a direct impact on the quality of the model output. P7, P9, and P10 explained that successful predictive models started with good quality data. Bănărescu (2015) noted that data quality plays an important role in data analysis within the audit process. According to Huerta and Jensen (2017), a model using inaccurate data could mislead users. Collecting quality data to build predictive models can be another challenge and represent a barrier that auditors and investigators must overcome. For example, data collected from third-party vendors may have missing or corrupted data. Further, data gathered from audit clients may have data entry errors or duplicate transactions (Shukla & Mattar, 2019). All these data issues must be resolved before the data can be used in a model.

**Understanding the Data.** To extract value from data, auditors and investigators first need to understand the data collected. By knowing how the data was generated and what type of information is included in a data set, auditors and investigators will be in a better position to
select the best data to use for predictive models (Amani & Fadlalla, 2017; Huerta & Jensen, 2017). According to Gupta and Mehta (2020), “business financial data is considered to be high-dimensional data” (p. 3), which means that the data includes many features. For example, a feature in a vendor master file could be a vendor phone number; depending on the audit or investigation, phone numbers may be important, but they may also represent noise that degrades a model. Knowledge of a client’s business operations and groups provides context and adds intelligence to the data that otherwise would not exist (Y. Chen et al., 2019). Raw data collected during audits and investigations includes a plethora of features. Data features consist of both numerical and textual data. What features are important depends on the data being analyzed and the type of investigation being performed.

P1 believed that it is important to understand the data and know what is included in the data. P7 suggested that all data points can provide insight, but the auditor must understand the data to be aware of what data points are available for analysis. Overall, an understanding of the data will help auditors and investigators make better decisions about the characteristics to program into the model. Huerta and Jensen (2017) recommended exploring the data to better understand it. Data exploration could consist of simple statistical analysis, including calculating the minimum, maximum, and mean values of invoice amounts, as well as summarization and stratification of certain database fields. Also, visual analysis of the data could go a long way to helping auditors and investigators to understand the data. P8 noted that basic visualizations with some logic could help users better understand what is in the data.

P7 explained that the general idea of predictive models is to conduct predictions by analysis, so all data points can be useful and provide insight that are not immediately apparent without at least considering the data. The key point here is that investigators must understand the
data first so they are aware of all the data available to investigate. However, P10 warned that even though all data has value, all data should not necessarily be gathered and analyzed. P10 indicated that there are constraints to what should be analyzed. Organizations only have so many people but an infinite amount of data, so not all data can be reviewed, integrated, modeled, and improved. Therefore, one must understand the data to select the best data for analysis. In the end, all data has value, but not all data is equal, said P10.

**Data Import and Preprocessing.** Beyond the lack of data needed to develop models, another challenge is getting data in a usable format. There are no standard data formats that all companies abide by. When data is collected to be analyzed, it may be in many different formats. P1 noted that getting the data from a client quickly and easily in a readable format is a big challenge. P2 stated that a fair amount of time is spent just getting the data in one spot and preparing the data to be analyzed. P7 believed that “there’s no best data” but “only the best format” to analyze the data. Once the data is received, it must be preprocessed or cleaned and placed in an analyzable format to be “digested” by a data analytics model or machine learning model. If the data cannot be formatted adequately, then the data is useless (Y. Chen et al., 2019; Dutta et al., 2017).

As auditors and investigators take the necessary steps to understand the data, they should begin to identify areas of the data that need to be preprocessed. The data cleansing aspect of data analytics and model development should not be underestimated. P1 stated that data scientists spend much of their time manipulating and processing the data. Also, P1 pointed out the challenges with data preprocessing, such as a lack of clarity concerning how the data should be cleaned and processed. By gaining an understanding of the data, the auditor will be in a better position to transform the data into a format that can be used by a data analytics model, machine
learning model, or deep learning model. Preprocessing steps include identifying duplicate transactions and missing data in data sets (Dutta et al., 2017). To address the issue of missing data, Moepya et al. (2016) developed techniques to estimate missing values using learning algorithms that incorporate imputation procedures to estimate the missing values. Output from models that do not use full data sets may mislead auditors and investigators to make incorrect judgments. However, it should be noted that the bad output is not due to a flaw in an algorithm, but rather caused by the use of incomplete data fed to the model (Moepya et al., 2016).

For unstructured data sets such as letters to shareholders, the preprocessing steps include breaking sentences into meaningful terms or deleting unimportant terms. Current text mining tools can help automate this manual preprocessing effort (Boskou et al., 2019). Also, preprocessing data can include processing data sets to resolve class imbalance problems by splitting majority classes into multiple data sets to be included with a minor class data set (Byungdae & Yongmoo, 2020; Y. Chen et al., 2019). Ultimately, P7 warned that auditors and investigators must take care not to manipulate the raw data too much during the data preprocessing stage because what looks to be errors may be clues to fraud. P7 explained that any data point can be useful and should not be prejudged or viewed with too much skepticism in the beginning of an investigation, because this may bias the process. Data cleansing should be very limited. The original raw data should be kept intact to create a model for analysis. P7 provided an example to illustrate how a data set could easily be manipulated incorrectly. Suppose there was a customer with the name of “Tom” and suddenly there was a “Ton.” An analyst might think this was a typo and just change the name from “Ton” to “Tom.” However, the analyst might just have deleted a clue and a valuable data point, said P7. Several of the participants warned against processing the data too much. P4 stated that the less an investigator processes the raw data, the
better the chance of detecting the subtle patterns in the data. By not processing the data too much, the investigator may avoid inadvertently hiding patterns due to summarizing or using other data processing techniques.

**Structured and Unstructured Data.** In exploring research question two, which asked how financial, nonfinancial, and textual analyses can be used by auditors and fraud investigators to improve the detection of accounting fraud warning signals, the researcher found that the use of unstructured data could be a powerful new weapon in the fight against accounting fraud. Until recently, auditors and investigators used mostly structured data in the form of spreadsheets, general ledgers, and other formats consisting of columns and rows. Veering away from structured data, semi-structured and unstructured data consists of emails, text messages, news articles, corporate press releases, webpage content, analyst reports, and social media posts (Boskou et al., 2019; Lewis & Young, 2019). Towards the end of the unstructured data spectrum lies data such as images, videos, and telephone calls. As auditors begin to use less structured data, the auditors have more difficulty capturing the value of the data. But with new tools and techniques, the value of the unstructured data can be extracted once the data is placed in an analyzable format (Gandomi & Haider, 2015; Sun, 2019).

Throughout the research, unstructured data was consistently put forth as the most underutilized data with high potential (Boskou et al., 2019; Gandomi & Haider, 2015; Richins et al., 2017; Sun & Vasarhelyi, 2018). As the name suggests, unstructured data does not have structure such as columns and rows. Unstructured data takes the form of emails, free text in journal entries and business expense reports, and comments on vendor invoices. In the past, it was difficult to extract unstructured data. Text mining tools are much more advanced now and can be used for more than just keyword searches (Yoon et al., 2015). P5 believed that one of the
best ways to identify corrupt intent is to review the free text field of a journal entry. However, keyword searches are not enough to take advantage of the unstructured data. By just using keyword searches, auditors and investigators are not extracting the full value of the unstructured data. Using advanced technology such as Natural Language Processing (NLP) allows much more information to be gleaned from the data, such as sentiment and context. P6 noted that in some industries such as insurance, unstructured data may represent almost 80% of the data generated by an organization. P6 added that if investigators only use structured data, then they are only using approximately 20% of the available data to build a model. By only using 20% of the available data, the model is probably not going to generate great results, said P6. Further, included in unstructured data is metadata. P6 explained that just by looking at the “To” and “From” fields of an email, investigators can extract valuable information. P6 provided the following example to illustrate the value of metadata. By looking at email traffic, an investigator may identify issues that need additional follow-up, like an unusually high volume of email traffic between a senior executive and a low-level employee. This level of traffic between these two employees may point to a bigger issue that needs to be investigated further, said P6. In addition, P9 stressed the importance of using unstructured data to analyze a company’s data for fraud. Like the other participants, P9 believed that unstructured data has been underutilized when it comes to fraud investigations. P9 found that NLP is the best tool to mine the unstructured data in text fields.

**Labeled Data.** To successfully build a supervised classification model, labeled data is required. For accounting fraud detection, two classes generally exist—fraudulent and nonfraudulent. Without data labeled as fraudulent or nonfraudulent, the chances of building a predictive model are low (Mohammadi et al., 2020; Sarkar et al., 2020). Companies such as
Google, Yahoo, and Netflix have successfully deployed predictive models. Unlike accounting fraud detection models, the models used by these companies have plenty of labeled data. In general, customers of Google, Yahoo, and Netflix give permission to use their data. Further, for several decades, the credit card industry has successfully used predictive models to detect credit card fraud. P6 explained that the success of the predictive models used in the credit card industry is due, in large part, to the vast amount of labeled data available to train their detection models. Every day, credit card customers generate high volumes of credit card transactions, so there is no lack of data to use. Also, the data is properly labeled as fraudulent or nonfraudulent. Customers are more than willing to let banks know when fraud has occurred on their accounts. Customers proactively contact their banks to let them know which transactions are fraudulent and nonfraudulent. Essentially, credit card companies have a vast number of annotators that are willing to provide the annotating service for free. Because the labels are generated by individuals with firsthand knowledge of the fraud, the credit card companies have a high level of confidence that the data is labeled accurately.

P6 pointed out that the training data used in the credit card industry is much better than the data used for accounting fraud detection models. P6 believed that the business community has seen a lot of good quality value come out of the predictive models used by the credit card companies. Because of this, auditing firms have attempted to build and deploy similar detection models but have not realized the same level of success as the credit card companies. Audit firms do not have anywhere near the same level of labeled data as do the credit card companies, mostly because audit clients do not normally admit to accounting fraud or provide data that identifies which transactions are fraudulent. Businesses do not have data capture fields in their system
stating whether a transaction is fraudulent, said P9. Accounting fraud is concealed, and there is usually no incentive to disclose the fraud to the auditors or regulators.

Label quality and label quantity matter more than the model itself. According to P3, the lack of labeled data for accounting fraud detection models may represent the greatest challenge to building a successful predictive model. P3 suggested that challenge is due to the rare instance of fraud or at least the potentially vast underreporting of fraud. Further, P3 wondered how predictive models could be properly trained given the underrepresentation of the fraud class. The scarcity of labeled data presents a significant roadblock to building successful accounting fraud detection models. Unfortunately, the problem of limited labeled data does not seem to be going away. Given the strict data privacy laws and regulations, financial data is very difficult to obtain, especially data that is labeled as fraudulent (Huerta & Jensen, 2017).

For now, researchers use publicly available data from the SEC that identifies fraudulent companies. However, P3 pointed out that by building a model based on labeled data from only the SEC, you may be only picking up on cases that the SEC cares about. There may be other types of accounting fraud that may be missed given the limited amount of labeled data, said P3. P5 reinforced the point that the biggest limitation to predictive modeling is the lack of seed data. P5 recognized that due to the concealment of fraud, there is not a lot of labeled data available, which directly impacts the ability to properly train and test an accounting fraud detection model. Further, P8 added that before you can classify transactions or companies as fraudulent or not fraudulent, you must have labeled data to train a model.

**Data Security.** Having data in one spot can facilitate an audit or investigation. However, there is a downside to accumulating data, especially confidential data, into one repository. The data may be more vulnerable to cyberattack. Normally, sensitive data is maintained in various
silos dispersed throughout a company. By gathering sensitive data in one repository, the risk of a
data breach may increase. P9 stated that “no one person has a key to all of the vaults.” When all
data is in one repository, it may become more vulnerable to attack. Unlocking all the data vaults
and accumulating the data into one repository may increase the risk that data will be
compromised, said P9. The success of a predictive model depends on access to data. However, if
companies do not trust that auditing firms can protect their data, then they may not provide
access to the data. Maintaining the security of significant amounts of confidential data provided
by clients is a concern for auditors (Alles & Gray, 2016). Joshi and Marthandan (2018)
recognized that this is a challenge for auditors because they need data to perform audits. Shukla
and Mattar (2019) believed that the concern about data security is preventing the adoption of big
data analytics.

