THE EFFECTS OF ADVANCED ANALYTICS AND MACHINE LEARNING ON THE TRANSPORTATION OF NATURAL GAS

by

BJ Stigall

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Liberty University, School of Business

May 2021

Abstract

This qualitative single case study describes the effects of an advanced analytic and machine learning system (AAML) has on the transportation of natural gas pipelines and the causes for failure to fully utilize the advanced analytic and machine learning system. This study's guiding theory was the Unified Theory of Acceptance and Use of Technology (UTAUT) model and Transformation Leadership. The factors for failure to fully utilize AAML systems were studied, and the factors that made the AAML system successful were also examined. Data were collected through participant interviews. This study indicates that the primary factors for failure to fully utilize AAML systems are training and resource allocation. The AAML system successfully increased the participants' productivity and analytical abilities by eliminating the many manual steps involved in producing reports and analyzing business conditions. The AAML system also allowed the organization to gather and analyze real-time data in a volume and manner that would have been impossible before the AAML system was installed. The leadership team brought about the AAML system's success through transformation leadership by encouraging creativity, spurring innovation while providing the proper funding, time, and personnel to support the AAML system.

Keywords: advanced analytics, machine learning, natural gas, data warehouse, training, proper resources, UTAUT, transformational leadership

THE EFFECTS OF ADVANCED ANALYTICS AND MACHINE LEARNING ON THE TRANSPORTATION OF NATURAL GAS

by

BJ Stigall

Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Liberty University, School of Business

May 2021

Approvals

______________________________ ______________
BJ Stigall, Doctoral Candidate Date

______________________________ ______________
Dr. Gayle R. Jesse, Dissertation Chair Date

______________________________ ______________
Dr. Dennis Backherms, Committee Member Date

______________________________ ______________
Edward M. Moore, Ph.D., Director of Doctoral Programs Date

Dedication

This research project is dedicated to my wife who had her life turned upside down during the last three years while I was in school. Also, to my parents for all they have done, thank you.

Acknowledgments

I would like to thank the best chair anyone could have, Dr. Gayle Jesse, for her guidance through this research study. I would like to acknowledge the wonderful input of Dr. Dennis Backherms throughout this process. I am also grateful to Dr. Edward Moore for listening to me and helping me through some tough times when I was discouraged.

I would like to acknowledge my employer and co-workers for their support during the doctoral journey, and a special thanks to the participating organization and the individual participants.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

Natural gas is a natural resource produced by drilling into reservoirs containing vast quantities of the gas. Once produced, the natural gas must be transported safely to the end-user of the product. Leakage of natural gas can cause environmental damage, and a catastrophic failure of the pipeline system can cause loss of service, or worse, loss of life and property.

The purpose of this qualitative case study was to discover the characteristics of a successful implementation of advanced analytics and machine learning (AAML) to improve the transport of natural gas through the interstate pipeline system. This study intends to add to the literature methods that contribute to the success of implementation, the benefits of implementing, and roadblocks to implementing an AAML system.

Section 1 describes the background of the problem, defines the problem, states the purpose, describes the nature of the study, discusses the method and design, presents the research questions, lays out the conceptual framework, defines terms, presents the assumptions, limitations, delimitations, defines the significance of the study, and presents a literature review.

Background of the Problem

There has been an increasing demand for improvements to processes and understanding of the processes through data (Kornecki & Strube, 2018). However, monitoring and collecting data for AAML can be expensive and time-consuming (Lechevalier et al., 2019). Most current supervisory control and data acquisitions (SCADA) systems do not support the level of data collection and analysis necessary to process the vast amounts of data collected by modern sensors (Shinozuka et al., 2015). Implementing complex advanced analytical services can be difficult for organizations (Giannino et al., 2018). The process control field has largely ignored advanced analytics in recent times (Dong & Qin, 2018). Therefore, organizations underutilize the volume of data made available by the various sensors placed along the production path to monitor and detect the system's failures (Peres et al., 2018).

However, organizations that implement advanced analytics can increase the performance of their systems (Giannino et al., 2018). For example, the biotechnological industry utilizes real-time analytics to control the production process to ensure the production process is as efficient and well-controlled as possible (Kornecki & Strube, 2018). Some in the utility industry utilize advanced analytics to monitor for failures along their distribution network to detect anomalies and predict the customer's usage to have supply where needed (Farzaneh-Gord & Rahbari, 2016; Shinozuka et al., 2015).

The failure of a natural gas pipeline can have a devastating effect on customers and the population surrounding the pipeline facilities (Farzaneh-Gord & Rahbari, 2016). Utilizing predictive methods, through the use of advanced analytics and machine learning, will help avoid service interruption by reducing catastrophic failures and ensuring supply is transported to the customer endpoints on a timely basis (Farzaneh-Gord & Rahbari, 2016; Shinozuka et al., 2015). Furthermore, predicting the customer's volumetric flow needs requires monitoring and prediction facilities to accurately anticipate the usage requirements of the end-users of the pipeline system (Farzaneh-Gord & Rahbari, 2016).

Problem Statement

The general problem to be addressed is the underutilization of advanced analytics and machine learning for process monitoring resulting in reduced system reliability. Flouris et al. (2017) stated that complex event processing systems are utilized to detect and alert operators when adverse operating conditions occur by applying rules to combinations of primitive events. Furthermore, Peres et al. (2018) supported the use of predictive systems to alert operators to potential failures within the systems before a critical malfunction disrupts production. However, Giannino et al. (2018) found that monitoring complex systems can be difficult for organizations, and Shinozuka et al. (2015) stated that most supervisory control and data acquisition (SCADA) systems are limited in the types of data collected from sensors installed on production systems. Peres et al. (2018) also discovered that organizations underutilized the volume of data generated by industrial systems to monitor and predict the failure of production processes. However, when AAML has been adopted, an increase in a production system's performance can be realized (Giannino et al., 2018). The specific problem to be addressed is the underutilization of advanced analytics and machine learning for process monitoring production systems within the interstate natural gas pipeline industry in the United States, resulting in reduced system reliability.

Purpose Statement

The purpose of this research project is to discover the improvements realized by implementing advanced analytics and machine learning (AAML) and the reasons for failure to fully utilize such a system for interstate natural gas pipelines in the United States. While the benefits seen by implanting an AAML are great, implementing AAML systems can be a complex and expensive task for organizations, and past failures of advanced analytics have caused many in the process control field to ignore advanced analytics (Dong & Qin, 2018; Giannino et al., 2018; Lechevalier et al., 2019). Further hampering the full utilization of AAML, supervisory control and data acquisition (SCADA) systems cannot process the amount of data required for AAML (Shinozuka et al., 2015).

Nature of the Study

This study was conducted utilizing a qualitative case study method. The purpose of qualitative research is to find the how and why of a research problem, whereas quantitative research looks for relationships among variables to arrive at a conclusion (Yates & Leggett, 2016). This study sought to understand how an advanced analytic and machine learning system could enhance the operations of a production system, specifically an interstate natural gas pipeline. Therefore, a qualitative research method was chosen.

The primary instrument utilized within this case study was interviews of information systems, business, and pipeline professionals at various levels within the organization. The interview questions focused on the operational effects of the advanced analytic and machine learning system in the pipeline system's operation. The interviews were coded and analyzed to determine how advanced analytic and machine learning system affected the operation of the pipeline system and what the effect of utilizing the advanced analytics and machine learning systems were.

Discussion of Method

There are three methods to choose from when designing a research study: quantitative, qualitative, and mixed methods. The researcher must choose the best design that fits the problem under study. The choice of method depends on the central phenomenon, sample size, and data collection methods (Yates & Leggett, 2016).

Quantitative. Quantitative research is numerical-based research utilizing measurement and quantification (Robson & McCartan, 2016). Quantitative research employs instruments (i.e., surveys) that are predetermined before the study begins (Boeren, 2017). These instruments are analyzed after all the data has been collected (Boeren, 2017). Statistical methods are used to analyze the data to find insight into the hypothesis (Robson & McCartan, 2016).

Qualitative. Qualitative research is utilized when the research study aims to understand the meaning of a human or social problem (Creswell & Creswell, 2018). According to Stake

(2010), qualitative research searches for understanding a problem through the perception of the human, through interpretation. The researcher builds an understanding through the participant's viewpoint in a natural setting (Creswell & Poth, 2018). Qualitative research is utilized when the how and why of a problem are attempted to be discovered (Yates & Leggett, 2016).

Mixed. Creswell and Creswell (2018) described mixed methods as “an approach to inquiry involving collecting both quantitative and qualitative data, integrating the two forms of data, and using distinct designs that may involve philosophical assumptions and theoretical frameworks” (p. 4). By utilizing mixed-method research, the similarities and differences of a problem can be brought to light (Östlund et al., 2011). Mixed-method research is appropriate when greater insight is required than what quantitative and qualitative alone will provide (Creswell & Creswell, 2018).

Discussion of Design

Five qualitative research methods were considered for this research study. The methodologies considered for this study included: narrative, grounded theory, phenomenology, and ethnography. The following is a discussion of each method, the basis for utilizing the selected method, and the reasons for rejecting other methods.

Narrative. Narrative research is utilized to understand an experience (Clandinin & Caine, 2008). The narrative research methodology began to take shape in the 1990s (Clandinin & Caine, 2008). Narrative research typically focuses on one or two individuals (Creswell & Poth, 2018). The researcher collects data through the subjects' stories, and the stories are organized in a chronological manner (Creswell & Poth, 2018). Narrative inquiry is an ongoing methodology whereby the researcher probes the experience before, during, and after the interview (Clandinin & Caine, 2008). The interviewer may also decide to embed into the environment to conduct research. The researcher may utilize visual media or observation to conduct the research while embedded in the environment (Clandinin & Caine, 2008). Even with different methods to conduct a narrative study, researchers are looking for a deeper understanding of the lived story of the subject being studied (O'Grady et al., 2018). In addition to the phenomenon under study, researchers should be concerned with the way the study is conducted (O'Grady et al., 2018). This study attempts to understand how advanced analytics and machine learning systems enhance the value of the organization. Because this research is focused on the effects of a computer system on an organization's value instead of an individual person, this methodology was ruled out.

Phenomenology. Phenomenology studies attempt to discover the common thread among a group of individuals who have experienced a phenomenon (Creswell & Poth, 2018). While phenomenology focus on ideas and essences, it recognizes that a real-world exists (Moustakas, 1994). Phenomenology has been utilized to bring to the forefront phenomena experienced by humankind that has been ignored in the past (Wojnar & Swanson, 2007). Phenomenology reconciles the conflict between idealism and realism by utilizing phenomenological methods to discover the essence of the phenomena (Moustakas, 1994). It is vital that the researcher enters into a phenomenological study without any preconceived notions of the phenomenon and approaches the subject with a fresh perspective to discover the essence of the phenomena (Moustakas, 1994). The researcher moves towards a fresh view of the phenomena by utilizing bracketing (Wojnar & Swanson, 2007). By reviewing the notes made in the field, the researcher can reflect on the observations and assumptions to remove any assumptions brought into the research (Wojnar & Swanson, 2007). The phenomenology method was ruled out for this research study. This study focuses on the benefits that a successful implementation of an AAML system provides an organization instead of focusing on the individuals themselves.