**Theme 2: Data Mining Techniques**

The theme of data mining techniques focused on emerging technology that has the
potential to provide auditors and fraud investigators with a significant advantage in the fight
against accounting fraud. The literature has described the new technology using terminology that
seems more in line with a science fiction movie than with accounting and auditing procedures.
Auditors using artificial intelligence to mine for accounting fraud is a major leap forward,
especially for a conservative industry that has been slow to change (Richins et al., 2017). The
research abounds with statements of how data mining techniques outperform traditional audit
techniques or how statistical analysis pales in comparison to machine learning (Appelbaum et al.,

According to the findings from Omar et al. (2017), the logistic regression model is not as
good as a simple decision aid when it comes to classifying fraudulent and nonfraudulent
transactions. Moepya et al. (2016) believed that data mining has an advantage over statistical analysis because data mining methods do not impose arbitrary assumptions. Further, Yeh et al. (2016) suggested that statistical analysis is restricted due to its one-dimensional linear view. Mongwe and Malan (2020) added that data mining has an advantage over statistical methods because data mining does not need to make assumptions about the statistical distribution of the data. S. Chen (2016) explained that the error rate is relatively high for conventional statistical methods and that using data mining techniques would provide more accurate results as opposed to performing a conventional regression analysis. The research also suggested that traditional methods cannot handle big data, especially the vast amounts of unstructured data. Further, these traditional methods require a significant amount of time and are expensive to execute (Mark et al., 2019). Gepp et al. (2018) stated that traditional regression models do not fit big data because the models are highly structured.

Ultimately, the traditional audit procedures generate far too many false positives requiring manual review and resolution, which is very expensive and time consuming. To limit the number of false positives, more sophisticated techniques began to be deployed. Research has begun to incorporate new technology such artificial intelligence, machine learning, deep learning, and text mining to improve fraud detection accuracy (Appelbaum et al., 2018). Sun (2019) explained that artificial intelligence is a general category of computerized activity of which machine learning, deep learning, and text mining are subsets. For this theme, the researcher focused on machine learning and text mining techniques. Deep learning is outside the scope of this research primarily because of the advanced nature of the technology. Deep learning is based on deep neural networks, which makes it extremely difficult for auditors and fraud investigator to explain the output to clients and regulators.
Machine Learning. Kokina and Davenport (2017) believed that machine learning adds another level of intelligence when it comes to analyzing numbers. Because machine learning can digest vast amounts of data, its predictions are more accurate compared to regression analysis, according to Jan (2018). Machine learning is already working to automate mathematical modeling outside the accounting arena (Kokina & Davenport, 2017). Research has shown that the use of machine learning can help extract greater insights from data analytics (Dbouk & Zaarour, 2017). However, even with these accolades, machine learning is no panacea for detecting accounting fraud. The research has made it very clear that there is no one best machine learning model. There is no consistent or conclusive winner when it comes to the best machine learning model for detecting accounting fraud (Qian & Liu, 2018). M. Ahmed et al. (2016) concluded that the best fraud detection model has not been found because of the lack of labeled data.

Mongwe and Malan (2020) appropriately summed up the prevailing thought with respect to the best machine learning model:

From the studies surveyed, it was found that there is no overall best method, with different methods outperforming on different data sets. This result is in line with the No-Free-Lunch theorem for learning algorithms that states that no completely general-purpose learning algorithm can exist, so one can assume that there exists no best machine learning algorithm for all problem instances. (p. 102)

Finally, P7 added that there is no best machine learning technique. There are a lot of different techniques, such as decision tree, random forest, self-organizing maps, and back propagations. Overall, there are many kinds of models. P7 suggested that each model can be treated like an opinion from an audit team member, and like opinions, model predictions are not
always correct. Even information from an Excel spreadsheet can be just another opinion but may 
not be weighted very heavily, said P7. Machine learning consists of supervised learning, 
unsupervised learning, and semi-supervised learning (Gepp et al., 2018; Margagliotti & Bollé, 
2019). The main difference among supervised, unsupervised, and semi-supervised learning 
relates to the labeling of training data sets.

**Supervised Machine Learning.** Because supervised learning requires analysts to train 
and test models manually, the models are considered supervised (Lewis & Young, 2019). 
Supervised machine learning requires labeled data sets to train a model. The auditor uses 
historical data and labels the data set as fraudulent or nonfraudulent. The labels are known, and 
the algorithms can learn from the data by using examples provided by the labels. However, 
because supervised learning requires labels, collecting labeled data sets may be challenging, 
limiting the use of the supervised learning technique (Sun & Vasarhelyi, 2018). When it comes 
to accounting fraud, the labels for the data do not have to be that complicated—fraud and non-
fraud or proper and improper. However, it is not always easy to obtain labeled data, as discussed 
above. P6 added that because supervised learning uses known fraud to label data, this technique 
is not so good at detecting unknown frauds. An unsupervised learning approach using clustering 
is better at discovering new frauds, said P6. P6 stated that, in the end, a “combination of a 
supervised and unsupervised approach tends to be more powerful.”

**Classification.** The research showed that a supervised machine learning approach using 
classification techniques was the most popular model for accounting fraud detection (Amani & 
Fadlalla, 2017; Bao et al., 2020; Byungdae & Yongmoo, 2020; Y. Chen et al., 2017; 
Mohammadi et al., 2020; Mongwe & Malan, 2020; Omidi et al., 2019; West & Bhattacharya, 
2016a). In basic terms, fraud detection boils down to classifying a company or transaction as
fraudulent or nonfraudulent. In other words, the classification problem is binary—fraud or not fraud (Bao et al., 2020). Because fraud detection fits neatly into a binary classification problem, the supervised classification model is one of the most popular models used for accounting fraud detection. Mongwe and Malan (2020) found that the most common classification methods included:

- support vector machines (svm),
- decision trees, including
  - CART,
  - CHAID,
  - C4.5, and
  - C5.0,
- random forests,
- genetic algorithm (GA),
- artificial neural networks (ANN),
- Naïve Bayes, and
- Bayesian belief networks.

**Unsupervised Machine Learning.** Given the significant challenge of obtaining labeled data for accounting fraud detection, unsupervised machine learning may be the answer. However, the research involving this technique is in the beginning stages compared to the research relating to supervised classification methods. Most of the accounting fraud detection research related to supervised classification models and only a handful of the research projects related to unsupervised clustering methods (Omidi et al., 2019). Unlike supervised learning, unsupervised learning has a better chance of identifying new frauds. Because unsupervised
clustering does not require labels based on past frauds, this technique has a better chance of identifying new fraud schemes as compared to supervised classification techniques. P1 explained that if investigators used unsupervised learning techniques, they would have a better chance of keeping up with the fraudsters because unsupervised learning does not rely on data labeled using known frauds.

**Clustering.** Mongwe and Malan (2020) indicated that clustering techniques are used with unsupervised machine learning. As mentioned earlier, clustering is helpful when data is not labeled. If one or two clusters are identified, then clustering is not that difficult. However, when thousands of clusters are generated, auditors will likely have a difficult time telling which clusters have good transactions and which clusters have bad transactions. P10 stated that some researchers believe that unsupervised machine learning may be the answer to the labeling problem. However, this type of machine learning uses clustering to identify anomalies. When clustering techniques produce 10,000 clusters that need to be reviewed to determine which clusters represent high-risk groups, then it may not be worth it at some point, said P8.

P10 stated that “clustering is that dark art…how do you possibly know how many clusters you are supposed to have?,” P10 asked. Further, P10 suggested that there is a curse of dimensionality with respect to clustering. “If you have billions of data points…there is no way your clusters are going to be meaningful,” said P10. Clusters by themselves do not mean anything. “You still need a human to investigate them,” said P10. Finally, P9 explained that even though the clustering techniques do not require labeled data like a classification technique, clustering still requires the clusters to be manually reviewed to determine which ones represent fraud, so in essence, there still exists a labeling problem.
**Semi-Supervised Machine Learning.** Semi-supervised learning, which exists between supervised and unsupervised learning since some of the data is labeled while a much more significant portion of the data is unlabeled. In this situation, the goal is to train the classifier using labeled and unlabeled data (Abdallah et al., 2016). P1 believed that the semi-supervised approach would be great when labeled data is not available and an analyst has “a good gut feeling of which transactions could potentially by fraudulent.” Further, P1 explained that the investigator could use clustering to identify high risk cluster groups. Once the clustered groups are labeled, the labeled data could be used to find other potential fraud. However, as discussed above, it may be difficult to discern which clusters have high risk transactions, especially if there are thousands of clusters.

**Text Mining and Analysis.** The research clearly shows that combining the analysis of unstructured data with structured data can improve the fraud detection process (Boskou et al., 2019; Dong et al., 2018; Joshi & Marthandan, 2018; Richins et al., 2017; Sun, 2019). In exploring research question two, how to use financial, nonfinancial, textual data to identify red flag warning signs better, the researcher found data mining techniques that allow auditors and investigators to extract valuable information from unstructured data to highlight warning signals. As discussed above, auditors and investigators use mostly structured financial data, which represents potentially only 20% of the available data. By only using structured data, auditors and investigators are looking for fraud through a keyhole instead of opening the door to get the full view of the potential fraud. The upside to using unstructured data for clues of accounting fraud is enormous and well worth the effort required to analyze the data.

There are many reasons why auditors and investigators have not used unstructured data in the past. Two main reasons include limited access to unstructured data and the lack of systems to
handle and store the unstructured data (Joshi & Marthandan, 2018). With the proliferation of unstructured data, in part because of social media and the internet, auditors can now readily access unstructured data. Further, with advancements in technology, unstructured data can be processed in a meaningful way to provide auditors with insights that did not exist before (Alireza, 2019). However, the challenges of working with unstructured data do not end because of the advancements in technology or availability of data. Extracting knowledge from unstructured data is still difficult and requires some understanding to address the challenges of working with unstructured data. First, the unstructured data must be formatted so that it can be analyzed. In its raw textual form, computers cannot understand the nuances of human language, also known as natural language. Because text is created by humans, there is quite a bit of ambiguity in the natural language that humans can discern but computers cannot. These ambiguities are lexical, syntactic, and referential. Without addressing these ambiguities, computers have no chance of extracting meaningful data from natural language (Guo et al., 2016; Throckmorton et al., 2015).

Text mining has evolved over the years, and its supporting technology and techniques continue to evolve. Lewis and Young (2019) found that past research relied mostly on simple text mining techniques using keyword searches, word counts, and dictionaries. P3 noted that if several words have the same meaning, such as “photograph” and “picture,” a simple text mining approach will not capture the nuances of these two words, and a more sophisticated text mining tool such as NLP is required. The use of NLP is helping move from relatively simple approaches to more sophisticated techniques. NLP is commonly used to process unstructured data so that computers can understand language, but more importantly so that models can process unstructured data like they do with structured data (Goel & Uzuner, 2016). NLP is considered
part of artificial intelligence and is used to interpret natural language both in written and spoken form (Boskou et al., 2019).

Part of the data processing for unstructured data includes splitting the words into tokens to be viewed individually. Analyzing words without considering word sequence or semantics is referred to as a bag-of-words approach, since the words are essentially scooped up in a bag and viewed without the context or word sequence (Loughran & McDonald, 2016). By combining bag-of-words and generic dictionaries, positive and negative words can be identified, which can help discover the tone of unstructured data.

Lewis and Young (2019) provided a clear explanation of how to determine the tone of a document:

Consider, for example, the assessment of a document’s tone. One first counts the number of positive or negative words based on a specific dictionary and then typically scales the counts to create word proportions. Documents with a relatively high frequency of positive words are considered optimistic and likewise those with a relatively high percentage of negative words are labeled pessimistic. A commonly used measure of tone is the difference between the scaled positive and negative tone words. (p. 598)

By not capturing the order and sequence of the words, bag-of-words cannot determine the context or semantics of words (Sun & Vasarhelyi, 2018). If context is added to the data, then the value of the data can be increased and provide much more in terms of data analysis, said P8 and P10. Additional processing of the natural language includes parts-of-speech tagging, stemming, and removing stop words (Boskou et al., 2019; Dong et al., 2018; Hájek, 2018; Sun & Vasarhelyi, 2018). P9 found that NLP and text analysis worked best for detecting accounting fraud by mining memo fields or the unstructured portions of a transaction record, but P9 also
added that text mining was more than just keyword searches. Further, P9 believed that text mining is much better at detecting accounting fraud than using a rules-based system with structured data. With respect to bribery and corruption and looking at improper payments, P5 believed that “the number one indicator of corrupt intent is what people put in the free text field of a journal entry.” P5 added that “the good thing about accountants is they have to track things, even when they’re committing fraud. Mining free text fields is one of the best ways to perform data mining for fraud, or at least improper payments.”

P10 pointed out another exciting area where NLP could provide additional intelligence for auditors and investigators. Information extraction using NLP can extract facts from documents to be cross-referenced with facts from other documents. P10 stated that auditors “could look at whether all facts across texts and financials corroborate each other or whether there are differences which could give you confidence in a way to kind of pinpoint fraud.”

Ultimately, great strides have been made over the last few years to improve the technology and techniques to extract information from unstructured data and identify warning signals hidden in financial statements. Hájek and Henriques (2017) found that fraudulent financial statements tend to have both positive and negative sentiment. Further, Goel and Uzuner (2016) noted that the writing styles in fraudulent and nonfraudulent annual reports are different. By discovering the subtle clues in the language used in corporate disclosures, auditors and investigators will be in a better position to identify the warning signals hiding in plain sight.

**Algorithms.** Algorithms provide the primary engines for the artificial intelligence and machine learning techniques to operate. Algorithms can handle the big data used for the models (Margagliotti & Bollé, 2019). The research discussed many algorithms, and more than one
algorithm was recommended to have the best chance of detecting fraud. Some of the algorithms provided by the participants and literature included:

- logistic regression
- random forest
- neural networks
- recurrent neural network
- convolution neural networks
- decision tree
- C5 decision tree
- Naïve Bayes
- Bayesian interference
- Support Vector Machine
- self-organizing maps
- gradient boosting

This list is just a small number of the algorithms noted by the researcher from the research gathered. When asked to provide the best algorithm for accounting fraud detection, the participants were unanimous in their response that there was no one best algorithm. P6 stated that “it really depends on, again, the quality of the data that you have and the problem you’re trying to solve.” P3 noted that analysts use algorithms such as random forests and regularized logistic regressions because of their ease of use. Ultimately, the data will help you select the most useful algorithm to use, said P3. Further, P3 stated that essentially a model boils down to two main ingredients—data and features selected. Once the two ingredients are added to the model, it is just a matter of selecting the algorithm that performs the best. If a random forest has an accuracy
of 80% and a logistic regression has an accuracy of 60%, then the random forest algorithm would be the best choice. There will be some iterative work like this to find the best algorithm to use, given the outcome you are looking for, said P3. Even though a vast number of algorithms exist, the participants stated that they only really use about five algorithms because those five work. The literature supported the participants regarding the best algorithm to use. Researchers agreed that there was no single algorithm that performed the best for each fraud scenario and data set (Dbouk & Zaarour, 2017; Moepya et al., 2016).

**Theme 3: The Model**

Beyond data and data mining techniques, successful predictive models require balanced data sets, the right features, proper model training, maintenance, and growth. Failure to adequately address each one of these items could result in a poorly performing model that provides output that is no better than an educated guess by the auditor or fraud investigator (RQ1). Theme 3 focused on building and maintaining machine learning models and the challenges users can expect to encounter as they build predictive models. The concepts in Theme 3 provided the researcher with additional information to understand better why successful predictive models are difficult to build and require quite a bit of upfront work, time, and patience to achieve accurate predictions (RQ1).

**Imbalanced Data Sets.** Most of the companies used in the research relating to accounting fraud detection models were nonfraudulent, while a minority of the companies used were fraudulent. Perols et al. (2017) referred to this problem as the “Needle in a Haystack Problem” because fraudulent companies are relatively rare compared to nonfraudulent companies (Mohammadi et al., 2020). Majority and minority classes create difficulties for algorithms to accurately classify fraud and non-fraud (Abdallah et al., 2016; Dutta et al., 2017;
Oussous et al., 2018). P1 recommended developing a strategy to handle imbalanced data sets and stated:

You’re not expecting a huge portion of fraud in a very large lake of data. And so how do you deal with something like that. It’s a very challenging problem to tackle because in any predictive analytics, whether it’s machine learning, deep learning, or simple regression type algorithms, if you do not have a strategy for handling how the model is going to learn the characteristics of what a fraudulent activity looks like it will get swamped out by everything else.

One way to address this problem is to create a balanced data set from imbalanced data. A balanced data set contains an equal number of fraudulent and nonfraudulent companies. Researchers developed several techniques to create balanced data sets including methods such as undersampling and oversampling (Bao et al., 2020; Omidi et al., 2019). With an undersampling approach, researchers include the same number of fraudulent and nonfraudulent companies to train the model, regardless of the actual number of nonfraudulent companies in the data set collected. However, the research has indicated that undersampling is the least effective method for handling imbalanced data sets because the method limits the amount of data used to train a model (West & Bhattacharya, 2016a). For example, if a data set contained 5,000 companies, of which 1,000 represented the minority class and 4,000 represented the majority class, only 1,000 of the nonfraudulent companies would be included in the data set and used to train model with undersampling. Therefore, in the end, only 2,000 of the total 5,000 companies would be used to train the model. Using less data is counter to what the research recommends. When training predictive models, more data is needed, not less data. Therefore, the undersampling technique is believed to be the least effective approach to the imbalanced data set problem (Dutta et al., 2017).
On the other hand, oversampling has its drawbacks as well. By oversampling the minority class, the researcher risks overfitting the model, which can bias the model. One oversampling technique splits the majority class into smaller data sets and uses the same fraudulent instances in each of the newly created data sets. To avoid overfitting a model using the oversampling technique, researchers developed the SMOTE. With SMOTE, instead of using the same fraudulent instances to pair with the nonfraudulent instances, additional fraudulent instances are synthetically created using the k-nearest neighbors technique (Abdallah et al., 2016; Dutta et al., 2017). However, Byungdae and Yongmoo (2020) noted a drawback in the SMOTE technique as it does not use real data to train models. Overall, imbalanced data sets are just one more challenge that auditors and fraud investigators must overcome as they build predictive models that accurately and consistently detect accounting fraud (RQ1).

Features. A big part of any successful predictive model involves selecting the best features to use in a model. P1 suggested that certain features have very little predictive value, and therefore, features must be selected carefully and with purpose. Gupta and Mehta (2020) stated that features must be selected before any data is fed into a model. Like machine learning algorithms, there are no best features that should always be included in accounting fraud detection models, said Jofre and Gerlach (2018). A feature selection technique can help select the best features to use in models (X. Tang et al., 2018). If too many features are used, this may create too much noise for the model to make accurate predictions. This scenario is referred to as the curse of dimensionality. Fraud prediction can be hampered by the curse of dimensionality (Gupta & Mehta, 2020; Mohammadi et al., 2020; Perols et al., 2017). A high number of features can increase the complexity and the execution time of the model, ultimately reducing the
effectiveness of the model’s output and classification accuracy (Boskou et al., 2019; Gupta & Mehta, 2020).

P2 warned that users must be careful to limit the features that are included because too many features might cause too much noise and limit the accuracy of the model. Some features may not provide valuable information. P2 stated that too many variables may make things worse and make isolating the most important features to use in predicting fraud more difficult (Mohammadi et al., 2020). Further, P2 suggested that when data comes in as a feature, users must be thoughtful about how to identify the right features to use. For example, the time and date of a transaction might not be as important as the interval between the transactions. By understanding the data and thinking through the various scenarios, auditors can build better models, said P2. Also, P8 suggested that investigators not only select features to use but also understand the context of the feature. For example, a date by itself may not be that telling, but the day of the week of the date could add context to the data and be more valuable in terms of making accurate predictions. To avoid working with too many features, Abdallah et al. (2016) recommend deploying dimensionality reduction strategies such as data compression, feature selection, and feature construction. Choosing which features to use can be challenging because it is not always easy or intuitive to identify relevant and irrelevant variables (Badih et al., 2019).

Model Training. Before any data is fed into a model, an analyst must execute many steps, all of which have inherent challenges. P5 stated that a key characteristic of a successful predictive model is adequate seed data to train the model. “If you can’t tag the transaction as relevant, not relevant, fraud or not fraud, whatever it is, if you don’t have a way to tag, then you can’t train the model,” said P5. Further, P1 stated that knowing “how to train a model and when
to stop training a model is just as important as the quality of the data.” P1 explained that there is “a lot of art to the data science” of training a model. P1 stated:

One of the most common errors with people who are just entering into predictive analytics or predictive modeling is the overfitting problem. When you’re creating a model and you’re training it on a subset of your data, how do you know if that model is gotten to 100% accuracy because it’s learned the exact characteristics of every single one of your rows? And how you go about analyzing that problem with your loss curves and your accuracy curves is incredibly important because if you overfit then that model will never ever predict any real-world scenario with any accuracy.

Finally, Jofre and Gerlach (2018) explained the difficulty in training a model to classify fraud and non-fraud instances. They stated that “learning from these rare events is a very challenging task given the small amount of observations available to train predictive models, hence especially difficult to further discriminate between fraudulent and nonfraudulent instances” (p. 8).

Model Bias. Believing that predictive models cannot be biased because they are run by computers using algorithms is a common mistake made by model designers and builders (Veale & Binns, 2017). However, this thinking is shortsighted since a significant amount of human judgment goes into selecting the data sets, algorithms, and features used to build the model. These judgments may inadvertently create bias in a model that no machine learning algorithm can eliminate (Huerta & Jensen, 2017). Further, P9 stated that “accidentally encoding bias into the model [is] going to skew your output.” Researchers found that studies might have been biased because only publicly available data was used and did not include data from private firms, which is difficult or impossible to collect (Omar et al., 2017). Finally, P6 stated:
I think there is a need to proactively address bias in modeling and be thoughtful about that. I think we’ve seen a lot of examples where people have built models and unintentionally introduced bias based on things like the training data set was biased. There is a need to proactively address bias in modeling and be thoughtful.

**Concept Drift.** Although it is impossible for auditors to also be data scientists, they must understand the data used by a model and how models change over time. There may be seasonality in the data, which may cause a model to drift or degrade. Models cannot be implemented and left alone for years at a time. As data changes, the model changes, and without proper model maintenance, the model will be useless (Abdallah et al., 2016). P6 provided the following example to explain concept drift. During the COVID-19 pandemic, the buying habits migrated from in-person shopping to online shopping. Without updating the predictive model to reflect the change in buying habits, models began to produce incorrect predictions. Models started flagging transactions as fraudulent because a person’s buying habits changed from never purchasing items online to always purchasing items online. As the data changed, the model did not change. Given the pace of change, predictive models cannot be developed and forgotten. They need to be maintained. P6 further explained that concept drift can limit predictive models. “A lot of people will implement a model, and then they’re done and then let it run in production for three years,” said P6. Over time, models degrade without being maintained. As the data changes, processes need to change as well, explained P6.