Grounded Theory. The purpose of grounded theory is to discover a new theory or generate a new theory (Creswell & Poth, 2018). Grounded theory is popular among the social sciences (i.e., social workers), medicine (i.e., nursing and theory professionals), education, and management professionals (Bryant & Charmaz, 2019). Students find grounded theory popular since it allows them to turn their practical experience into theory; therefore, providing more efficient treatments (Bryant & Charmaz, 2019). The fit of a researcher to grounded theory may be more effective depending on the researcher’s age, experience, and stage of career (Bryant & Charmaz, 2019). Researchers who perform a grounded theory study must be able to conceptualize data, be tolerant of confusion, and tolerate confusion’s attendant regression while waiting for the concepts to emerge from the data (Bryant & Charmaz, 2019). This study is primarily focused on the effects of implementing advanced analytics and machine learning system rather than developing a theory for advanced analytics and machine learning systems, grounded theory is not appropriate for this study.

Ethnography. Ethnography is the study of a larger shared culture group than the number of participants typically found in a grounded theory study (Creswell & Poth, 2018). Ethnography focuses on the shared patterns of values, behaviors, beliefs, and language of a group that shares a culture (Creswell & Poth, 2018). Examples of culture groups include schools, churches, and workplaces (Creswell & Poth, 2018). An ethnography study's primary purpose is to provide a complete picture of the culture’s views and actions. In addition, an ethnography provides a picture of the group's environment under study (Reeves et al., 2008). Researchers conducting an ethnography study utilize informal methods of gathering information about the culture to collect the complex data necessary to find the solution to the research questions (Reeves et al., 2008). An ethnography allows the researcher to embed into the culture to observe the group's behaviors and actions and perform the informal interviews (Reeves et al., 2008). Because ethnography is a study of culture, it is not appropriate for this study.

Case Study. A case study focuses on contemporary real-world events and answers the questions of why and how (Yin, 2018). A case study may involve the study of one case or multiple cases to draw conclusions (Creswell & Poth, 2018). A case study is an exploratory methodology that leans towards answering the why and how questions of research (Yin, 2018). The case study methodology is a preferred method when the environment cannot be manipulated, and the event is a contemporary, ongoing event (Yin, 2018). A case study utilizes documents, physical artifacts, direct observation, and interviews to answer the why and how questions (Yin, 2018). A case study is typically focused on one case or a small number of cases to answer the how and why questions (Dul & Hak, 2007). The case study was the most appropriate method for this research study. The research focused on the how and why questions of a contemporary event, and the researcher cannot manipulate the events.

Summary of the Nature of the Study

This study is a qualitative study utilizing the case study methodology. Robson and McCartan (2016) defined a case study as “a strategy for doing research which involves an empirical investigation of a particular contemporary phenomenon within its real-life context using multiple sources of evidence” (p. 150). This study sought to discover why companies fail to utilize AAML systems to improve their production systems' performance. This study also sought the contributing factors of the failure to utilize AAML systems and to discover success factors for those who implement AAML systems. The narrative, grounded theory, phenomenology, and ethnography methodologies are focused more on the shared culture of a group of people. Therefore, they were not appropriate to answer the how and why questions a case study allows the researcher to answer. The case study allows the researcher to employ multiple artifacts and interview individuals involved in the case. Therefore, the case study was the most appropriate method for this research study.

Research Questions

When breaking the problem statement apart, the following emerged as the basic concepts from the problem statement: (a) the failure to fully utilize advanced analytics and machine learning, (b) reduced reliability through failures, and (c) the natural gas industry. The research questions developed from the problem statement include:

RQ1: Why do companies fail to fully utilize advanced analytics and machine learning for process monitoring?

RQ1a: What factors contribute to the failure to utilize advanced analytics and machine learning?

RQ1b: What factors contribute to the success of utilizing advanced analytics and machine learning?

RQ2: How do advanced analytics and machine learning affect the production process?

RQ3: How do advanced analytics and machine learning improve service in the natural gas industry?

Conceptual Framework

The conceptual framework analyzed the effect of implementing an AAML package on an organization by comparing the organization's previous state with the state of the organization post-implementation. The concepts that increase customer satisfaction and an increase in profit were examined in this study. The implementation was compared to the Unified Theory of

Acceptance and Use of Technology model and the Technology Acceptance Model. The organization leadership was examined and compared to the Transformational Leadership Model.

AAML Failure and SuccessThere has been an increasing demand for improvements to processes and understanding of the processes through data (Kornecki & Strube, 2018). However, monitoring and collecting data for advanced analytics and machine learning can be expensive and time-consuming. Most current supervisory control and data acquisitions (SCADA) systems do not support the level of data collection and analysis necessary to process the vast amounts of data collected by modern sensors (Lechevalier et al., 2019; Shinozuka et al., 2015). Implementing an AAML system can be complex and challenging for organizations (Giannino et al., 2018). Therefore, organizations underutilize the volume of data made available by the various sensors placed along the production path to monitor and detect failures of the system (Peres et al., 2018).

However, organizations that utilize advanced analytics can increase the performance of their systems (Giannino et al., 2018). For example, the biotechnological industry utilizes real-time analytics to control the production process to ensure the production process is as efficient and well-controlled as possible (Kornecki & Strube, 2018). Additionally, some in the utility industry utilize advanced analytics to monitor for failures along their distribution network to detect anomalies (Shinozuka et al., 2015) and predict the usage of the customer in order to have supply where needed (Farzaneh-Gord & Rahbari, 2016).

Effects of an AAML

Advanced analytics provides real-time analysis of the production process, thereby showing the changes in the process, allowing for the detection of emerging problems (Dong & Qin, 2018). The growing presence of the Internet of Things (IoT) technology allows for insight into processes not previously available (Krumeich et al., 2016). Continually monitoring the processes allows organizations to adjust according to the current business environment and the data produced by the IoT devices, therefore, reacting more quickly to changes than organizations that do not employ advanced analytic techniques (Krumeich et al., 2016). In addition, the collection of this massive data allows for the prediction of future states to allow organizations to act proactively instead of reactively to process changes (Krumeich et al., 2016).

Improved Service

Reliable gas supply to customers requires a timely view of the gas pipeline operating conditions (Su et al., 2018). The failure of a natural gas pipeline can have a devastating effect on customers and the population surrounding the pipeline facilities (Farzaneh-Gord & Rahbari, 2016). The system load on a natural gas pipeline was traditionally consumer heating load and industrial load that was easily predictable (Chertkov et al., 2015). However, with the changing mix of end-users of natural gas, such as power generation, the pipeline system's load is not as predictable and steady as it once was (Chertkov et al., 2015; Zhang et al., 2016).

Increased operational knowledge leads to less unexpected downtime of production facilities giving way to higher reliability, increased customer satisfaction, and increased profits (Giannino et al., 2018; Shinozuka et al., 2015). The higher fidelity data produced by an AAML allows for predicting flow in the system, allowing for more accurate flow prediction; thus, increasing efficiency (Su et al., 2018; Zhang et al., 2016). Furthermore, higher fidelity process data allows for predicting production failure of systems resulting in the avoidance of unplanned downtime and better prediction of end-user needs (Chertkov et al., 2015; Farzaneh-Gord & Rahbari, 2016).

Utilizing predictive methods through the use of advanced analytics and machine learning will help avoid service interruption by reducing catastrophic failures and ensuring supply is transported to the customer endpoints on a timely basis (Farzaneh-Gord & Rahbari, 2016; Shinozuka et al., 2015). Furthermore, predicting the customer's volumetric flow needs requires monitoring and prediction facilities to accurately anticipate the usage requirements of the end-users of the pipeline system (Farzaneh-Gord & Rahbari, 2016).

Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology (UTAUT) theory is a unified theory that combines elements of eight separate theories into one unified theory of acceptance of information technology initiatives (Venkatesh et al., 2003). This theory considers the participants' personal traits to determine the acceptance of new information systems (Venkatesh et al., 2003). The UTAUT also considers whether the system's adoption is mandatory or voluntary (Venkatesh et al., 2003). Additionally, UTAUT takes into account the performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2003).

Technology Acceptance Model

The Technology Acceptance Model is widely utilized in the technology field (Scherer et al., 2019). This model looks at core variables, outcome variables, and external variables (Scherer et al., 2019). Core variables include (a) perceived ease of use – degree the technology will be free of effort, (b) perceived usefulness – degree the technology will enhance the job, and (c) attitudes toward the technology – evaluation of behavior associated with the use (Scherer et al., 2019). Outcome variables include (a) behavioral intention – intention to use the technology, and (b) technology use – actual use (Scherer et al., 2019). External variables include (a) subjective norm – the perception of co-worker's approval or disapproval of the action, (b) computer self-efficacy – belief that the person can use the technology, and (c) facilitating conditions – the degree to which the organization provides the support of technology (Scherer et al., 2019).

Transformational Leadership

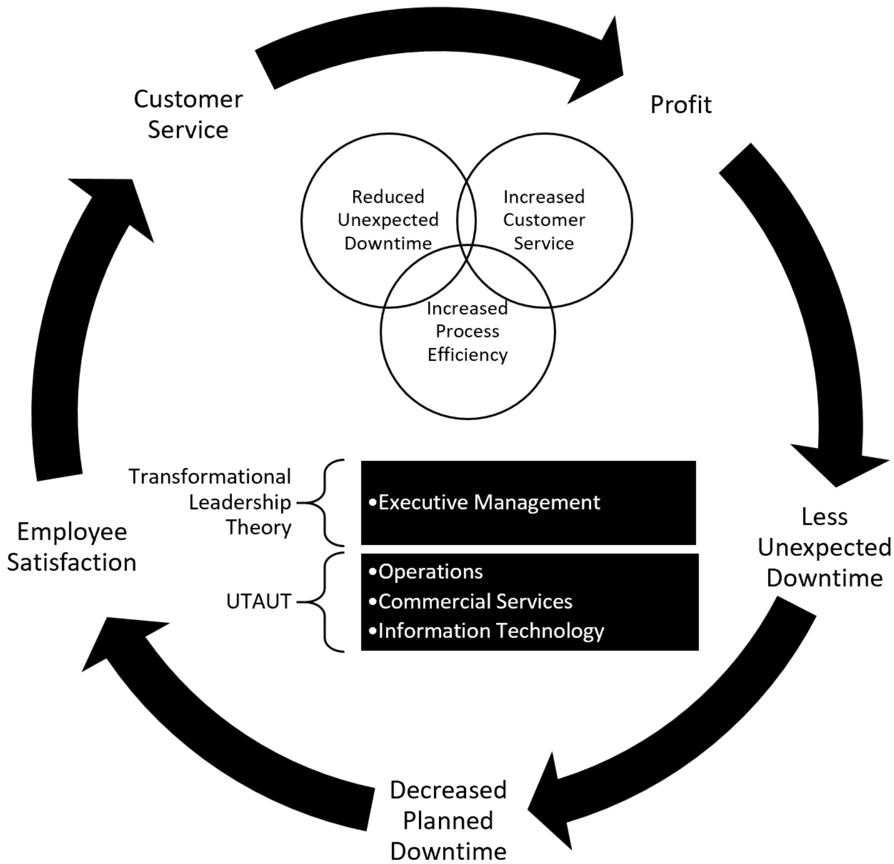
Transformational leadership has shown a positive effect on the internal development of innovative solutions to problems an organization may face (Gumusluoğlu & Ilsev, 2009). Teams have become an increasing unit of organization within the workplace. Those teams are working on a larger number of intelligence and cognitive tasks, processing more information than ever before (Schippers et al., 2008). Transformational leadership encourages people to go beyond their self-interest and motivates them to grow in their skills by encouraging them to remove themselves from the 'always been done this way' box (Schippers et al., 2008). Transformational leaders allow for the proper funding, time, and personnel to support advanced innovation within an organization (Gumusluoğlu & Ilsev, 2009). Transformational leaders also encourage creativity, which allows for increased innovation (Gumusluoğlu & Ilsev, 2009).