**Model Performance.** According to P6, the biggest complaint about predictive models is the high false positive rate that some models generate. False positives are one of the main limitations of predictive models. Some organizations do not understand how modeling works and have a false sense of security, believing that models are 100% accurate, which models are not,
said P6. Essentially, model predictions are just reasonably good guesses. Therefore, it is important to measure the accuracy of a predictive model. P9 believed that a high false positive rate means that the right sequence or data points have not been identified and a better-defined outcome is needed. There must be some other data point that can help train the model to get a better outcome. To understand how well a model is performing, users must measure the output. For an accounting fraud detection model, where the two classes are fraud and not fraud, true represents fraud and negative represents not fraud. Therefore, model output can be measured in terms of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP refers to a models’ sensitivity, and TN refers to a model’s specificity. When the sensitivity of a model is increased, the specificity of a model is decreased. Conversely, when the sensitivity of a model is decreased, the specificity of a model is increased. Dutta et al. (2017) provided five measurements to measure the performance of a binary classifier: (1) sensitivity, (2) false positive rate, (3) accuracy, (4) precision, and (5) specificity. Dutta et al. illustrated the performance measurements as follows (Equations 1, 2, 3, 4, and 5):

\[
\text{Recall or sensitivity} = \text{TP} = \frac{TP}{TP+FN} \quad (1)
\]

\[
\text{False positive rate} = \frac{FP}{FP+FN} \quad (2)
\]

\[
\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)
\]

\[
\text{Precision} = \frac{TP}{TP+FP} \quad (4)
\]

\[
\text{Specificity} = \text{TN} = \frac{TN}{FP+TN} = 1 - \text{false positive rate} \quad (5)
\]

The Receiver Operating Characteristics (ROC) curve and area under the curve (AUC) can also be used to evaluate the model output (Dutta et al., 2017). J. Han et al. (2012) described the ROC curve as follows:
For a two-class problem, an ROC curve allows us to visualize the trade-off between the rate at which the model can accurately recognize positive cases versus the rate at which it mistakenly identifies negative cases as positive for different portions of the test set. Any increase in TPR [true positive rate] occurs at the cost of an increase in FPR [false positive rate]. The area under the ROC curve is a measure of the accuracy of the model. (p. 374)

Depending on the objective of the model, the sensitivity or specificity of a model may need to be adjusted. If finding every fraudulent transaction is more important, then the model’s sensitivity will be increased. If, on the other hand, limiting false positives is more important, then the model’s specificity will be increased (Bao et al., 2020; Dbouk & Zaarour, 2017; Jofre & Gerlach, 2018). Tuning a model may generate a new set of headaches for the auditor or fraud investigator. P9 explained that sometimes tweaking a model in one area causes alerts to go from 100 down to two. When that happens, the model must be reevaluated to understand what rules must be adjusted to make the model right. This will take time, which auditors and investigators are not usually afforded. By continuously updating the model based on feedback from the results of the model, the model gets smarter. Unfortunately, this growth takes time and is not practical for external financial auditors or fraud investigators, who are working under tight deadlines, for the most part.

**Model Growth.** A successful predictive model must be able to grow. Without growth, a model will not be able to improve its accuracy rate (Brown-Liburd & Vasarhelyi, 2015). Models that are continuously learning will produce better results. P6 believed it is important to understand how the model will operate. P6 posed important questions such as:

We’ve done the math and we can score things. But what do I do with the results? Do I put it in a spreadsheet? Do I send them an email? Do we have a system that tracks them? How
do I record? The evaluation of whether or not this was a false positive. Is there a case management system? You need that feedback loop in order for this thing to self-learn, and putting that into production is actually a significant effort.

It takes time to put a model into production. “A lot of models get built and then just put on the shelf, because nobody thought about actually how to implement them,” said P6. Further, P7 believed that models must have the ability to grow, because criminals are creative. According to P7, if models do not grow, then new frauds will not be detected. Successful predictive models continuously learn from data and results (Bhardwaj & Gupta, 2016). When there is a false positive result, the model must be able to learn new things to improve the results. This might include changing the weights or scoring to improve results, said P7.

**Trusting Models.** Trusting a model too much is a top mistake an organization can make when using predictive models, said P1. According to P1, companies “think that whatever the model predicts the model is correct.” Further, P7 explained that users believe that whatever a predictive model tells them is correct. “People think that predictive models tell you the truth,” said P7. However, users may not understand how predictive models work; “they may not understand how to ask their questions to the AI [artificial intelligence]. So, if you ask nonsense questions, surely the AI would give you nonsense answers,” said P7. Conversely, users may not rely on the results of a model because they do not trust the model. P6 provided an example where a company implemented an analytical model to score medical clinics. The output of the model showed that a clinic had a very high fraud rate. However, the company did not trust the model and did not believe that fraud was occurring, so it decided not to perform an investigation. However, one week after the company declined to investigate the clinic, “the owner of the clinic and all the employees were arrested for running a medical mill,” said P6. As it turns out, the
model was accurate, but the company did not trust the model. Finally, Huerta and Jensen (2017) explained that it is a balancing act between over- and under-reliance on predictive models. Huerta and Jensen state the following:

Over-reliance on Big Data is also detrimental, since accepting information without challenging it can lead to mistakes. The panelists indicated that, as with any tool to support decision making, the effectiveness of Big Data relies on its actual use. When accountants do not rely on the results produced by the system—under-reliance—the insights that could be gained are lost. On the other hand, when accountants blindly accept the results produced by the system—over-reliance—the healthy skepticism needed to evaluate information is lost. Relying on a system is a balancing act; both under- and over-reliance pose risks in the use of Big Data. (p. 111)

Theme 4: Human Agency

When thinking about artificial intelligence, machine learning, or predictive modeling, one does not automatically think humans. Without a deep understanding of technology, a layperson may believe that humans have very little interaction with the technology. However, they would be wrong. The research showed that quite a bit of human agency is required to build a successful predictive model (Appelbaum et al., 2017; Huerta & Jensen, 2017; Issa et al., 2016; Sutton et al., 2016). Unlike what is represented in science fiction novels and movies, computers do not have the cognitive abilities to think like humans or apply common sense to certain situations. If models are fed biased data, the model generates results that are biased. Overall, technology can complement human experts but not replace them, at least for now (Richins et al., 2017). Theme 4 focused on the human agency and the human interaction required to build successful predictive models.
**Professional Skepticism and Judgment.** Professional standards require auditors and investigators to use professional skepticism and judgment as they perform audits and fraud investigations. These same skills are required to enable auditors and fraud investigators to objectively collect and analyze data (Kaban, 2020). Applying professional skepticism requires auditors to be curious enough to understand how a model works and how a model scores transactions. Auditors must approach audits with a questioning mind even when using technology to analyze data. Without questioning how a model works, auditors will not be able to explain the model’s output or why certain transactions were flagged for further review. Without professional skepticism, auditors may rely too much on the results of the predictive models (Rose et al., 2017). As discussed above, there is a balancing act when it comes to trusting models. P1 explained that “if you don’t have enough skepticism about what the model is saying, you can fall into a deadly trap of making incorrect judgment calls.” P9 also believed that auditors and fraud investigators must exercise professional skepticism and judgment when using predictive models. “I think it definitely impacts the use of the model. At the end of the day, someone has to make a call as to whether or not this end result is useful,” said P9.

When using rules-based detection systems, auditors must possess professional judgment so they can properly adjust scoring and weights used in predictive models. Auditors must decide which rules to use to have the best chance of flagging fraud. Also, auditors must decide what should be considered a high-risk transaction and what scores should be considered too high. Risk scores that are too low will generate too many transactions for an auditor to review (Baader & Krcmar, 2018). P3 stated that the challenge becomes determining the cutoff for high-risk transactions. Audit resources will have to be expended to look more deeply at the transactions that were flagged by the model to determine whether the transactions represent fraud or whether
they were false positives. The more transactions that need to be reviewed, the greater the cost of
the audit engagement, which audit firms cannot not afford, since audits are usually fixed cost
with little margin for extra work. Further, P4 did not believe that auditors can trust models if they
do not possess a questioning mind. P6 believed that a good data analyst is curious and aware of
how data impacts the model. Using professional judgment, auditors will be able to do a better job
of avoiding bias in a model (Kokina & Davenport, 2017).

As discussed above, the quality of the data matters when it comes to building successful
predictive models, so an auditor must have a healthy skepticism of the data to ensure that
garbage is not going into the model. P10 believed that there is a misconception about being
skeptical of the model and stated:

You don’t need to be skeptical of the model; you need to be skeptical of the data that’s
feeding the model. Models are literally just statistical treatments. They’re really just a
function of the data you feed into them. You shouldn’t be skeptical of a logistic regression
model because that’s fully explainable.

No model provides absolute assurance of fraud detection. P7 stated that models provide
opinions as to whether transactions or companies are fraudulent, and it is up to the auditor to
make the final decision. Sun (2019) believed that the model makes predictions that the auditors
should evaluate using professional judgment. Finally, Richins et al. (2017) explained that big
data analyses can complement auditors’ professional judgment but not eliminate it. Richins et al.
also stated that “it is difficult to automate holistic assessments of management’s motivation,
opportunity, and ability to rationalize financial statement fraud, as these assessments require
social intelligence” (p. 73).
Data Analytics, Statistical, and Technology Skills. Beyond professional judgment and skepticism, auditors and fraud investigators must also possess technical skills including data analytics, statistics, accounting, and finance. Auditors with business and data science background do not usually exist, but if they do exist, they are rare. Several of the participants referred to an individual with business or financial experience and data science skills as a “unicorn.” Regardless of the tools an auditor or investigator uses, they will always need basic data analysis skills and be comfortable working with data and technology (Richins et al., 2017). Even basic statistics such as regression analysis and linear modeling should be a part of an auditor’s skill set when performing data analysis using technology (Shukla & Mattar, 2019). Again, this will help the auditor or investigator explain the output from predictive models. This is not to say that auditors and investigators should be data scientists. They should not be, but they must be familiar with data and technology and possess basic data analysis skills (Shukla & Mattar, 2019). These skills should not be taught on the job but in the classroom at the higher education level. To accomplish this, curriculums at colleges and universities will need to be updated to ensure this need is met (Al-Htaybat et al., 2018). Further, since the models use algorithms, which are based on a foundation of statistical principles, auditors and investigators will benefit from understanding basic statistics.

However, some regulators are worried that accounting firms will move too far towards data science and hire data scientists to perform data analytics. The regulators worry because it is likely that the data scientists will not have the same business acumen that auditors possess. The regulators are concerned that the audit quality may suffer if auditors are replaced by data scientist with no business experience (Earley, 2015). Richins et al. (2017) believed that to take
full advantage of data analytics, auditors must possess both data analytic skills and business acumen.

P4 believed that it is important for auditors and fraud investigators to have a data technology background when working with predictive models. Auditors “have to have some comfort working with data if you’re going to be responsible for making the data do something,” said P4. Further, P4 did not believe auditors need to be data scientists, but they do need to understand the difference between an alpha-numeric and a number or an integer and a character field. Having some basic level of data and technology skills will help the auditor explain the results achieved, explained P4. Even with technology, auditors and fraud investigators must be able to analyze data. P8 stated that “you still have to understand the fundamentals of what the data is trying to tell you.” Further, P9 believed that some of the challenges and limitations relating to models may be addressed with better education of the workforce.

A technology background is helpful for auditors and investigators, but a mathematical background would also be helpful to understand the complex math modeling. An understanding of math would help auditors and fraud investigators to explain the model results. Given the mathematics and statistics used in modeling, auditors and investigators may benefit from having these skills when working with predictive models (Appelbaum et al., 2018). Ultimately, auditors and fraud investigators using predictive models must have the appropriate data analysis and technology skills to avoid making mistakes (Dbouk & Zaarour, 2017). P7 strongly believed that to overcome the challenges with building and using predictive models, organizations must continuously educate their auditors and fraud investigators, as well as executives so they understand that data analytics is a good thing.
**Ethical AI.** One key discussion point that the participants brought up was ethical AI. Predictive models, especially ones using machine learning techniques, learn from the data fed to the model. However, if the model is fed biased data, then the model will generate a biased output, which companies will then use to make business decisions such as who qualifies for a loan and who does not. Biased model results can cause organizations to discriminate. Therefore, humans must be involved in the process. P10 stated that ethical AI is currently a key topic and a concern for organizations. P2 stated that the human agency component of predictive modeling is becoming very big. Organizations must understand how their models are “being applied to people and how are these people being affected, positively or negatively,” said P2. Artificial intelligence does not have the capacity to evaluate ethical issues. Therefore, humans must be in the loop to ensure that company decisions comply with regulatory and ethical standards (Deloitte, 2019).

**Theme 5: Approach**

For Theme 5, the researcher focused on the organization rather than the individual. By doing this, the researcher was able to gain a different view of artificial intelligence and predictive modeling. In this theme, the researcher explored how organizations can contribute to the success or failure of predictive models. The researcher used the term “Approach” to represent the various organizational facets that must exist to build successful predictive models. Before auditors can analyze data, they must have the appropriate hardware, software, and skills. Organizations must plan for and invest in the infrastructure required to perform data analysis. All of this starts with an organization’s philosophy towards intelligent automation.

**Philosophy.** Over the last 10 years, the push to implement artificial intelligence has increased dramatically, and companies are jumping on the bandwagon to deploy predictive
models. In some cases, companies invest millions of dollars without really thinking through what needs to be done to successfully build and use predictive models. Oussous et al. (2018) suggested that organizations should consider “technological compatibility, deployment complexity, cost, efficiency, performance, reliability, support and security risks” (p. 432). Each one of these items requires time, attention, and reflection to properly implement. However, because it takes time to address these items, some companies may skip a thoughtful evaluation and consideration of these items. By not understanding what is required to properly develop predictive models, organizations will likely fail in their attempts to successfully use artificial intelligence. Organizations must avoid the temptation to adopt artificial intelligence and predictive models before they are ready. Instead, companies must have a plan and the right infrastructure in place before they rush into adopting big data analytics (Shukla & Mattar, 2019).

Companies need to have a longer perspective. They need to develop long-term goals. P10 believed that companies are bound to fail when they do not understand the big picture of deploying technology. P10 stated:

One of the major mistakes [made by organizations] is thinking that AI will easily be deployed in a company, and usually AI deployment or AI models fail in companies, and the reason is because of the lack of understanding or technical expertise that would allow the model results to be adopted and trusted. Trust is the biggest blocker.

P10 also noted that the reason why companies cannot trust models is because they have unrealistic expectations about what they are deploying. When building models or a data analysis system to detect accounting fraud, the participants agreed that simpler is better. Simplicity is part of explainability, which is discussed in Theme 6 below.
Another point made by the participants is that companies should take smaller steps to implement predictive models and artificial intelligence. This way, if the project fails, it fails in a much smaller way. Small failures may prevent companies from being too scared or fearful to try again. The participants suggested that companies should not try to get too sophisticated too fast. They recommended keeping the model simple in the beginning. Companies should concentrate on the low-hanging fruit and small wins and make the business case for bigger projects moving forward, said P6. This is easier said than done because of the current push to develop and implement artificial intelligence and predictive models right away, especially for larger firms (Lowe et al., 2018). P6 suggested that companies should work on achieving small wins.

Companies are spending too much time, effort, and money setting up centers of excellence and hiring teams and buying technology, but they really should focus on getting “successful business outcomes, and then it makes it much easier to build a business case to expand and go forward,” said P6.

Because big data projects are disruptive, organizations require top-level executive commitment and leadership (Alles & Gray, 2016). To successfully deploy new technology such as artificial intelligence and predictive models, the corporate culture will likely need to be changed to handle the new processes. The entire company must be on board with the new initiative; otherwise, the big data projects will ultimately fail. P5 pointed out that one major challenge for users of predictive models is getting access to technology and the resources to build a predictive model. P5 also stated that firm-wide sponsorships are needed in order to be able to get the necessary resources. P2 suggested having an advocate at the highest level in the company to help facilitate the kind of change required to implement new technology such as artificial intelligence or predictive modeling. Further, P2 explained the need for an advocate:
You need an advocate who understands the business and the data, and the Chief Data Officer needs to be able to build those bridges and to effect change, structural change to where data is collected as matter of course; data is analyzed; decisions are made in a data-driven way as best as possible.

Finally, organizations need a plan so team members know where to start and how to coordinate and integrate a new project. P6 believed that the best thing a company can do is have a “plan to get the results quickly.” Without a plan, the project will likely fail (KPMG Center for Excellence for Data-Driven Technologies [KPMG], 2019). Previously, the researcher discussed how the lack of data, biased data, missing data, and unbalanced data sets can impact the accuracy of predictive models. Here, the lack of a strategic plan to implement and deploy technology can create challenges or roadblocks to building successful predictive models (RQ1). Another challenge for accounting firms is the slow pace of adopting new and innovative technology such as artificial intelligence and predictive modeling (Earley, 2015).

Beyond the challenges discussed above, Shukla and Mattar (2019) identified the following 15 barriers of adopting technology:

- poor business case
- financial constraints
- lack of top management commitment
- organizational resistance to change
- legacy system
- complexity of data management
- poor quality of data
- concerns for data security
- legal and ethical challenges
- lack of knowledge sharing
- lack of infrastructure readiness
- lack of skilled labour
- immature technology
- scalability challenges
- risk of system failure (p. 1019)

**Investment.** Companies are spending millions of dollars on intelligent automation which includes artificial intelligence and smart machines (KPMG, 2019). Developing artificial intelligence and predictive models is not cheap. To build good predictive models, organizations must be willing to invest by hiring the right people, such as data scientists and domain experts, as well as purchasing the appropriate hardware and software. P3 stated that there is a big up-front cost to set up a predictive model, but once the model is in place, the maintenance cost will be less. Since artificial intelligence and predictive modeling is expensive, management may need to be convinced that it is worth investing in the new technology. Shukla and Mattar (2019) found that companies that could not develop a convincing business case faced more challenges adopting advanced data analytics projects. Management must be shown the value of undertaking such a monumental project. Alles and Gray (2016) suggested that it may be more difficult to convince audit partners of the need to invest significantly in new technology because partners can calculate the direct impact the extra expense will have on their own pockets based on the firm’s profit distribution formula.

P2 stated that to change an organization’s culture, a company must invest in people by training them or hiring new employees with the appropriate skill sets. P7 also believed that
quality predictive models require a lot of investment to build. Companies must hire people including data scientists who have PhDs and invest in hardware and software, said P7. To change an organization, personnel will need to be retrained or even newly hired. Ultimately, hiring and retraining employees requires a significant investment by organizations. However, financial constraints such as inadequate budgets create an enormous roadblock and prevent companies from making the necessary investments (Shukla & Mattar, 2019). Overall, it takes time and money to build the right team and develop accurate detection models and systems. Boskou et al. (2019) noted that when introducing new data analytic techniques, organizations worry that the costs will be more than the acquired benefit.

**Patience.** When organizations invest vast amounts of money to develop predictive models, they want quick returns on their investments. Unfortunately, building a good model takes time to get the right data and train the model. Organizations must be patient as predictive models are built. One key limitation to developing a good predictive model is management’s impatience (RQ1). Models need time to grow. The significant number of iterative steps required to train a model takes time. If an organization is not patient, they may scrap a predictive model before it has time to yield value. Ultimately, time pressures can prevent models from being accurate and effective. When companies spend millions of dollars on new technology to build models, they expect results sooner rather than later. Unlike the credit card industry, the auditing and investigating communities have not been working with models for the last 20 years, said P9. It will take time to get it right. Further, P2 stated that organizations could build models quickly, but the quality of the model would likely be poor. It takes time to build a quality model that considers ethics, biases, social responsibility, accuracy, and resiliency, and it also takes time to go through multiple iterations to fine-tune the model, said P2. One critical limitation to
developing predictive models is trying to build models too quickly. It is almost impossible to build a quality model quickly because models need to be trained and improved, which takes time. Models need time to grow. However, not too many organizations understand this, which causes difficulties when trying to build high-quality predictive models. P7 believed that organizations must have a longer perspective and have long-term goals (Alles & Gray, 2016; S. Han et al., 2015; Richins et al., 2017; Shukla & Mattar, 2019).

**Teamwork, Organization, and Process.** Many of the participants believe that a team approach is required to build successful predictive models. A team may include a domain expert, a data expert, and a translator, one who can help facilitate the communication between the domain experts and the data experts. At the start of the technology revolution, IT departments oversaw the computing and data analysis functions. However, in today’s environment, the IT department should not be the sole provider of these services, according to P5. Predictive modeling and data analysis are not an IT function. The biggest mistake a company can make is thinking that advanced data analytics is an IT function, said P5. To be successful, audits and fraud investigations should use a team approach that includes a domain expert and a data scientist with an in-depth technology background including mathematics (Richins et al., 2017). Also, P2 did not believe that data analytics belongs in the IT department. P2 stated that the Chief Data Officer needs to help the business groups and ensure that data analytics projects are not impeded but facilitated by the IT department. Further, P2 believed that having the right team is as important as having the right algorithm. P2 explained that without domain experts, the data scientist may not understand the nuances of an industry where regulations are complex.

P7 suggested that a team must consist of team members with different backgrounds who will be required during the many phases of an audit or fraud investigation such as data collection,
data analysis, case management, model building and optimization, and report writing. For each phase of the audit or investigation cycle, different expertise is required. Ultimately, P7 stated that the team should consist of a subject matter expert who understands how to perform a fraud investigation, a data scientist who knows how to build a mathematical model, and an individual to help translate the language between the subject matter expert and the data scientist. The translator should make sure that the team is on the same page. Usually, the auditor thinks that the data experts just need to push a button to get an answer, and the data experts question why the auditors need a particular piece of data. The translator can link the two together, said P2.

Part of a successful model includes having access to all the data. Organizations sometimes compartmentalize their data. In the past, IT departments were the keepers of the data. P9 has seen some IT groups treat data as high-level intelligence with limited access. “No one person has a key to all the vaults,” said P9. No one knows the entire data population. There may be many gatekeepers that have keys to their individual data vaults, said P9. Also, P9 recommended that a process needed to be set up to gain access to the data and then a pipeline must be set up to continuously stream the data. However, having all the company’s data in one place may make the data more vulnerable to cyberattacks. Auditing and fraud investigation firms must have the appropriate data security protocols in place to address any security concerns (Alles & Gray, 2016). Finally, P6 believed that combining data science, auditing, and data analysis skills into one team works best. P6 suggested that these resources sit next to each other to collaborate, even if they do so virtually. Without this type of collaboration, the data scientist will “build a highly accurate model and then they discover that their results” do not have any business relevance, said P6. Further, P9 believed that beyond continuously educating the workforce, more collaboration and teamwork is required to help overcome some of the challenges of building
successful predictive models. Finally, P4 believed predictive models will improve as domain experts work with data experts. P4 explained:

I think as people who are experts in the design and use of predictive models spend more time talking with people who are experts in catching fraud, you will find more and more. You’ll be able to catch more fraud.

**Clear Goals and Objectives.** For predictive models to be successful, clear goals must be set. Organizations must know what they want to do. To develop a successful model, developers must be able to identify what they care about. It is critical to have a really good idea of what the auditor or fraud investigator is looking for. Machine learning algorithms will be different depending on the goals of the investigator (Margagliotti & Bollé, 2019). Auditors must have clear goals throughout the entire data analysis process. There should be clear goals from the start, such as what data should be requested and how will the data be cleansed. P7 stated that it is not enough to have a goal of detecting fraud. Investigators must ask additional questions to develop specific goals. They must understand what kind of fraud they want to detect, what kind of things they want to look for, and which area may have fraud. Without these types of specific insights and clear goals, auditors and investigators will not know the right questions to ask the predictive model or artificial intelligence, said P7. “They [auditors and fraud investigators] may not understand how to ask their questions to the AI. So, if you ask nonsense questions, surely the AI would give you nonsense answers,” explained P7. P10 stated:

It’s basically understanding what the specific problem is that you’re trying to solve. And then using the right menu of solutions to solve that problem. What do the users care about? Do users care more about finding every single instance of fraud?
Further, P10 warned that if the goal is to find every instance of fraud, then they must also take on the extra work of resolving the false positives. Also, by having clear goals and objectives, data analysts may be able to mitigate the risk of bias or nonrepresentative data sets (Margagliotti & Bollé, 2019). Finally, West and Bhattacharya (2016b) noted the importance of completely understanding the accounting fraud problem before attempting to solve such a complex issue.