Discussion of Relationships Between Concepts

Transformational leadership and UTAUT were the chosen frameworks to apply to this study. Both theories provide a view into the success or failure of implementing an AAML project. Transformational leaders encourage teams to work together to find creative solutions to issues facing the organization (Schippers et al., 2008). UTAUT is a combined theory of information technology implementation that looks at implementing technology in the light of user behavior and a technological lens by combining several different models into one unified model (Venkatesh et al., 2003). Combining transformational leadership and UTAUT, a full picture of the success or failure of implementing an AAML system will be illustrated.

Figure 1

Conceptual Framework



As Figure 1 illustrates, increased customer satisfaction and profits earned are the results of better utilization of production assets through decreased downtime, increased efficiency, and reduced costs.

Summary of Conceptual Model

The conceptual model for this study combines Transformational Leadership Theory with UTAUT to analyze the implementation of an AAML project from both the leadership perspective and the user perspective of implementing a technology project. Transformational leaders change the goals of an organization's employees from goals only satisfying the needs of the individuals into goals that provide the best results of the collective good (Tourish & Pinnington, 2002). Meanwhile, UTAUT examines the behavioral and technological aspects of implementing a technology project (Venkatesh et al., 2003). UTAUT combines several technology theories into one unified theory (Venkatesh et al., 2003). By examining both Transformational Leadership and UTAUT, a fuller picture can be drawn from implementing an AAML system.

Definition of Terms

Advanced Analytics. Real-time monitoring software system that attempts to help operators maintain or improve production quality (Kornecki & Strube, 2018).

Business Intelligence. The extraction of information from data (typically structured data) through statistical methods and data mining techniques (Chen & Sun, 2018).

Internet of Things (IoT). Internet-connected devices (Boyes et al., 2018).

IIoT. Industrial components connected through smart sensors to allow advanced analytics of the resulting data (Boyes et al., 2018).

Machine Learning. A technique utilized by computer systems to learn patterns in data (Louridas & Ebert, 2016).

SCADA. A Supervisory Control and Data Acquisition system is a centrally located monitoring and control system that allows operators to remotely control the operation of industrial equipment (Boyes et al., 2018).

Assumptions, Limitations, Delimitations

This study was designed with a careful evaluation of the study's assumptions, limitations, and delimitations. Assumptions are factors that may not be in the researcher’s control (Simon &

Goes, 2018). Limitations may introduce weak points in the study, while delimitations are boundaries placed around the study by the research designer (Simon & Goes, 2018).

Assumptions

An assumption is a conclusion about behaviors or outcomes based on inferences made by observing another's behavior (Hagger & Chatzisarantis, 2009). In research, relying too heavily on assumptions can lead to incorrect conclusions about the outcomes of the research conducted during the study. To mitigate the effects of assumptions on a research study, researchers can modify their tools and methods to account for assumptions generated by complex problems. Also, researchers should properly document assumptions, so the effect assumptions have on the study are clear (Nkwake, 2013).

The following assumptions were assumed to be true conditions for this study. The study participants represented the organization in which they were employed or were employed but do not represent the industry as a whole. The researcher assumed the participants were open and honest about the effects the AAML system had on the organization and the organization's conditions before implementation. It was assumed that the participants were familiar with the current and former working environment. While the participants may exhibit response bias by telling the researcher what they think the researcher wants to hear, the researcher will triangulate the participant responses to verify the participants' information (Leedy et al., 2016).

Additionally, the researcher took steps to mitigate any unwanted effects derived from the assumptions above. The participants were able to review any material provided to the researcher to verify accuracy (Leedy et al., 2016). All participation was voluntary, and each participant had the option to participate or withdraw from the study. All participants had the opportunity to review the research scope and understand how their responses would help define an AAML system's effect on the organization.

Limitations

Limitations are factors that constrain the research, which is out of the researcher's control but could, nonetheless, affect the study (Simon & Goes, 2018). Limitations of this research study included the COVID-19 pandemic and the number of possible participants. This study was conducted during the COVID-19 pandemic of 2020. The pandemic did not allow the researcher to travel to the participants to conduct face-to-face interviews. To mitigate the effects of the pandemic, the researcher conducted the interviews over teleconferencing software. The number of participants available for the study was limited by the number of interstate natural gas pipelines in the United States. There are approximately 30 major pipelines located in the United States (EIA, n.d.c). Further, as a qualitative case study, the focus was limited to a maximum of two participating organizations. All participants were either current or former employees of the participating organizations.

Delimitations

The researcher chose the delimitations of a study. These choices are driven by the methodology, questions, variables, and theoretical perspectives (Simon & Goes, 2018). Boundaries should be carefully chosen to accurately reflect the limitations of the study.

The participant group was limited to a maximum of two organizations within the interstate natural gas pipeline industry. There are approximately 210 natural gas pipeline systems in the United States. Thirty of the largest interstate natural gas pipelines transport about 80% of natural gas volume (EIA, n.b.b). Due to the size of the industry, one pipeline was chosen to participate in the study. Other industries were eliminated as participants in the study because of the nature of interstate natural gas pipelines.

Significance of the Study

Traditional SCADA systems are not equipped to handle the influx of data from the vast amount of sensors installed in today's production systems (Shinozuka et al., 2015). Furthermore, AAML systems are not fully utilized for optimizing the production process (Peres et al., 2018). AAML systems provide the resolution required for the advanced monitoring of production systems not provided by traditional control systems (Kornecki & Strube, 2018). This study specifically addressed the gaps in the literature describing the implementation of an AAML system better to control the production process, specifically natural gas transportation. This study also integrated biblical principles for implanting an AAML system into an organization.

Reduction of Gaps

This study attempted to provide a reduction in gaps in the literature by focusing on the success or failure of implementing an AAML system. Specifically, this study focused on the interstate natural gas industry. The body of literature extensively discusses the benefits of an AAML system; however, the body does not adequately describe what makes the implementation of an AAML system a success or failure. This study attempted to fill this gap.

Implications for Biblical Integration

When God created the Earth and placed man on this planet, He expected us to be good stewards of this planet (*New International Version Bible*, 1978/2011, Genesis 1). Any production process should be monitored to ensure that the process works as efficiently as possible. Natural Gas, being a natural resource, is no different. Natural gas transporters should take the necessary precautions to ensure the gas is transported as efficiently as possible with minimal loss.

Traditional tools utilized for monitoring the production process, such as those used in controlling the transportation of natural gas, are not typically advanced enough to monitor the system at the granularity level that a modern sensory network can provide (Giannino et al., 2018; Krumeich et al., 2016; Shinozuka et al., 2015). This study attempted to demonstrate how an AAML system can be implemented to better monitor the data produced by a modern sensor network and better control natural gas flow within an interstate natural gas pipeline system.

Relationship to Field of Study

Implementing a system, such as an AAML system, requires many information technology systems to work in concert with each other (Flouris et al., 2017; Kornecki & Strube, 2018; Peres et al., 2018; Shinozuka et al., 2015). As these systems are heavily based on information technology platforms, this study relates to the information system field of study. This study aimed to add to the information systems literature by describing how successfully implementing an AAML system will help an organization succeed and point out pitfalls that may occur along the way.

Summary of the Significance of the Study

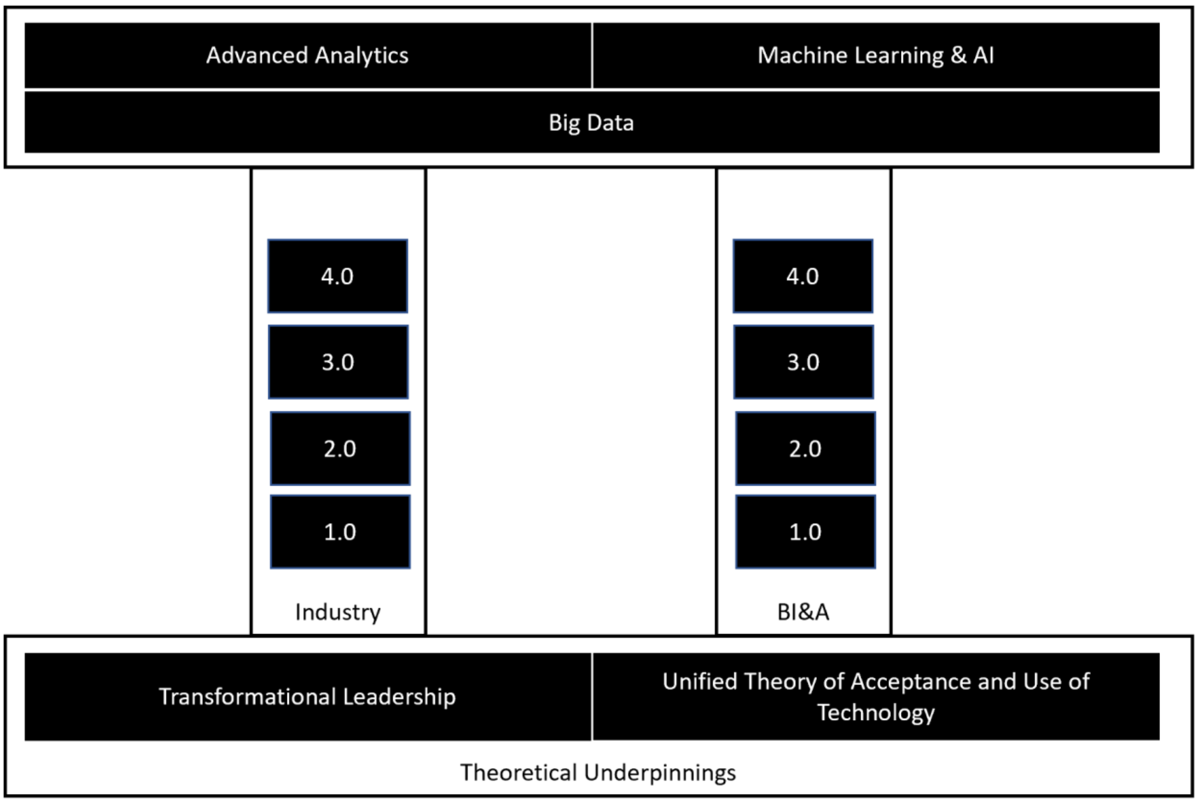
When God created the Earth and placed humans on this Earth, He expected humans to make the best use of the natural resources available (*New International Version Bible*, 1978/2011, Genesis 1). Monitoring systems, such as an AAML system, helps optimize the production process by providing data traditional systems do not (Dong & Qin, 2018; Peres et al., 2018; Shinozuka et al., 2015). This study aimed to reduce the gaps in the literature describing a successful implementation of an AAML system and describe pitfalls that may occur while implementing an AAML system. While the body of literature is extensive in describing the benefits of an AAML system, the literature does not adequately describe the success factors and failure points of implanting such a system.