**Theme 6: Explainable AI**

At the start of the study, the researcher expected to delve into the various predictive models and learn which models were best for accounting fraud detection based purely on the accuracy of the results of the model. However, instead the researcher learned that predictive models should be viewed beyond just the accuracy of the output (Amani & Fadlalla, 2017). The research has shown that if the output of a model cannot be explained, then the model is useless, regardless of how accurate the model is (Lewis & Young, 2019; Margagliotti & Bollé, 2019). The term Explainable AI addresses the need for auditors and fraud investigators to be able to explain the results of predictive models. Throughout the participant interviews, the participants expressed the importance of explainable AI when building predictive models. The ability for auditors and fraud investigators to explain the output from a model was one of the most important concepts noted during this study. If the model output cannot be explained, then the model cannot be trusted. When asked to provide one key characteristic of a successful predictive model, P4 responded with a single-word answer—explainability. P4 pointed out that Google does not need to worry about whether a user understands how the results of a Google search are derived. However, this is not the case for an auditor or a fraud investigator. They must be able to explain their results to audit clients, regulators, or even judges (Margagliotti & Bollé, 2019).
With artificial intelligence and complex predictive models, explainability becomes much harder than a simple rules-based model.

P6 stated that “an underappreciated aspect of modeling is that you have to make sure that the outputs of the model are understandable by the people who are going to act on the results.” Ultimately, the participants recommended keeping models simple and using basic algorithms such as logistic regression, neural networks, decision trees, and Naïve Bayes. Regardless of what algorithm is selected, the auditor or investigator must be able to explain the output; otherwise, the algorithm will be useless. For this reason, model developers may choose a simpler algorithm that has an accuracy rate of 80% compared to a much more complex algorithm that has an 85% accuracy rate. In other words, it is better to use a less complex algorithm such as logistic regression to be able to explain the model versus using a much more complex decision tree algorithm, which would be much more difficult to explain, said P6. The less complicated algorithm can be explained, and the model can be trusted, whereas the more complicated algorithm cannot be readily explained, and the model cannot be trusted. According to P10, “the big challenge is how do you get highly predictive results and also explain them.” P10 further explained that there is a “tradeoff between having explainable models and predictive models and so auditors and businesses in general are erring on the side of lower predictive accuracy, so that they can explain things.”

**Black Box.** The more complex algorithms are considered black boxes because it is very difficult to follow how an algorithm processes the data from the beginning to the end, causing a lack of transparency. The difficulty to interpret the results of models that use machine learning techniques may impact the willingness of the audit and fraud investigation communities to adopt predictive models (Mongwe & Malan, 2020). The processing performed by the more complex
algorithms occurs in a black box, and most of the time auditors cannot explain how the algorithm works. Without being able to properly explain how an algorithm works, end users of the model will not trust the results, rendering the model useless (Sun, 2019). The participants stressed the importance of understanding the model and the output generated by the model. Without understanding why a model scored a particular transaction as high, fraud investigators will not know where to start investigating the transaction. P6 noted that models are black boxes because they do not explain why certain transactions are flagged as potentially fraudulent. The models just tell users that a transaction has a high risk for fraud. “But you hand that to an investigator, and they would say, well, I don’t know where to start looking,” said P6. “We found that if you can turn that into a scorecard with a series of business rules effectively that translate the results of the model into okay these are the high-risk factors and these are the rules that triggered,” P6 explained. The scorecard provides “a to do list of things that you want to go look at,” said P6.

P9 believed that one of the challenges of working with models is the explainability piece. Because models are using more and more complex calculations, the auditors and fraud investigators are having difficulties explaining why the model selects a particular set of transactions as fraudulent. It would be helpful if models generated information to show what triggered the transactions, what rules were broken or what data points led the model to flag the transaction, said P9. Further, P2 recommended simplicity when it comes to building models. However, P2 did not advocate for simplicity just for the sake of simplicity. Keeping models simple allows auditors and fraud investigators to explain the model, said P2. Also, P2 recommended starting with simple solutions and then moving forward with more complicated solutions. P5 stated that if an investigator cannot document the process used to identify the fraud,
then the model is worthless, even if the technique is successful. “If it’s a black box, you’re never going to sell it,” P5 stated.

Finally, P10 stated that “accounting is a domain where explainability is super important.” When auditors or fraud investigators say a transaction or a company is fraudulent or not fraudulent, they must have supporting evidence and be able to explain the “line of reasoning” as to how they arrived there, said P10. In terms of approaches or models for auditors to use, deep learning is probably off the table because it may be too difficult for auditors or fraud investigators to explain how they arrived at their conclusion, explained P10. “Different disciplines have different risk tolerances, and advertising has a very low risk tolerance, so it errs on the side of being more predictive and less explainable and audit is the opposite,” said P10. Margagliotti and Bolle (2019) explained the need to avoid adopting a black box approach:

Because there is a decision to make at the end (by investigators, analysts or even judges) based on the results of the various analysis carried out, there is a strong need of transparency. For this reason, avoiding the black box effect of the process applied constitutes a major challenge to integrate machine learning approaches in forensic science. The need for transparency and understanding of the algorithms also applies to the data sets used to train those algorithms. A biased or a nonrepresentative data set will lead to biased results and pose significant risks. (p. 139)

**Summary of the Findings**

Accounting fraud continues despite the efforts by the government, corporations, and professional organizations to stop it. The vast amount of data generated by organizations can be used to hide fraud, but the data can also be used to detect fraud. However, even with the significant volume of data and advanced technologies, fraud detection models do not provide
absolute assurance that fraud will be detected (Kaban, 2020). The best approach for developing predictive models that guarantee accounting fraud detection is still an open question (Gepp et al., 2018; Mark et al., 2019). Without the necessary data to build the right model, accounting fraud will still plague companies and shareholders. To understand the problem better, the researcher conducted 10 interviews with experts from accounting, auditing, technology, and academia. Additionally, the researcher collected literature relating to artificial intelligence and predictive modeling used for accounting fraud detection. Further, the researcher reviewed surveys, articles, and blogs from accounting and technology firms as well as professional organizations relating to artificial intelligence and predictive modeling. The themes and subthemes identified from the research data collected are displayed in Figures 2 to 7.

Research question one concentrated on understanding better why predictive models do not accurately and consistently detect accounting fraud. There are several factors that prevent auditors and fraud investigators from accurately predicting accounting fraud. The fact that labelled data is not readily available, reported fraud is rare, and a vast number of data features are available to select contribute to inaccurate and inconsistent predictive models (Bao et al., 2020; Mohammadi et al., 2020). Unfortunately, building quality predictive models is not as easy as selecting data and an algorithm and then pushing a button. Building quality predictive models is much more complex and requires technical skills beyond what are taught in the traditional accounting curriculums currently provided in higher education. Artificial intelligence and predictive models are in the early stages of development and require more research and action before models are ubiquitous and included in standard audit procedures. Figure 2 displays four key areas where auditors and fraud investigators are likely to encounter pitfalls that prevent them from building quality predictive models.
Figure 2. Research Question One and Applicable Themes

Figure 3 displays the key factors at the data level that act as roadblocks preventing auditors and fraud investigators from building and deploying quality predictive models.
Figure 4 displays key technology concepts that auditors and fraud investigators must understand before they can build and deploy quality predictive models.
Figure 4 displays the key themes identified relating to auditors and fraud investigators. Because of the technical nature of artificial intelligence and predictive modeling, analysts must possess certain skills relating to mathematics, statistics, data science, technology, and data analysis, as well as professional skepticism and judgment. However, this is not to suggest that one individual must possess all these skills, but rather that an audit or fraud investigative team must represent these disciplines.
Figure 5. *Research Question One and Key Individual Themes*

Figure 6 displays the key themes that organizations must understand if they hope to have any chance at developing and deploying quality predictive models.
Research question two concentrated on how to use financial, nonfinancial, and textual analyses to detect accounting fraud warning signals. Traditionally, audit and fraud investigative techniques were used mostly financial data to identify accounting fraud. More recently, linguistic data is being combined with financial data to illustrate a more complex story than had been told with only financial data (Hájek & Henriques, 2017). In the past, unstructured or textual data could not be easily processed to extract information. However, over the last 10 years, new techniques such as NLP have been developed to process unstructured data so that valuable
information can be extracted to highlight warning signals that could not be seen before. Figure 7 displays the two key themes that were developed relating to research question two.

In summary, an overview of the study and presentation of the research findings analyzing the data exploring the perceptions of subject matter experts who have vast experience in auditing and investigating accounting fraud as well as building predictive models has been presented. Recommendations for actions and further study, along with personal reflections and study conclusions are discussed below.

**Application to Professional Practice**

In this qualitative case study, the researcher explored the perceptions of subject matter experts for building and deploying artificial intelligence and predictive models to detect accounting fraud. This research study identified the challenges that auditors and fraud investigators can expect as they build and deploy predictive models. The lack of labeled data to train predictive models is the most significant obstacle to overcome. However, the lack of labeled data is not the only roadblock to building quality predictive models. This study identified

![Figure 7. Research Question Two and Applicable Themes]
the gap between current predictive models and desired models that can predict accounting fraud accurately and consistently. Oussous et al. (2018) noted that additional research is needed before analytical results from predictive models can be optimized. By divulging the challenges and roadblocks with respect to building and deploying quality predictive models, organizations and the accounting profession can take realistic steps towards resolving the gap between current predictive models and desired predictive models. In the meantime, using the knowledge provided in this study, the accounting industry can develop processes and techniques that incorporate best practices from traditional data analytics, current predictive modeling, and machine learning techniques. This researcher believes it is important to make an honest assessment of where the industry stands with respect to using artificial intelligence and predictive models to detect accounting fraud. Overall, artificial intelligence and predictive models should not be thought of as a singular solution but part of a larger accounting fraud detection system.

Based on the marketing information provided by some organizations and accounting firms, one would believe that predictive models can predict accounting fraud accurately and consistently. Organizations and accounting firms are investing significant sums of money to acquire technology and models that most likely will not live up to the hype, at least not now. Sound wisdom would suggest that instead of going all in on artificial intelligence and predictive modeling, organizations should take a more tempered approach and use the best of traditional data analytics, expert systems, and machine learning techniques. The researcher believes this approach would yield better results without overspending on technology and models that still need to be improved. For several years, artificial intelligence and predictive modeling have been growing in popularity. Auditing firms are jumping on the bandwagon and announcing their desire to partner with technology companies and use artificial intelligence. KPMG announced
that it is working with IBM’s Watson to assist on audits (Kokina & Davenport, 2017). Technology companies are hyping the benefits of artificial intelligence and predictive modeling. P1 and P8 warned against falling prey to this hype and forgetting about traditional data analysis using statistics, mathematics, and a rules-based approach.

However, P1 and P8 did not say to avoid predictive models but recommended using them as part of a combined approach to detect accounting fraud. Artificial intelligence and predictive models can make a significant contribution to the detection system as they help the expert view and analyze 100% of a data set and learn patterns and trends, which they were not capable of doing with past static technologies. Therefore, it is important to include the use of predictive models to intelligently evaluate financial and textual data to identify warning signals generated by high-risk areas and transactions. By analyzing data in this intelligent manner, false positives can be reduced. Auditors and fraud investigators can focus on the exceptional exceptions, instead of wasting time resolving false positives. The accounting profession must not fall behind but continue to make strides in developing advanced technology to facilitate the detection of accounting fraud.

F. Tang et al. (2017) explained that data mining uses many techniques, including statistics, mathematics, artificial intelligence, and machine-learning to extract useful information from data. Relying solely on artificial intelligence and predictive modeling to detect accounting fraud is not a practical approach, at least at this point. In other words, if auditors attempted to only use artificial intelligence or predictive modeling to combat accounting fraud, they would miss valuable warning signals right before their eyes. A combination of traditional techniques and new techniques is the best approach for detecting accounting fraud. Given the current state of data and technology, predictive models are not an all-encompassing solution for accounting
fraud detection. Auditors and fraud investigators must possess many tools in their proverbial toolbox to detect the subtle clues generated by perpetrators committing accounting fraud. The notion that artificial intelligence and predictive models always produce correct predictions is incorrect as more work must be done before models can predict accounting fraud accurately and consistently.

P8 recommended having a more diverse set of analytic capabilities within the offerings of an audit firm so that predictive analytics is not the only component. Further, P1 warned auditors and fraud investigators not to overlook the value of simple analytics such as a rules-based logic system. Also, P7 and P10 suggested that the results from a rules-based analysis could be used as inputs into a predictive model, which would provide a more robust system than just using a rules-based system or a predictive model separately. P10 recommended viewing the detection system like a funnel. At the top of the funnel, auditors could use coarse tools such as a rules-based approach, where transactions are flagged because rules were broken. However, P10 believes this likely will yield too many false positives to be a viable approach by itself. Therefore, within the detection system, the rules-based approach may be the first step, with the next step becoming a little more refined, consisting of non-learning algorithms, such as a Benford’s Law algorithm and text mining. Next, a learning algorithm such as a supervised machine learning algorithm may be used to take a more refined approach. In the end, the model would generate transactions with the highest risks, and thereby reducing the number of false positives to review and resolve. By understanding which rules were triggered, the auditors and fraud investigators will be in a better position to explain the results of an audit or investigation to clients, regulators, and judges. In other words, this sequential approach will create explainable artificial intelligence and avoid the black-box effect.
Currently, no predictive models exist where auditors or fraud investigators can feed data into a model and predict accounting fraud with any degree of certainty. There is no such technology or predictive model currently. “I have not seen output from a machine learning platform that discovered something without having a human being having to go through thousands of alternative scenarios to say this is the best one,” says P8. Until the perfect model can be developed, a detection system should be used. The goal of a detection system is to layer intelligence to help identify potential warning signals pointing to possible fraudulent activity. With the addition of artificial intelligence and predictive models to the available tools, auditors and fraud investigators can cover much more ground with fewer false positives. Overall, accounting fraud continues to be a significant problem for corporations and audit firms. As it turns out, there is no one solution or perfect model to detect accounting fraud. Prior to this study, the researcher thought it possible to develop a predictive model to detect accounting fraud accurately and consistently. The researcher believed that with enough research, the best model could be identified. However, this study enlightened the researcher. Currently, it is not possible for predictive models to detect accounting fraud on a consistent basis with any accuracy because labeled data is not readily available. Even though predictive models do not exist as the researcher first believed, there are ways to generate labeled data to help improve the accuracy of models, as noted below.

Recommendations for Action

The researcher in this qualitative study discovered several significant concepts that could be proposed to the accounting and auditing profession. The lack of labeled data prevents the development of quality predictive models for accounting fraud detection. Therefore, the recommendation presented in this section focused on how best to overcome this obstacle. The
researcher found that supervised learning was the prominent technique used to build predictive models for detecting accounting fraud. However, to train supervised learning models, a significant amount of labeled data must be collected. Unfortunately, data with fraud labels does not exist in significant quantities. To this point, researchers have used labeled data from government agencies such as the SEC, Tehran Securities Exchange (TSE), and Ministry of Corporate Affairs for the Government of India (Bao et al., 2020; Gupta & Mehta, 2020; Mohammadi et al., 2020). But this labeled data is in short supply and not enough to adequately train a supervised learning algorithm. P5, P7, and P10 presented several techniques to generate labeled data, which are discussed below.

In a perfect world, researchers would like to have data sets that are completely and accurately labeled. However, these type of data sets do not exist, so one way to generate labels for an unlabeled data set is to use passive learning. In this scenario, transactions are randomly selected from the unlabeled data set so that a domain expert can manually label, code, or annotate the randomly sampled data (Miller et al., 2020). Depending on the task, the domain expert reviews the data to identify transactions that were fraudulent and not fraudulent or high risk and low risk. In other words, the expert labels the transactions as okay or not okay. The characteristics associated with the two classes are used to classify the remaining data set population. Based on the review of the expert, the model is updated to eliminate the false positives, and then the model is run again. This technique requires several iterations to tune the model to achieve an acceptable accuracy rate. This approach is labor intensive and likely to generate a significant number of false positives.

An example of passive learning using random sampling includes the following. An expert selects a random sample of 1,000 transactions out of a data set consisting of 100,000
transactions. Next, the expert reviews the 1,000 transactions and labels each transaction as fraudulent or not fraudulent. Finally, the labeled data is fed into a supervised machine learning algorithm to review the remaining 90,000 transactions. The results of the assisted review are not considered the final answer. Additional analytics are performed to increase the accuracy of the results. The passive learning approach could be used as part of the data analytics solution. However, this technique is extremely weak, especially when the data set includes an imbalance of classes like fraud and not fraud. By selecting a sample randomly, the domain expert wastes a significant amount of time coding irrelevant transactions, which is extremely expensive and time consuming (Miller et. al., 2020).

P5 and P7 discussed ways to improve the passive learning or assisted review techniques. Instead of randomly selecting transactions to code, which wastes the expert’s time labeling irrelevant transactions, the sample could be selected based on heuristics such as rules. P5 suggested selecting a sample by initially scoring the transactions based on a rules-based approach. Further, P5 stated that scoring should be based on an understanding of the data and a thoughtful consideration and development of the rules that would indicate fraud. P5 explained that triggers could be round-dollar invoice amounts, invoices paid on the same day, payments to vendors in corrupt countries, and other specific rules. P5 explained that points would be assigned to these triggers and modified if need. “Every time a transaction hits one of those test criteria, it gets a point or 5 points or 10 points or whatever I give it,” said P5. This also allows the investigator to be able to explain exactly how transactions and vendors were identified for further review, explained P5. Further, P5 noted that the scoring only uses a rules-based approach and there is no predictive ability at this stage of analysis. An expert would then review the results generated based on the rules-based approach, and the transactions that were flagged would be
labeled to include or exclude transactions, said P5. In other words, the transactions would be labeled as fraud or not fraud. Based on the transactions selected as fraud, the model would then select other transactions from the full population based on the characteristics of the transactions selected. Essentially, the model is asked to find more transactions like the ones manually selected, says P5. Over time, based on scoring transactions and then labeling the transactions, the model is being trained and becomes smarter and smarter each month, noted P5.

P10 discussed a process to take the passive learning or assisted review method one step further by using weak supervision and active learning. P10 recommended using weak supervision to generate noisy, low-quality labels (Helmstetter, 2021; Hernández-González et al., 2016; Zhou, 2018). Transactions would be flagged, at this point, based on a list of rules relating to financial and textual data. The flagged transactions would be labeled as okay or not okay. The labels are considered noisy because the ground-truth of the labels has not been verified by an expert. The expert would perform a high-level review to ensure that the rules were capturing the appropriate transactions. In other words, the expert would perform a gut check to ensure that the results of the weak supervision process were reasonable and heading in the right direction. According to P10, “weak supervision basically looks at the ensemble of those rules to assign probable labels, so they have a starting point for learning algorithms. So that seems like a perfect kind of mash-up of machine learning for auditing.”

The data with the noisy labels generated by the weak supervision process are fed into an active learning system. At this point, the machine does not have a lot of confidence in the labels generated by the weak supervision process. In other words, the machine does not have a high level of certainty with respect to how to label each transaction (Helmstetter, 2021; Zhou, 2018). There is still a large gap between the two classifications. To help close the gap, the machine asks
the expert for assistance on transactions about which the machine has the most uncertainty. The expert would then review the transactions requested by the machine to provide the appropriate label. With the additional input from the human, the machine can learn how best to label transactions and thereby reduce uncertainty and increase confidence (Jans & Hosseinpour, 2019; Miller et al., 2020).

Active learning reduces the amount of annotating and labeling required by the expert and ultimately reduces costs. Because of the intelligence provided by the human during this labeling process, false positives can be reduced as well. Finally, this process allows the machine to continuously learn from the expert and create explainable artificial intelligence. P10 explained the use of weak supervision and active learning as sequencing the models. For example, an auditor could take millions of claims and run them through an initial scoring model to reduce the number of claims to be reviewed. The auditor would still not be certain of the subset of claims, so a secondary model would be used to reduce uncertainty and increase confidence to help funnel the claims down to the exceptions to review. P10 stated that Instagram and Facebook use models with different goals at different levels of the process. P10 explained the process as follows:

So they might use a very coarse model when they're dealing with billions and billions of records, just to get down to some subset of records that are of interest, and then they'll use more fine grain models that may take longer to run once they have less data…. There’s no reason why you have to just pick one model to solve your problems…. So what you could do is prioritize or curate a set of likely fraudulent transactions for human reviewers to then manually annotate and review.
Recommendations for Further Study

The use of artificial intelligence and predictive models to detect accounting fraud continues to be a subject that is open to future research. This study identified that labeled data is a key component of predictive models. Given this, this researcher recommends further research relating to weak supervision and active learning to generate labeled data. Also, sharing company data could be another way to satisfy the need for the vast amount of data needed for research. Further, additional research should be performed to ensure that accounting, auditing, and data science courses provide the required skill sets for future auditors and fraud investigators to perform sophisticated data analysis techniques using predictive models and artificial intelligence. Finally, additional research should be conducted to evaluate whether third-party providers are a better alternative for accounting firms when it comes to building predictive models. In other words, would accounting firms be better served by outsourcing and working with technology experts rather than performing all aspects of the model development and deployment in-house?

Further research into employing weak supervision and active learning techniques for labeling financial data could assist in the development and deployment of predictive models for accounting fraud detection. The medical industry has made great strides using weak supervision and active learning to develop labels for medical diagnosis (Zhou, 2018). Further, weak supervision and active learning techniques are used to cull through vast amounts of social media information, labeling tweets as fake news, for example (Helmstetter, 2021; Miller et al., 2020). The accounting and auditing communities could benefit from the research and progress made by other industries with respect to these types of innovative labeling techniques.

Throughout this research, the researcher discussed the significant issue caused by the lack of labeled data. Given the sensitive nature of accounting fraud, auditors and investigators cannot
simply share examples of fraud among accounting firms. M. Ahmed et al. (2016) acknowledged that there are limited publicly available data sets of financial fraud with which to perform research. Further, M. Ahmed et al. noted that because of privacy and competition, it is extremely difficult to obtain actual financial data from companies to perform research. West and Bhattacharya (2016a) also found that because of the private nature of financial data, companies are unwilling to share fraudulent information. Sun (2019) believed it is an urgent problem for auditors and regulators and that an information-sharing process that aggregates data from clients within an audit firm should be established to protect the privacy of the client data.

P7 believed that because businesses operate globally, sensitive financial data will need to be shared among countries, which will require governments to get involved and be part of the solution. Without government support, the sharing of sensitive financial data among countries will most likely not occur, said P7. Finally, P10 suggested a way that the raw financial data could be shared and kept private. P10 stated:

Apple…uses a technology called differential privacy…. The idea is you release data up to the point where you can't tie that data back to an individual. And so if it gets down…to three records and you're trying to get…specific statistics about three people, it gets too close to home and you don't give those statistics, but if you're looking at aggregate data across 50 individuals or 1000 individuals, it's fine to report it because you can't pinpoint it to one person. So…from an audit perspective, that's an approach that could be very useful…to share data to keep it anonymous.

Short of sharing raw financial data, P5 suggests sharing the data analysis techniques that have successfully detected accounting fraud “in a consortium type forum, such that you're not sharing the data, you're just…sharing the algorithms, so that other companies can benefit from
what you've learned,” says P5. Further, P5 stated that if a “consortium of other companies [are] all contributing in an anonymous secure way through a distributed ledger, that could really be beneficial." Finally, P5 suggested that in addition to sharing algorithms, companies should also share features that have worked best for detecting accounting fraud. “What were the risk triggers or what were the triggers that described that improper payment? And so the data is not being sent, but the algorithms that we're hitting will be,” stated P5. Overall, further research should be conducted to determine the best way for companies to share data that does not violate confidentiality agreements. Given the need for a large volume of labeled data to train models, it is imperative that further research be performed to determine how best to share sensitive financial information for research purposes.

Another recommendation for further study includes academia. Accounting curriculums must keep up with the demand for advanced data analytics skills. Future research should be conducted to ensure that universities provide students with the necessary skills required by corporations and auditing firms. Today’s students should be comfortable working with data and technology. Auditors and forensic accountants must be well versed in data technology. Sun (2019) discussed the need for an accounting class that incorporates training in statistics, machine learning, data analytics, and programming. Students do not have to be experts in all these fields, but they should be familiar with the concepts so that they are not blindsided while working in the field. Earley (2015) believed that by providing students with data analytics skills at the university level, academics can help close the skills gap that currently exists in the audit arena. Further, higher education courses should be developed to include students from data science and accounting departments. By allowing students from these two departments to interact, academia
can create an environment that allows for an exchange of ideas and skills among these two
groups of students.