A Review of the Professional and Academic Literature

As the need for better process monitoring and prediction of natural gas flow increases within the interstate natural gas pipeline industry within the United States, companies transporting natural gas have a need to increase their monitoring of the pipeline system to reduce costs and environmental effects from the transportation of natural gas. However, monitoring and collecting data to better utilize the pipeline system and reduce the environmental effects of transporting natural gas can be complex, expensive, and time-consuming. Current SCADA systems typically do not have the processing capabilities within the software to provide the additional needs for the operators to make the decisions required to process the information in a timely manner (Kornecki & Strube, 2018; Shinozuka et al., 2015). There is a lack of studies on the effects of machine learning and advanced analytics within the interstate natural gas industry. The review of literate begins by building a foundation for the theoretical underpinnings of the study by examining the transformational leadership style and the Unified Theory of Acceptance and Use of Technology. These theories provide the view used throughout the study to understand why implications of advanced analytics and machine learning implementations fail and what makes them successful. Next, the stages of industry are examined to build upon the modern industrial revolution brought about by the advances in industry, beginning with the first industrial revolution in the 1700s. After exploring the development of industry, the development of business intelligence and analytics is explored to show the path from the earliest stages of computerization to the current state of machine learning and advanced analytics. The literature review then focuses on advanced analytics and machine learning technologies by exploring big data, advanced analytics, and machine learning to define their use within industry and analytics. A visual representation of this literature review's building blocks is presented in Figure 2 to show the foundation and pillars that support advanced analytics and machine learning.

Figure 2

Literature Review Diagram



Transformational Leadership

One of the key features of the successful implementation of projects is leadership (Richter et al., 2016). Leadership is a crucial function for any organization to become more competitive in the ever-increasing global marketplace (Keskes et al., 2018). Leadership’s primary functions are to put together the appropriate teams at the right time and with the appropriate resources to achieve the organization’s goals. It is the leader's responsibility to direct the employees' behavior in such a manner to achieve these goals (Keskes et al., 2018).

Leadership provides the necessary guidance, motivation, feedback, and resources to the employee to successfully implement a project. Transformational leadership has been identified as an efficient leadership style for change within an organization (Richter et al., 2016). For information systems projects based on the agile methodologies, Kelle et al. (2015) found that employing a transformational leadership style had a significant influence on the success of a project, no matter the project's size. Open communication was found to be an essential aspect of transformational leadership that contributed to the success of project implementation (Kelle et al., 2015).

At its root, transformational leadership is about transforming self-centered individuals into committed team members working to achieve a common purpose (Alvesson & Kärreman, 2015). Alvesson and Kärreman (2015) identified individualized consideration, intellectual stimulation, idealized influence, and inspiration as the basic transformational leadership elements. They also point out that some assume that the leader can exert influence to modify the employee’s self-confidence, enthusiasm, place in the group, and compliance with group norms (Alvesson & Kärreman, 2015).

Idealized influence. Idealized influence is described as leaders trusting and respecting their followers, appealing to the followers at an emotional level (Bottomley et al., 2016). Idealized influence is also referred to as charisma. Additionally, idealized influence encourages leaders to stand up to challenging situations because they believe it is the right and ethical thing to do (Bottomley et al., 2016). Inspirational motivation encompasses the leader’s ability to communicate the vision in an attractive manner that energizes the followers to internalize the vision and take on the challenge of the vision to perform at their best to achieve the goals necessary to fulfill the vision (Bottomley et al., 2016). Leadership encourages intellectual stimulation by allowing the followers to question the previously held assumptions and beliefs (Bottomley et al., 2016). By allowing the followers to question the status quo, the follower's creativity will be stimulated. New and better ways will be discovered to accomplish the work, and the necessary risks will be taken to accomplish the goals to meet the vision (Bottomley et al., 2016). Leaders take into consideration the individual's needs through individual consideration (Bottomley et al., 2016). Leaders help an individual accomplish the goals by coaching and mentoring the followers and keeping the followers in line with the vision set forth (Bottomley et al., 2016).

Influence on Work. Transformational leadership has an influence on employees' work by providing meaningful work, trust in leadership, and self-efficacy (Hildenbrand et al., 2018). Transformational leadership can help reduce the negative effects of work employees experience through their careers.

Burnout. One such condition is burnout. Burnout is a condition that occurs when an employee's capacity is exhausted through an intense work cycle, and the employee is not allowed to replenish themselves (Schaufeli et al., 2009). Burnout happens when a person experiences chronic emotional and interpersonal job stressors and is recognized as a serious problem within the workforce (Maslach & Leiter, 2016). According to Maslach and Leiter (2016), the components of burnout are exhaustion, cynicism, job detachment, reduced efficiency, and reduced accomplishments. A properly executed transformational leadership style can help mitigate burnout's effects by supplying the employee with the proper resources needed to accomplish their job in a meaningful manner (Hildenbrand et al., 2018).

Safety. Safety is an important aspect of the work environment. Work-related accidents can have long-term, lasting effects on an employee's physical, mental, and economic status (Kim

& Jung, 2019; Shen et al., 2017). Work-related stress is a significant contributor to work-related illness (Muthamia et al., 2015). The stressors may be environmental (e.g., noise, temperature, amount of work, and privacy levels) or may be caused by life events (e.g., divorce, new child, death in the family, and unemployment), which can lead to safety incidents (Kim & Jung, 2019; Muthamia et al., 2015). Hoffmeister et al. (2014) linked leadership’s support of a safety culture to the success of a company’s safety results. They stated that leadership’s behavior is directly related to the worker’s response to an organization's safety programs. Kim and Jung (2019) found that applying transformational leadership can reduce job stressors, improving the safety environment of an organization. Additionally, Shen et al. (2017) found that leaders defining the vision for safety, culture, and practicing safe work practices influenced worker safety through the use of transformational leadership.

Employability. Transformational leadership has the potential to improve the employee’s skill set, making them more employable. In a study conducted by Yizhong et al. (2019), employees who worked under leaders utilizing transformational leadership were able to increase their skills, thereby becoming more valuable to the organization. This improvement in an employee’s skills did not come without risk. As the employees gained skills, they became more mobile in the workforce, which brought about a risk of losing a valuable employee to another organization (Yizhong et al., 2019).

Quality. In professional settings, transformational leadership helps boost the quality of the employees' output (Andersen et al., 2017). In the study conducted by Andersen et al. (2017), it was found that leaders with a medium-sized circle of influence had the best results in improving quality among the professionals that reported to them. The medium-sized groups were still a manageable size to perform effective communications of the employees' goals and desired outcomes. Whereas leaders with a larger circle of influence were not as effective at communicating the requirements for performance, and quality was diminished in these groups (Andersen et al., 2017).

Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology (UTAUT) model was developed in 2003 for the purpose of integrating the competing models of information technology acceptance theories into one unified model (Venkatesh et al., 2003). Eight models were reviewed, and a unified theory was integrated from elements of the eight models. The eight models reviewed were (a) Theory of Reasoned Action (TRA), (b) Technology Acceptance Model (TAM), (c) Motivational Model (MM), (d) Theory of Planned Behavior (TPB), (e) Combined TAM and TPB (C-TAM-TPB), (f) Model of PC Utilization (MPCU), (g) Innovation Diffusion Theory (IDT), and (h) Social Cognitive Theory (SCT).

Theoretical Models

Theory of Reasoned Action. TRA is a theory that integrates behavioral and normative beliefs (Goodarzi et al., 2019). The theory holds that the most significant factor behind predicting behavior is intention (Goodarzi et al., 2019). The development of TRA came about to “better understand relationships between attitudes, intentions, and behaviors” (Montano & Kasprzyk, 2008, p. 68). Before developing TRA, many theorists wanted to eliminate attitude as a factor for predicting underlying behaviors (Montano & Kasprzyk, 2008). However, Fishbein separated the attitude about an object and the attitude about the behavior with regard to an object (Montano & Kasprzyk, 2008). The components of subjective norms are normative beliefs and the motivation to comply with cultural behavior (Montano & Kasprzyk, 2008). However, one aspect ignored by TRA is the demographics of the population under study (Elahe Kordi, 2018).

For example, the elements of brand-switching behavior may vary from location to location based on demographic characteristics in addition to personal and normative beliefs (Elahe Kordi, 2018). Similarly, if an employee considers the new software product useful in their daily work, they may be more willing to accept the new system. In the same vein, if other workers are beginning to use the new package, the new social norm becomes the use of the new software, and the employees who are resistant to the new software package may begin to utilize the product in their daily routine through the construct of social norms.

Technology Acceptance Model. Davis (1986) introduced TAM as a new model for describing technology acceptance, basing the model on TRA. The TAM model is a specialized model designed to only apply to information technology projects (Davis et al., 1989). Davis (1986) proposed the model to improve the user acceptance process's understanding and provide a theoretical basis for user acceptance testing. The TAM model proposes that perceived usefulness and perceived ease of use are determined by the users' attitude, intentions, and actual adoption of the software (Davis et al., 1989). Ease of use, defined as using the system being free of effort, is an important aspect of adopting a new software system (Deslonde & Becerra, 2018). The less effort a user has to expend to learn how to use the system and the effort over time required to use the system should ease adopting a system by the users (Deslonde & Becerra, 2018). Deslonde and Becerra (2018) also identified usefulness as the software package's ability to enhance the user's job performance. Technology that was easy to use was also perceived to enhance the end-user's job performance (Deslonde & Becerra, 2018).

Motivational Model. There are two types of motivation within the motivational model: intrinsic and extrinsic (Vallerand, 1997). Intrinsic motivation is a type of motivation brought about by deriving pleasure and satisfaction from performing the activity (Vallerand, 1997).

Extrinsic motivation is when a person engages in an activity to gain something from outside the activity (Vallerand, 1997). Within the information technology environment, Fagan et al. (2008) found that extrinsic and intrinsic motivations have an effect on the adoption of a new information technology system. Fagan et al. (2008) also found that behavioral intention and ease of use have relationships with intrinsic and extrinsic motivation in adopting an information technology system.

Theory of Planned Behavior. The Theory of Planned Behavior (TPB) is a model of human behavior based on an individual's behavior intention. The behavioral intention is based on three key factors: attitude, subjective norm, and perceived behavioral control and is a function of a person’s willingness to perform the behavior. Based on a person’s behavior intention, their actual behavior can be predicted (Fu & Juan, 2017). A person’s attitude is the measure of a favorable or unfavorable view towards the behavior. The subjective norm factor is the perception the person has that other people think about performing the behavior. Perceived behavioral control is the person’s perception of how easy it will be to adopt the new behavior (Shi et al., 2017).