Finally, P10 suggested that one way to overcome some of the limitations and challenges
of using predictive models and artificial intelligence by auditors and fraud investigators is to
outsource certain tasks to vendors that specialize in advanced data mining and data analysis.
There is an “emergence of third-party services that could be used as vendors for auditors or
software platforms that could synthesize this type of data,” says P10. Earley (2015) also believed
that one way to address gaps in auditor expertise with respect to data analytics and technology is
to outsource the data analysis tasks. As mentioned above, the larger auditing firms are starting to
work with technology companies such as IBM. Additional research is required to determine the
best way for smaller accounting and auditing firms to work with technology companies as well.

Reflections

This researcher believes that focus and dedication is required to complete a study. The
researcher understood that sacrifices would be required to complete this study. However, the
researcher did not understand that it would be the researcher’s family who would be doing the
sacrificing. The researcher’s wife endured a period without tangible companionship while the
researcher’s son and daughter lost time with their father. The study required quite a bit of time to
perform, and many time management decisions had to be made to the detriment of the family.
The long hours working on the project highlighted how fortunate the researcher is to have a
loving family that wants to spend time with him. By taking away the time normally spent with
family, the researcher realized how valuable it is to experience moments with the family,
whether big or small. There were many family vacations during which the researcher was
required to work on the study while the family ventured out to explore. The researcher missed
those moments and now realizes how precious family time really is. In the past, opportunities to bond with the family may have pass by the researcher without notice. However, now the researcher notices and craves family time and will try not to take it for granted going forward.

Professionally, the researcher learned a new world relating to artificial intelligence and predictive modeling. The researcher began this project believing that predictive models were the epicenter of the technology universe. However, it is clear now that the epicenter is artificial intelligence and that predictive models are only a part of the larger picture. Prior to the study, the researcher believed that only a few predictive models existed and that it would be a matter of collecting the right data and selecting one of a few models to be able to accurately detect accounting fraud. By the end of the research, the researcher came to understand that predictive models are much more complicated to build and deploy. The machines must be trained based on examples as they learn from experiences. The experiences or examples are provided to the machine in the form of labeled data. Also, the research made clear that machines do not possess common sense or intuition. Machines are not able to intuitively identify biased data, and it is up to the expert to use data that would not bias the model. Further, the research highlighted the need for ethical artificial intelligence. Not all data can be used to train models. If past business practices had biased processes, then the data collected from these business systems would likely be biased and, therefore, bias the model. Ultimately, this would perpetuate bias in the business processes. The lesson learned is the need to understand the data used to train models. Experts must understand the examples or experiences reflected in the training data. By doing this, the expert will be in a better position to determine whether the data is biased.

Finally, because of the complexities of the machine learning algorithms, especially neural network algorithms, it is very difficult for the expert to know what triggered the model to flag
certain transactions. In other words, a model can be a black box, making it impossible for experts to explain the results of a predictive model to management, regulators or judges and ultimately lose the trust of the decision makers. Therefore, it is important to simplify models where possible, even if it means losing a few percentage points in the accuracy of the prediction. By learning about the technology that exists under the artificial intelligence umbrella, the researcher understands better how to use models to detect accounting fraud. This study has provided this researcher with a more powerful tool than the current data analysis software that exists today. With unsupervised machine learning algorithms as well as NLP techniques, fraud investigators have a better chance to catch up to the fraudsters.

**Biblical Perspective**

“Can a blind man lead a blind man? Will they not both fall into a pit? A disciple is not above his teacher, but everyone when he is fully trained will be like his teacher.” (*English Standard Version Bible, 2001/2016*, Luke 6:39–40). In this parable, Jesus teaches His followers to understand one’s own faults before judging others. Before one can lead and guide others spiritually, leaders must understand God’s kingdom and related principles. By enlightening the Jewish leaders on leadership principles such as selflessness and humility, Jesus was trying to remove spiritual blindness from his disciples. If the Jewish leaders did not recognize their own spiritual blindness then how could they possibly follow Jesus? This parable relates to the research in terms of humans guiding machines. Before humans can train a machine to detect accounting fraud, they must understand what accounting fraud looks like and have the right examples to train the models. If the experts are unsure of what accounting fraud looks like, then they cannot expect models to detect accounting fraud accurately and consistently. Experts cannot teach or guide machines on something they do understand themselves. If leaders do not
understand the word of God, then how can they lead and be right with God? With learning algorithms, specifically active learning, models are expected to continuously learn and ultimately detect accounting fraud and then when the machines are fully trained, they will be like the teacher.

Jesus teaches his followers that by obeying His word they will know the truth which will set them free from the guilt of sin: “and you will know truth and truth will set you free” (English Standard Version Bible, 2001/2016, John 8:32). Here Jesus is not talking about the lies or falsehoods a child tells his parents, or a husband tells his wife. Jesus is talking about the truth that He is the son of God and is spreading the word of God to the people. By placing faith for salvation fully on Jesus, who died for the sins of man, His disciples will no longer need to be controlled by the power of sin. The crucifixion of Jesus set man free from sin but only if man has faith in Him. For this study, the research showed that the ground-truth of labels is important to understand. As discussed, labels generated by weak supervision contain noise. These noisy labels are not always accurate. In other words, the labels do not always represent the ground-truth because the source of the labels are weak. Without the truth, models will always be inaccurate and provide false results. Without believing in Jesus and the truth He tells about God and salvation, man cannot be set free from sin.

“For just as the body is one and has many members, and all the members of the body, though many are one body, so it is with Christ.” (English Standard Version Bible, 2001/2016, 1 Corinthians 12:12). Jesus is inclusive of everyone including “Jews or Greeks, slaves or free.” (English Standard Version Bible, 2001/2016, 1 Corinthians 12:13). All who have faith in Jesus are one spirit and not divided based on ethnicity or social class. Paul uses the analogy of the body to explain that all members of the church are dependent on each other to be a healthy and
effective body. Each member possesses different spiritual gifts, and together they can operate as one body, under the direction of the Holy Spirit. In a strong accounting fraud detection system, all parts of the detection system need to work in concert to be effective. Artificial intelligence and predictive models are only one part of an effective detection system, along with the human-in-the-loop and other systems such as an expert system using a rules-based approach. Presently, predictive models cannot provide the complete picture of potential fraud in companies. Ultimately, models are only one tool to be used with other data analysis techniques.

**Summary and Study Conclusions**

In summary, the researcher explored the perceptions of 10 subject matter experts with expertise in auditing, fraud investigation, or predictive modeling. Prior research evaluated the performance of selected algorithms to detect various financial frauds including accounting fraud in publicly traded firms (T. Ahmed & Naima, 2016; Bao et al., 2020; Mohammadi et al., 2020). This study attempted to expand the scope of prior research to identify the best machine learning algorithms to detect accounting fraud consistently and accurately. This research focused on two research questions, and six main themes were identified to substantiate the two research questions. Research question one focused on understanding why predictive models were not able to detect accounting fraud consistently and accurately. Research question two focused on how to use financial, nonfinancial, and textual analyses to improve the detection of accounting fraud warning signals. Themes developed from the conclusions in this research study included (1) Data, (2) Data Mining Techniques, (3) Model Input and Output, (4) Human Agency, (5) Approach, and (6) Explainable AI.

The results of this research revealed that there is no one best machine learning algorithm to use to detect accounting fraud. The basic algorithms such as SVM, ANN, decision trees, and
Naïve Bayes perform reasonably well. However, the researcher found that predictive models are influenced much more by training data than by algorithms. The researcher found that data has the biggest impact on how well a model makes predictions. This makes sense, because the model is learning from the examples provided by the expert, and if there are limited examples to feed the model, then the model will most likely not be able to detect accounting fraud consistently or accurately. In the end, the researcher found that data, more specifically labeled data, is much more important to building accurate models than any machine learning algorithm or any other aspect of the model building process. However, because of the sensitive and confidential nature of financial data, especially fraudulent data, publicly available data to perform research in this area is extremely limited. Without labeled data to train machine learning algorithms, experts have little chance of improving the performance of predictive models to detect accounting fraud.

To overcome the challenge of limited labeled data, the researcher noted several techniques used in other industries that could be beneficial to the auditing and investigating communities. Weak supervision and active learning are techniques that can be an effective way to generate labeled data to build better predictive models. Further, to gain access to a larger volume of financial data, the researcher recommended conducting research to determine the best way to share company data without violating confidentiality agreements. The sole purpose of the data sharing would be to perform research to build better predictive models. The United States and international governments will likely need to be part of the process to help facilitate the sharing of financial data. Also, the researcher believes that academia has a role to play in improving the accounting fraud detection system. Higher education courses should be designed for future auditors and fraud investigators to ensure that subjects such as mathematics, statistics, data analysis, and machine learning are taught. Finally, the research revealed that text mining
combined with financial analysis can be a powerful technique for sifting through company data to identify exceptional exceptions (Hájek, 2018; Sun & Vasarhelyi, 2018). By layering the intelligence gained from financial and textual data analysis, auditors and fraud investigators have a better chance of identifying more red flags leading them to detect accounting fraud.

The interview participants acknowledged that artificial intelligence and predictive models require vast amounts of data to be able to train models properly. The participants also acknowledged that the auditors and fraud investigators must possess the appropriate technical skills as well as professional skepticism and professional judgment. Further, organizations must be willing to support efforts to expand their operations using new technology such as artificial intelligence and predictive models. Because of the nature of this endeavor, corporations will need to invest in technology and personnel to build and deploy quality predictive models.

Overall, to survive the constant scheming by fraudsters, companies and the accounting profession must continuously improve their techniques and methods for detecting accounting fraud. This includes using the latest technology to mine data and developing experts who have the necessary skills to detect accounting fraud. Other industries such as the medical profession have used artificial intelligence and predictive models to classify certain illnesses. By detecting the warning signals of illnesses earlier, lives are saved. The accounting profession must not fall behind the fraudster but continue to make advances so that one day perpetrators will think twice before committing accounting fraud because of the high likelihood of detection.
References


https://doi.org/10.1016/j.acclit.2018.01.001


https://www.fasb.org/exp dash/Page/ImageBridgePage&cid=1176169257359#section_7


https://doi.org/10.1007/s10796-016-9647-9


https://doi.org/10.2308/isys-51799


https://doi.org/10.3390/su10020513


Kaban, İ. (2020). Central audit activities as a continous audit approach in the Turkish banking
https://doi.org/10.14783/maruoneri.676406


https://doi.org/10.1016/j.eswa.2016.06.016


https://doi.org/10.2308/jeta-51730


https://doi.org/10.1016/j.visinf.2018.11.001

https://doi.org/10.1080/00014788.2019.1611730

analysis of the differences between data mining techniques and experts’ judgments.


Tang, F., Norman, C. S., & Vendrzyk, V. P. (2017). Exploring perceptions of data analytics in


Appendix A: Participant Interview Questions

Background of Experts
1. What is your position?
2. How long have you worked in this position?
3. What is your educational background?
4. What work experiences and training relate to predictive modeling?

Predictive Models
5. What are the key characteristics for successful predictive models?
6. What is the best financial, nonfinancial, and textual data to use in predictive models to detect accounting fraud?
7. What machine learning technique works best to detect accounting fraud?
8. What are the best algorithms in predictive models for accounting fraud detection?
9. How do scoring models fit into predictive models?

Technology Skills
10. How important is it for users to have a data technology background when working with predictive models?
11. How does professional judgment and skepticism impact the successful use of predictive models?

Limitations and Challenges
12. What limits predictive models from making accurate and consistent predictions?
13. What are the challenges for users of predictive models?
14. What mistakes do organizations make when it comes to using predictive models?
15. Do time pressures prevent the effective and accurate use of predictive models?
16. What can be done to overcome these limitations and challenges?

**Future**

17. What improvements need to be made to predictive models to improve accuracy and consistency relating to accounting fraud detection?

18. What can organizations do to improve the use of predictive models?

19. What future research do you recommend to improve predictive modeling and artificial intelligence?

20. What potential changes do you expect over the next five years relating to predictive modeling and artificial intelligence?