Combined TAM and TPB (C-TAM-TPB). Combined TAM and TPB merge TAM and TPB into one model to form C-TAM-TPB. The two models intercept by showing the effects of attitude, behavioral intention, and self-reported use. TAM contributes perceived ease of use and perceived usefulness, while TPB contributes subjective norms and perceived behavioral control. The combined model provides a more complete picture of acceptance than the two individual models provide separately (Pynoo & van Braak, 2014).

Model of PC Utilization (MPCU). The Model of PC Utilization attempts to predict personal computers' utilization by users, not the intention of usage (Venkatesh et al., 2003). The core attributes of the MPCU model are affect, job fit, long-term consequences of use, facilitating conditions, and habits. Affect is the positive or negatives feelings towards the use of the PC. Job fit defines the belief the employee has that a PC will enhance job performance. Long-term consequences of use define the belief that using a PC will benefit the employee over the long term. Facilitating conditions is the support received by the employee from the employer to learn how to use the PC in their job functions. Habits are behaviors that occur without conscious knowledge that the behavior is occurring (Thompson et al., 1991).

Innovation Diffusion Theory (IDT). The Innovation Diffusion Theory (IDT), also known as the diffusion of innovation, is a social process where a few early adopters implement innovation into their operation. Over time others start to add the same technology into their processes. This process continues until most, if not all, of the social groups have adopted the technology into their processes (Valente, 1996).

Social Cognitive Theory. Social Cognitive Theory is a widely used model to explain individual behavior. Compeau and Higgins (1995) adapted SCT to information systems by adding acceptance and the use of information technology (Venkatesh et al., 2003). The behaviors are affected by a continuous relationship between the environment, behavior, and the person exhibiting the behavior (Compeau & Higgins, 1995).

UTAUT Development. From these eight models defined above, the UTAUT model emerged with five core determinants of intention and usage (performance, expectancy, effort expectancy, social influence, and facilitating conditions) and four moderators of key relationships (gender, age, voluntariness, and experience) to build the foundation of the UTAUT model (Venkatesh et al., 2003). The UTAUT model is considered to be a robust model to evaluate the acceptance of a new information system (Khechine et al., 2016). Research shows that UTAUT shows that those responsible for developing and deploying new information systems should focus on the functionality and features that will improve the acceptance by users (Khechine et al., 2016).

Industry

The development of the modern industrial complex can be divided into four stages. Industry 1.0 is another name for the first industrial revolution. Industry 2.0 began as industry introduced electricity to power manufacturing equipment, and the concepts of division of labor began to emerge. Industry 3.0 is marked by the introduction of electronics into industrial processes in the 1970s. Finally, Industry 4.0 emerged with the addition of the Internet of Things to the manufacturing process (Chen & Sun, 2018).

Industry 1.0. The modern industrial era began in England in the late 1700s when modern (for the time) manufacturing processes were developed on a large scale (Desmet et al., 2020). Several authors argue the reasons for the first industrial revolution beginning in England and give reasons for this. Spear (2014) argued that coal availability as a fuel helped England begin industrialization of their economy. Until the 1700s, most coal in England came from surface mines, but this source of coal was limited. Mining pits were susceptible to flooding and could be dangerous to the miners. The first operational draining pump was developed in 1714. The new pulping system allowed the pits to be mechanically drained. Even though this device was not efficient, it was effective in keeping the mines drained and used cheap fuel to power the pump. Further, coal spurred the transportation industry to develop new methods of transporting coal. The transportation of coal leads to an expansion of the shipping industry and the invention of the railroad. The broader use of coal allowed for the plentiful supply of raw materials of iron, copper, tin, and other smeltable materials available within the United Kingdom to be manufactured into products for use by the population. Before the efficient mining of coal, these materials were produced by burning wood as fuel. By the time of the industrial revolution, the supply of wood had been diminished to the point where the government had banned the use of wood in many places and limited the number of smelters in other places throughout the United Kingdom (Spear, 2014).

As labor became more expensive and demand began to rise, it became more economically viable to mechanize the manufacture of goods. As the demand increased for goods, the value of building machines made more economic sense. During this time, capital and energy were cheaper to deploy than adding additional labor to fulfill the goals. The increase in demand for new machines spurred inventors to increase their efforts to produce new machines in response to this increasing demand for mechanized production methods (Bottomley, 2014; Desmet et al., 2020).

Industry 2.0. The second industrial revolution, or Industry 2.0, was ushered into existence with the invention of electricity in the late 1870s (Chen & Sun, 2018). During this period, industry began to power its machines with electricity. Also introduced during this period was the concept of Scientific Management (Yin et al., 2018).

Electrification. During the Industry 2.0 era, the volume and variety of products produced by factories exploded. Mass production developed with the use of the assembly line popularized by Henry Ford and Taiichi Ohno began the development of the Toyota Production System (Yin et al., 2018). The development of electrical distribution led to the development of electric lighting, generators, motors, and new railroad technologies (Thomson, 2011). Electricity had a greater effect on moving industry forward than any other invention before electrification. While the steam engine helped to move manufacturing, and to some degree transportation forward, electricity's effects were felt on a broader market than just manufacturing. Electricity helped both farmers and manufacturers increase output to the degree that their output could not find a market. Manufacturing output increased in existing factories because of the electrification of the production process.

Scientific Management. Frederick W. Taylor developed Scientific Management. Taylor, an engineer in the steel industry, became interested in the economic use of resources. Taylor sought to increase productivity and reduce waste through time studies, eliminating wasted motion, and setting production standards. He also described the functional management model, management by exception, and worker training, among other concepts. Taylor applied the term scientific management to his methods after a 1910 rate case proceeding involving the eastern railroads when a witness referred to Taylor's methods as scientific. After the publicity of the testimony that proclaimed the railroad could save up to 1 million dollars per day, the method took hold and was applied to manufacturing across the country (Wren, 2011).

Industry 3.0. The information technology developments of the 1950s and 1960s led to the development of computerized machine control in 1968. The combination of information technology and the conversion from analog machine control to computerized machine control lead to the third era of industry, Industry 3.0 (Kurt, 2019; Yin et al., 2018).

Electronic Controls. Programmable Logic Controllers (PLC) were a key invention to the third industrial revolution. Before the PLC, machinery was controlled by manual intervention or through large relay banks. PLCs introduced efficiencies into the manufacturing process by improving reliability and simplifying maintenance of equipment. The PLC is also programmable, so minor changes in the machine's functionality would only require changes to a program instead of physical relocation of wires and relays (Tasca et al., 2018).

Materials Requirement Planning Systems. Along with the PLC, Material Requirements Planning (MRP) systems were developed in the 1970s. Before the technological developments of the 1960s, the computing platforms available could not handle the data processing requirements needed to fulfill the demands to develop an MRP (Wilson, 2016). MRP systems primarily focused on the bill of material, work orders, shop floor scheduling, production activity control, and inventory management. These systems allowed for the efficient use of the resources available on the factory floor through detailed scheduling of available raw materials and available machine capacity (Kiran, 2019).

Industry 4.0. The smart factory marks industry 4.0's arrival. Smart factory technology is built on the Internet of Things, the Industrial Internet of Things, wireless sensor networks, cloud technology, mobile computing, and big data technology (Wang et al., 2016). According to Wang et al. (2016), a factory that is transitioning into an Industry 4.0 factory should consider three key factors: (a) horizontal integration through value networks, (b) vertical integration and networked manufacturing systems, and (c) end-to-end digital integration of product design across the entire value chain.

Smart Manufacturing. Smart manufacturing originated in the United States but has spread globally to describe the manufacturing technologies based on smart sensor networks that provide data to control systems based on the additional data provided by the new technologies (Mittal et al., 2017). Smart manufacturing systems rely on a network of sensors to provide information to the control systems. The control systems may be in the cloud or local to the factory or piece of equipment. However, all of the equipment is tied together through the manufacturing plant's network and potentially to systems located remotely away from the local plant. These systems communicate through standard protocols allowing for interconnectivity between different manufacturers’ equipment and control and intelligence software (Kusiak, 2018).

Key Factors. Horizontal integration through the value networks of manufacturing means that the manufacturers coordinate their manufacturing capabilities through electronic means, in effect, creating one virtual factory across the entire supply chain (Wang et al., 2016; Xie et al., 2019). Conversely, vertical integration within the factory requires the information flowing from the factory-level sensors on various machines to fully flow the information through the enterprise to the highest level systems, such as the enterprise resource planning (ERP) system (Xie et al., 2019). Finally, the product's design should be integrated electronically through the entire supply chain (Xie et al., 2019). Not only should the process be integrated internally through all the departments within the organization, but the external participants in the design of the product should also be integrated into the process of product design, manufacture, distribution, and recycle (Liu et al., 2018).

Business Intelligence and Analytics

The term business intelligence came about in the 1990s. However, business intelligence and analytics (BI&A), as a sector of information technology, has existed since the 1970s with the development of decision support systems (Watson & Wixom, 2007). As technology has evolved, so have business intelligence systems. Business intelligence has been described in four phases divided along with the technological changes through the years (Chen et al., 2012; Chen & Sun, 2018). No matter the type of BI&A system employed by the organization, the goal of the system is to present transactional data in a manner that is usable to the analyst to extract knowledge from the data to make informed business decisions that support the goals of the organization and further the decision-making process based on that data (Gallinucci et al., 2018).

BI&A 1.0. BI&A 1.0 has its roots in the decision support systems of the 1970s (Watson & Wixom, 2007). This era of BI&A is heavily rooted in the database management systems (i.e., relational databases (RDBMS) and online analytical platforms (OLAP)) for processing and combining data into usable forms for analytics and business intelligence (Chaudhuri et al., 2011; Watson & Wixom, 2007). Business intelligence systems based primarily on structured data located in RDBMS and OLAP systems are considered BI&A 1.0 systems (Chen et al., 2012). These types of systems are the most widely supported in the industry (Chen et al., 2012). Data are loaded into 1.0 systems through the process of extraction, transformation, and load (ETL) and explored through graphical data exploration tools (Chen et al., 2012).

RDBMS. In his seminal paper on the relational model for storing data, Codd (1970) introduced the relational database model and normalization. In addition to developing the database, Codd developed a language based on mathematic symbols to query the data from the database (Chamberlin, 2012). Chamberlin and Boyce saw this type of language as a barrier to entry for many users. They devised a new language called Sequel, which was more readable than the relational algebra developed by Codd (Chamberlin, 2012). As a product class, RDBMS products have become an indispensable category of software in modern-day enterprises (Huang, 2019). The software that runs on top of the RDBMS is mission-critical software applications that businesses depend on for their daily survival in today’s business world (Huang, 2019).

OLAP. As opposed to Online Transactional Processing (OLTP) systems that focus on one application or department, an OLAP system is designed to bring data from multiple sources and analyze the data from a multidimensional viewpoint via interactive interfaces (Wang et al., 2019). OLAP systems rely on data from the source system to be periodically loaded into their data warehouse (Djiroun et al., 2019). Therefore, an OLAP system may not have the most up-to-date information like an OLTP system. However, an OLAP system allows analysts to visualize the data from multiple perspectives drilling up or down into the data to gain greater insight into the information presented by the OLTP system (Djiroun et al., 2019). In recent times, organizations bring data from outside the organization into their OLAP systems to assist in data analysis. The external data bring additional context to the data generated from within the organization (Gallinucci et al., 2018). However, bringing in outside data is not without risk. The data may not be as reliable as data from within the organization since the organization is not in control of the outside data (Gallinucci et al., 2018).

ETL. Extract-transform-load is the process for bringing data into the OLAP system for further use in analytics (Karagiannis et al., 2013). The ETL process is the backbone of the OLAP system. The ETL process is more than just simple querying of data from source databases (Karagiannis et al., 2013). The ETL process extracts the data from the source systems, brings together data from multiple systems into a new schema, and stores the data into the OLAP system (Karagiannis et al., 2013; Machado et al., 2019). The workflows involved in ETL can range from simple, fast-performing operations to very complex, long-running processes (Machado et al., 2019). There are two basic strategies for loading data into the OLAP system. The most common is batch-orientated methods, where the data are loaded into the system one or more times during the day (Machado et al., 2019). The second method is to stream the data into the OLAP system as it arrives in the primary system (Machado et al., 2019). The method used depends on the source of the data to be loaded into the system. When loading data, either by batch or streaming, it is important that the ETL process has a minimal impact on the source systems so user performance is not impacted (Machado et al., 2019).

BI&A 2.0. With the introduction of the commercial worldwide web, the type of information gathered by companies changed (Chen et al., 2012). The type of information generated by the new systems is not structured data that can easily be placed in an RDBMS and analyzed by traditional reporting tools (Chen et al., 2012). The type of data generated from these new web platforms is classified as unstructured data (Goh & Sun, 2015).

Unstructured Data. Unstructured data are classified as data with no defined structure, in contrast to data stored in an RDBMS system using a structured design. Sources of unstructured data within an enterprise include e-mail, failure reports, customer reviews, audio, photos, videos, and software logs (Chen et al., 2012; Kassner et al., 2015; Zhuang et al., 2016). In addition to data collected by organizations, data can be collected from external sources that contain unstructured data such as blogs, wikis, and social networking sites (i.e., Twitter and Facebook) to gain insights from areas not typically collected by an organization (Bhattacharjya et al., 2018). It has been estimated that unstructured data can be upwards of 80% of the data owned and stored by an enterprise and is one of the fastest-growing types of data an organization collects (Bhattacharjya et al., 2018; Kassner et al., 2015). Because of the lack of a defined structure within the data, unstructured data can be hard to analyze; therefore, many organizations do not take full advantage of the wealth of data available to them (Kassner et al., 2015; Korfiatis et al., 2019).

Unstructured data have challenges when being stored for future retrieval (Zhuang et al., 2016). Since traditional RDBMS systems require data to be in a structure, new types of storage systems have been developed to store unstructured data in a form that is retrievable in the future and capable of storing any relationship data or metadata to help retrieve the unstructured data

(Bacchelli et al., 2017; Zhuang et al., 2016). These platforms typically rely on a NoSQL platform or a distributed platform, such as Hadoop or Storm (Zhuang et al., 2016).

Analyzing Unstructured Data. When analyzing unstructured data, one must take great care when evaluating the quality of the unstructured data (Bacchelli et al., 2017). Because this data does not follow any structure, methods must be derived to ensure the quality of the data before analytics can be reliably performed on the data (Bacchelli et al., 2017). With the advancements in technology, data quality has become an ever-increasing issue within industry and academics (Vandepitte et al., 2015). Quality control tools help ensure data are appropriate and in the form needed for analytics. Quality control tools can assist in data format checks, verifying completeness, and assessing validity (Vandepitte et al., 2015). Unless data are accurate to the standard defined by the organization, reliable, relevant, and are fit for the use defined by the question at hand, the data may not produce the results required by the analytics, thereby giving erroneous results and leading to business decisions that can have disastrous consequences on the business (Muthee et al., 2018). Poor quality data can lead to providing misleading performance metrics to management. The incorrect performance indicators can lead management to make bad decisions that can result in putting resources into some initiative that should be eliminated or eliminating a project that may be performing well (Muthee et al., 2018).

After the data have been sufficiently cleaned, a number of methods and software can be utilized for analyzing the data. A key tool for analyzing unstructured data is text analytics (Zhuang et al., 2016). Tools such as MapReduce work with the storage engines (i.e., Hadoop and Hive) to process the data to return results mined from the stored unstructured data (Zhuang et al., 2016). For example, an open-source package called RapidMiner has text processing capability built-in to analyze data from unstructured sources (Goh & Sun, 2015). Many of these tools can also combine results from unstructured data analysis with structured data to gain additional insight into the enterprise data to assist in business decisions (Goh & Sun, 2015; Zhuang et al., 2016).

BI&A 3.0. With the development of mobile phones and the spread of sensors on various types of equipment throughout the world, a new generation of BI&A has emerged (Chen et al., 2012). BI&A 3.0 data sources include mobile phones, RFID tags, bar code readers, Internet of Things based sensors, and other similar devices (Chen et al., 2012). Together with Big Data, the data produced by the BI&A 3.0 devices allow organizations to analyze data in quantities not possible until recently. These new resources available to organizations allow the incorporation of BI&A 3.0 analytics into applications that feed analytical requests from the business to make informed decisions based on data generated by these new types of devices (Roßmann et al., 2018).

Mobile Data. Mobile data allows researchers and businesses to collect data like never before. With the availability of cell phones and wearable sensors, researchers can track participants' data with ease and with little intrusion into the participant's day-to-day life (Savage, 2015). Before the smartphone and wearable sensors, researchers were lucky to have a few hundred participants; however, with the new technology, one researcher was able to include over 40,000 participants in a study (Savage, 2015).

Radio Frequency Identification. Radio frequency identification (RFID) technology has opened up a new set of possibilities in tracking products and people (Xiao et al., 2018). As small devices attached to objects, the technology allows products to be tracked through RFID readers placed throughout warehouses, manufacturing plants, and retail stores (Zhou et al., 2017). With the cost of an RFID tag below 10 cents per tag, the reality of tracking high-cost items has become a reality for organizations that warehouse many thousands of items over millions of square feet of space. The usage of RFID tags has lowered the error rate of inventory of organizations through the automated reading of inventory levels and locations, replacing the manual system of counting each item and recording it into a computer system by hand or tracking on paper (Shahzad & Liu, 2015).

RFID devices can save organizations large sums of money by replacing manual recording and tracking of inventory with an automated system allowing the physical location of a product to be traced through a computer system (Zhou et al., 2017). In the retail world, RFID tags and readers have allowed companies that rent modes of transportation to set up return locations throughout the city of service and reduce the inconveniences of returning to a central point. The RFID and remote return locations can allow the company to analyze the movement of the vehicles to place them at convenient locations for the customer when they are required, increasing the chance that a vehicle is located where the customer need is in demand (Zhou et al., 2017).

Internet of Things. The term Internet of Things was coined in the early 2000s at the Massachusetts Institute of Technology, having origins in the RFID research being conducted at the time. Since then, the definition has been expanded to include a global network of sensors based on internet technologies that interconnect to provide information from sensors, audio, and video devices that provide data to a host system for storage and analysis (Wortmann & Flüchter, 2015). A combination of IoT and Industrial Automation and Control Systems (IACS) produced the Industrial Internet of Things, which is utilized in industrial and utility-based settings monitoring and controlling industrial equipment (Boyes et al., 2018). A common place to store the data produced by the Internet of Things and related technologies is in the cloud. Cloud infrastructure offers the technology to store and analyze the vast amounts of data produced by IoT devices (Čolaković & Hadžialić, 2018). Depending on the frequency of the reporting interval of the device, onsite technology may not be able to handle the bandwidth required to transport and store the data produced by IoT devices (Čolaković & Hadžialić, 2018).

BI&A 4.0. The development of technologies to support the era of artificial intelligence has ushered in BI&A 4.0. BI&A 4.0 builds on the usage of big data in BI&A 3.0 and adds the technologies of artificial intelligence. Artificial intelligence systems, combined with big data processing and analytics, allow technology to automatically make informed decisions without human intervention (Chen & Sun, 2018).

For example, in the medical field, artificial intelligence is being utilized to diagnose diseases that may be missed by traditional methods. The artificial intelligence modules are trained on the specialty data required for the diagnosis of the disease. Utilizing methods such as deep learning or other pattern matching algorithms, the machine can identify patterns in the test images or results and alert the doctors to the possible presence of a disease (Jha & Topol, 2016; Krittanawong et al., 2017).

Autonomous driving is another area using technologies from BI&A 4.0 software systems. When driving autonomously, a vehicle must be trained to obey the traffic laws and avoid crashes by using computer vision and artificial intelligence methodologies. These technologies are trained on big data platforms, and the trained models are placed in vehicles enhanced with the artificial intelligence hardware and software to control the vehicle (Hengstler et al., 2016). In addition to vision, the vehicle may be enabled to take commands from the passengers via voice command. The vehicle must be enabled with voice recognition and the ability to understand the voice commands and turn the voice command into actions taken by the vehicle in response to the occupants' command (Phelps, 1986).

Big Data

The amount of data generated in organizations today has been compared to the discovery of oil over 100 years ago. The unrealized value contained in the data can unleash new areas of profit for an organization by providing the company a competitive advantage over others in the industry who do not possess the same type of data (Sun, Cegielski, et al., 2018). Big data allows managers to make decisions based on data analysis instead of tribal knowledge and intuition (Hofmann, 2017).

5Vs. Big data was defined as a technology in 1997 as large data sets that challenge computing technologies' computational resources, including main memory and local and remote storage (Hopkins & Hawking, 2018). The data sets classified as big data are much larger than traditional data processing systems can handle and have complexity that makes it very difficult to fit the data into a traditional processing system. The data within the big data set are generated by a machine, human, or nature and recorded for later analysis. Big data is characterized by the five V’s: volume, velocity, variety, value, and veracity (Kalbandi & Anuradha, 2015; Rehman et al., 2017).

Volume. The big in the title big data refers to the volume of data being stored and analyzed. The volume involved in a big data set could be in the petabyte range or more, depending on the organization's size and the collection methods. The big data available to organizations allows for a more in-depth exploration of the data set to find patterns previously not available for discovery with conventional technology. Companies that can take advantage of the volume of data generated by big data can have a competitive advantage in the marketplace if they make discoveries early in the product cycle and take advantage of those discoveries (Johnson et al., 2017). This data can come from a variety of sources, including mobile phones, robots, satellites, drones, IoT, and many other similar types of devices. It is estimated that data volume will exceed 30.6 Exabytes by the end of 2020 (Rehman et al., 2017). The majority of the data growth is generated from unstructured data. Unstructured data is typically human content. Human content is classified as videos, photos, movies, financial transactions, e-mails, telephone conversation transcripts, tweets, and logs from systems generated by human interaction. The vast majority of unstructured data is generated by social media platforms, such as Facebook, Twitter, and YouTube (Khan et al., 2014).

Velocity. Big data velocity is the speed at which the data are ingested into the data warehouse infrastructure. The growth rate of data depends on the system and the spread of the sensor network or sources of the data (Al-Salim et al., 2018). The data coming into the data warehouse may be processed and analyzed in real-time to identify risks to the business, from customer churn to process failures within a production process. The streaming of this data requires a different processing type than traditional data processing, where high volume, low latency storage was required. Data coming into big data systems requires high-throughput computing to process the data in a timely fashion (Sun, Yan, et al., 2018).

Technologies available to apply to the big data system for high-velocity systems are Complex Event Processing systems and Stream Processing systems. These systems allow the real-time processing of data required in a modern big data system, processing data from many sources of data, such as sensors on a manufacturing system (Higashino et al., 2016). The speed of processing data is more critical to some applications than the volume of data that can be processed. The speed of processing can enhance agile decision-making in real-time. The increase of data processing velocity has been driven by the technological advances achieved through increased processing power along with the speed of the monitoring devices and the associated networks moving the data into the big data system (Alzyadat et al., 2019). Having high-velocity data allows the company to make decisions in real-time instead of looking back at historical trends, enhancing their ability to react to changes in processes or markets to serve their customers better. The support for high velocity of data within an organization can drive the organization to be more innovative with its customers, therefore serving the needs of the customer at a higher level than with traditional historical data ingestion and processing (Ghasemaghaei et al., 2018).

Variety. The type of data generated by systems and placed into big data warehouses vary greatly. The variety of data may be the most critical characteristic of big data processing (Abawajy, 2015). The data placed in a big data warehouse can be structured or unstructured data. A majority of the data in big data warehouses is considered unstructured data (Kalbandi & Anuradha, 2015). Much of the unstructured data does not fit well in a traditional relational database and, therefore, is hard to analyze with traditional tools (Nastasoiu et al., 2019). Some unstructured data can be classified as semi-structured data. For example, an e-mail has some structure in the form of the to, from, subject, and date sent attributes. However, the body of the message is unstructured data. Semi-structured data is typically processed using the same methodologies as unstructured data when analyzing the dataset's unstructured portion (Abawajy, 2015). The unstructured data can come from many different types of sources and require different processing techniques to analyze the data. The data may require different amounts of CPU, memory, and I/O requirements to process. This data may come from sensors, smart devices (i.e., smartphones and tablets), social media sources, e-mail, and documents (i.e., Word, PDF, and spreadsheets). Each type of source requires a different method for processing the data (Nastasoiu et al., 2019).

The variety of data processed within big data processing can be classified into four categories. Data are processed in batches, streaming, interactively, or by graph processing (Abawajy, 2015). When data are processed offline, it is processed in batch mode. Batch mode is used to find new patterns in existing data. A common tool used for batch processing is Hadoop. Streaming data are usually processed in real-time without storing the data first. Examples of stream processing include fraud detection and network intrusion detection. Using tools such as Apache HBase and Google BigQuery are examples of interactive processing. These tools process the data as the user submits the query to the engine to return results from the data set. Graph processing evaluates data based on relationships between objects. These relationships describe how the objects are related to each other. The relationships allow the query engine to process the data and return results based on the objects' relationships. The relationships represented by graph databases are more complex than those represented by traditional relational databases (Abawajy, 2015).

Value. For big data to have an impact on business or society, the data in the big data warehouse must be turned into valuable information (Kalbandi & Anuradha, 2015). Value for big data warehouses can be defined as social value or economic value. Social value for big data use includes uses by government, healthcare, education, or public safety to enhance the health and safety of the population. Use in this fashion would include preventing crime, support the wellbeing of the population, or improve national security. The economic value of big data improves the profit, growth, and competitive advantages of an organization. Using big data for economic value adds to the value of an organization giving the stakeholders more value for the investment in the big data solution (Günther et al., 2017).

Through economic value, Brinch (2018) discovered that big data allowed the business to make better decisions through their processes when using big data. He also found that using big data will change business processes when using the big data warehouse to explore and exploit the data (Brinch, 2018). Using big data to make business decisions positively affects the outputs and quality of new jobs and tasks. If the business process management tasks are used in conjunction with big data for decision-making processes, the data's value is realized through increased utilization of the big data system (Brinch, 2018).

Big data's social value is a new category of value for big data (Arnaboldi, 2018). For big data to provide social value, the data scientists and decision-makers must work together to derive value from the data housed in the big data warehouse. The data scientist's role is to extract the information from the data and present it to the decision-makers. The decision-makers work with the data scientist to refine the data and extract the most value from the datasets (Arnaboldi, 2018).

Veracity. Big data's veracity refers to the trustworthiness, availability, accountability, and authenticity of the data. The data's veracity is especially important with the volumes of data within the big data warehouse since the data may be utilized to build decision systems that are automated and have little or no human involvement in the decision process (Jamil et al., 2015). Especially with data coming from the IoT environment, data quality issues can arise from data acquisition. Sources of data errors include environmental factors, vandalism, faulty sensors, resource constraints, and security vulnerabilities (Perez-Castillo et al., 2018).

One of the main problems with big data is the lack of trust in the data collected and stored in the data warehouse. Before the proliferation of sensors and other big data sources, data were collected within a controlled process, and the data were easier to verify as accurate (Al-Jepoori & Al-Khanjari, 2018). Processes should be put into place to ensure the data entering the data warehouse are of the best quality to ensure the analysis resulting from the data is a high-quality analysis. Data governance can help ensure that data are of the best quality available for future analysis. If the proper data governance is put into place, then the data can be trusted to be of the best quality available, and if not, the people accountable for the data can be notified, and corrections are taken to ensure the data are corrected or eliminated from the data warehouse. A data governance plan puts into place the standards for quality data for the organization to follow when acquiring and cleaning data for the data warehouse. Data governance plans should identify data usage policies, who has access to the data, any quality metrics, and the units of measure if applicable (Koltay, 2016).

HACE Theorem. The HACE Theorem further describes big data as "large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data" (Wu et al., 2014, p. 98). Unless the analyst has the full picture of the data being analyzed, bias may be introduced into the analysis. Using only fragments of the data for analysis and failing to fully understand the proper perspective, the data may lead to an incorrect or misleading conclusion based on the limited understanding of the data (Drosio & Stanek, 2016).

Large Volume, Heterogenous Dataset. A large volume, heterogeneous dataset can describe one object. For example, the human body can be described by the typical demographic information. However, upon further analysis, x-rays or other imaging sources can be used to view the same body. Further examination of the body can be accomplished through the analysis of DNA. These various forms of information describe the same object, a human body, but have very different perspectives. A large amount of data with different schemas can make it difficult for an analyst to thoroughly analyze an object when the data comes from many heterogeneous datasets (Wu et al., 2014).

Autonomous Sources. Autonomous sources with distributed and decentralized control add another complication to the analysis of big data (Drosio & Stanek, 2016). Autonomous and decentralized sources are resources that provide data on devices that can stand alone and not rely on a centralized system to function. The autonomous and decentralized nature will allow the system, as a whole, to continue functioning even if one resource is not available (Wu et al., 2014). Examples of autonomous sources include cameras, sensors, retail transaction recording devices, smartphones, and other smart-enabled technology (Ma-Lin et al., 2018).

Relationships and Complexity. As data grows in size, the relationships become more complex and evolve over time (Wu et al., 2014). As the number of sensors grows, feeding big data through the technology of the Internet of Things, the relationship between the objects becomes greater. The amount of data collected and the diverse number of sensors now available to monitor a single object or process have grown considerably recently. The complexity of the relationship between the data becomes more difficult to manage (Gil et al., 2019).

Components of Big Data. Big data components can be broken down into two major categories. Compute-based models store their data on physical storage and recall the data from the storage as the data are processed. In-memory models store the data in the computer's main memory to retain fast processing not allowed by compute-based systems that store the data on physical storage (Rao et al., 2019).

Compute Based Models. Compute-based big data solutions can be broken down further into three categories: Massively Parallel Processing (MPP) systems, MapReduce (MR)-based systems, Bulk Synchronous Parallel (BSP) systems (Rao et al., 2019). MPP-based systems are implemented in software such as Teradata, Apache HAWQ, EMC Greenplum, and IBM Netezza. The data in MPP systems are divided among multiple systems that do not share physical storage. The individual systems process the data and return results after processing the data (Rao et al., 2019). MapReduce based systems are based on mapping tasks and reduce tasks. The systems process requests in parallel and summarize the data, and return it to the reduce task for determining the final results (Rao et al., 2019). BSP systems process data in iterations on individual nodes in parallel with information synchronized between the nodes during processing (Rao et al., 2019).

In-Memory Models. In-memory processing systems are up to 100 times faster than systems based on physical storage. These systems keep data in the main memory of the system for faster result processing. These systems typically process real-time data and streaming data. These systems work similarly to MMP systems but are limited in capability by the system's amount of main memory (Rao et al., 2019).

Advanced Analytics

Advanced analytics, also known as process analytical technology, is a combination of sensors and software used to measure, analyze, monitor, and control the production process (Kornecki & Strube, 2018). Advanced analytics allow organizations the ability to scale their process by controlling the factors within the process for better efficiency. The advanced analytic systems allow for real-time process monitoring for feedback on the process as well as improved consistency throughout the process (Watson et al., 2016). Many industries, such as the food industry, pharmaceuticals, manufacturing, and the energy industry, use advanced analytics to monitor their products (O'Shea et al., 2019). Furthermore, in some industries, such as pharmaceuticals, the increased use of advanced analytics was brought on by government requirements for quality control within the manufacturing process to ensure the highest quality product possible. The new requirements for higher manufacturing standards from the regulatory bodies required the manufacturers to have systems in place to monitor the process in order to bring the products to market (Nakano et al., 2018).

Supervisory Control and Data Acquisition. Industrial control systems like Supervisory Control and Data Acquisition (SCADA) systems are vital to controlling the world's industrial processes (Boyes et al., 2018; Castellani et al., 2017). SCADA systems allow a centrally located operator to monitor all the processes within the manufacturing plant or other industrial settings. For example, in the oil and gas industry, the centralized SCADA system allows the operator to start and stop compressors remotely, set control point values, and open or close valves (Boyes et al., 2018). The control systems the operator is interacting with through the SCADA system are usually located remotely throughout the plant or geographically dispersed in the case of oil and gas operations. The SCADA system is composed of two major parts; the control systems and the remote devices, which contain electronics to control the physical device and monitor the results to send back to the central SCADA system (Boyes et al., 2018).

However, SCADA systems are limited in their ability to provide advanced analytic within their capabilities. SCADA systems typically lack the processing capabilities to handle the advanced processing required of advanced analytics. The data processing requirements tend to overload the processing capabilities of the SCADA systems. SCADA systems are also limited in the types of sensor data brought into the system to keep the processing load from overwhelming the system and preventing the operator from controlling the system (Shinozuka et al., 2015). Another factor leading to the limitation of the SCADA system to process the wide variety of data required to make decisions on large datasets is the lack of attention from the process control industry. Since the mid-1990s, after failures of artificial intelligence efforts in the control industry, the process engineering efforts have not been restarted since the newer technology has been introduced into the artificial intelligence and machine learning fields (Dong & Qin, 2018).

Process Monitoring. Advanced analytic systems incorporate high-frequency data from many different sensors to analyze the process and show the operator recommendations or produce alarms when the process exceeds set limits or trends (Fissore et al., 2018). Keeping the process under control reduces the amount of waste generated from a product that is not usable or could require re-work, therefore, wasting raw materials, energy inputs, and processing time (Sánchez-Camargo et al., 2019). Advanced analytic systems monitor the process, apply multivariate analysis, and can control the process (O'Shea et al., 2019).

Reduction in Other Costs. In addition to the increased visibility into the current state of the processing environment, advanced analytic systems can reduce the need for offline activities, such as lab testing, reduction in recalls of products, more consistent and higher quality product, and reduce environmental waste by maximizing processing efficiencies (O'Shea et al., 2019). In addition to the costs associated with the normal production process, an advanced analytic system can reduce costs indirectly by optimizing the production process and allowing the manufacture to produce more product with the same equipment by optimizing the process to allow for better utilization of the existing equipment, therefore, avoiding the additional capital outlay required to expand existing facilities or build new facilities to account for additional production (Watson et al., 2016).

Processing Data. Advanced analytic systems ingest data into the data stores and similarly process the data that SCADA systems do, but typically do not control the underlying processing systems. The advanced analytic systems apply the rules set for the process and apply mathematical correlations to the data to monitor the system parameters (Pathak & Rathore, 2017). Advanced analytic systems usually work in real-time to analyze the data and provide feedback to the users, providing almost instantaneous feedback about the process's current state (O'Shea et al., 2019).

Some advanced analytics systems incorporate machine learning into the system to apply models to the process for a more accurate prediction of process variances. The machine learning models allow for processing more variable data from the analytical system, allowing the system to better predict process variation than simple multivariate analysis can typically perform (Nagy et al., 2019).

Machine Learning

As an idea, machine learning has been in the technological vocabulary as early as the 1940s; however, early machine learning systems were no more than expert systems trained with if-then-else type logic to produce an answer for a query based on the data input into the system. While the first conference on machine learning occurred in 1956, the technology to make machine learning a reality only came into existence in the early part of the 21st century with the introduction of technology from Google and others (Haenlein & Kaplan, 2019).

Machine learning is a technology used to find patterns in data through general-purpose algorithms, then make predictions from the patterns discovered by the chosen algorithm (Bzdok et al., 2018). In contrast to expert systems that make predictions on a set of rules, machine learning can process data with many variables, find patterns in the data, and link the data in interactive and nonlinear ways to predict the outcome based on the vast data supplied to the machine learning algorithm. However, when using machine learning to train models and produce predictions, the model may overfit the data and may produce a model that is not general enough to apply to a wide set of data beyond the training set (Obermeyer & Emanuel, 2016).

Machine Learning Types. Machine learning can be divided into three types: supervised, unsupervised, and semi-supervised. These types indicate how the data are managed and how features are processed through the selected algorithms (Cai et al., 2018).

Supervised Learning. Supervised learning consists of a data set with one column, or label, identified as the answer described by the rest of the columns of data. The supervised machine learning algorithm attempts to find the patterns to build a model to predict the labeled data as accurately as possible (Elforjani & Shanbr, 2018). There are two types of prediction methods within supervised learning: classification and regression (Louridas & Ebert, 2016). Classification algorithms attempt to classify the data into categories, while regression uses mathematical formulas to attempt to fit the data into a pattern to predict the outcome of the model's inputs (Herbrich & Graepel, 2015; Louridas & Ebert, 2016).

Unsupervised and Semi-Supervised Learning. On the other hand, unsupervised learning does not provide an answer, or label, to the machine learning algorithms. The chosen unsupervised algorithm attempts to find patterns in the data to provide a prediction (Elforjani & Shanbr, 2018). Semi-supervised learning is a hybrid of supervised and unsupervised machine learning. Data presented to semi-supervised machine learning will have part of the data labeled and part of the data unlabeled. The semi-supervised algorithms attempt to use both labeled and unlabeled data to find patterns in the data to provide a prediction (Naganathan et al., 2016).

Feature Engineering. Feature engineering is a precursor step to introducing data to a machine learning algorithm. Feature engineering is the process of selecting the data necessary to attempt to predict using a machine learning algorithm (Li et al., 2017). Li et al. (2017) described feature engineering as a fundamental process that defines the numerical fingerprints by selecting the data that describes the desired label. Feature engineering includes selecting the data, cleansing the data, and selecting the features (Dai et al., 2020). However, feature engineering is more than selecting data from the data warehouse. When engineering features for predicting with machine learning, features may need to be constructed from the data to produce a new factor from a set of existing features to help fit the data for the chosen algorithm (Dai et al., 2020). When engineering features are used in a machine learning problem, Principal Component Analysis (PCA) is one method to find the best features for training the model. Using PCA, the set of features is reduced to the most important features, thereby reducing the resulting model's complexity (Li & Tao, 2013).

Overfit and Underfit. Over and underfitting the data is a common pitfall when developing a regression model in supervised machine learning (Lever et al., 2016). Overfitting in a machine learning training session occurs when the resulting model is more complex than necessary. When overfitting a model, the algorithm takes into account noise to try to make the model fit all the data instead of a generalization of the data. The model may work very well with training data but perform poorly with test data (Lever et al., 2016). The overfit model treats outliers as part of the model's structure instead of outlying data or noise in the sampling (Ord, 2020). The opposite is true when underfitting the data. The algorithm presents a model that is too simple for the data presented (Lever et al., 2016). Changes in the underlying data may change a fit model into an over or underfit model over the model's lifetime. For example, changes in economic conditions may introduce outlying data into the dataset and cause a model to overfit the data and produce a model that will not perform well outside of the context of the training data (Ord, 2020).

Ethics of Machine Learning. There is a danger in misusing machine learning while trying to predict the outcome of input data. Careful consideration should be taken when trying to implement machine learning into the business context. For example, Facebook and Cornell University implemented an experiment to manipulate the news feed of users based on their posts' emotional content without the knowledge and consent of the users. This experiment resulted in bad press for both Facebook and the university and potentially harmed the users involved in the experiment (Mittleman & Druckenmiller, 2019; Stark, 2018).

Machine Learning in Natural Gas. Machine learning is utilized in forecasting the demands in many countries throughout the world; however, the literature focuses on either machine learning techniques (Merkel et al., 2018; Wei et al., 2019) or local distribution companies' (LDC) forecasting needs (Akpinar & Yumusak, 2016; Beyca et al., 2019; De & Gao, 2018; Izadyar et al., 2015; Papageorgiou et al., 2020). There is a gap in the literature describing natural gas forecasting for interstate natural gas pipelines in the United States.

Employee Training

Training has a direct effect on an employee's performance towards the organization's goals. As an essential activity, training plays an essential role in the competence and efficiency of an employee (Farhan Zeb & Imamuddin, 2018). Well-trained and committed employees contribute to the success of an organization through more efficient work and improved performance. Conversely, the lack of training contributes to unsuccessful initiatives (van Assen, 2021). Training is a purposeful activity to increase an employee's skill level and knowledge to perform their job more effectively. Training has many benefits though developing thinking skills and creativity to make better decisions and increase productivity (Sendawula et al., 2018).

Conclusion

The technological advancements that have led to the implementation of advanced analytics and machine learning systems have been in the making since the 1700s, beginning with the first industrial revolution. As technology has progressed from that point with the introduction of electricity, the invention of electronics, and the introduction of electronics into the manufacturing process, the need to analyze the various sensors' data has been increasing. The advancement of business intelligence and analysis, along with the reduction in cost and advancement of computing resources to process the data, has allowed business intelligence to progress from simple reports to high-speed analytics and machine learning to give feedback to the equipment operators and management almost in real-time.

Without leadership behind these efforts and the users willing to accept the technology and use it in their day-to-day jobs, the implementation of advanced analysis and machine learning can be challenging to accomplish. Transformational leadership and UTAUT will be the lens that provides the analysis of why advanced analytics and machine learning systems either fail or succeed.

Summary of Section 1

Section 1 lays out the foundation of the study. This section examines the background of the problem, problem statement, purpose statement, nature of the study, method, design, research questions, and conceptual framework, defining terms, assumptions, limitations, delimitations, and significance of the study. A review of professional and academic literature was presented describing the existing literature for business intelligence, big data, advanced analytics, machine learning, transformational leadership, and the Unified Theory of Acceptance and Use of Technology.

Section 2 of this study expands the study's foundation by examining a qualitative case study of the implementation of AAML. Management, front-line workers who utilize the system, and information system professionals will be studied to discover the effects of implementing an AAML system has on an organization.

Section 2: The Project

Section 2 describes the structure of the qualitative case study approach used to discover the factors of the successes and failures of implementing an advanced analytics and machine learning system at an interstate natural gas pipeline. This section begins by defining the purpose of the project. The role of the researcher follows and includes how access to the participants was granted. The research method and design are laid out, and a description of the population and sampling methods are defined. Following the discussion of the population and sampling methods, the data collection procedures are described. Within the data collection section, the researcher as an instrument is discussed. This discussion includes the interview guides and components of the interview. After the data collection section, the coding procedures are discussed. Finally, the reliability and validity of the data are examined.

Purpose Statement

The purpose of this research project is to discover the improvements realized by implementing advanced analytics and machine learning (AAML) and the reasons for failure to fully utilize such a system for interstate natural gas pipelines in the United States. While the benefits seen by implanting an AAML are great, implementing AAML systems can be a complex and expensive task for organizations, and past failures of advanced analytics have caused many in the process control field to ignore advanced analytics (Dong & Qin, 2018; Giannino et al., 2018; Lechevalier et al., 2019). Further hampering the full utilization of AAML, many supervisory control and data acquisition (SCADA) systems cannot process the amount of data required for AAML (Shinozuka et al., 2015).

Role of the Researcher

In case study research, the researcher may take on many different roles during the research. The researcher may be an insider to the group being studied or maybe a complete outsider to the group (Unluer, 2012). Whether an insider or outsider, a qualitative researcher is trying to understand the situation from an insider or emic perspective (Hancock & Algozzine, 2017). It does not matter if the researcher is an insider or an outsider; the researcher is one of the main instruments used during a case study research project. The researcher interviews participants, conducts surveys, gathers and analyzes documents, observes the environment, and writes the report detailing the ideas and observations (Stake, 1995). The researcher must be careful not to allow themselves to inject their bias into the research. This bias may come from cultural position or closeness to the research topic (Marshall, 2016).

Before the research could begin, the researcher was responsible for designing a proposal, submitting the proposal to the IRB, and obtaining approval from the IRB to conduct the proposed research. During this research project, the researcher was responsible for identifying the participants to interview, conducting the interviews, collecting any supplemental documents, and analyzing the collected information. The researcher is also responsible for securing the collected data and safeguarding the data to protect the interview subjects' confidentiality.

Identifying the subjects for this research project and gaining access to the participants was obtained through industry contacts, either through the researcher’s employer or through industry contacts. The researcher interviewed each participant, transcribed the recording of the interviews, gathered and coded the information, and analyzed that information plus other data to gain insights into the problem presented in this research.

Participants

The participants for this research project were solicited from the interstate natural gas pipeline industry. The participants were employees or retirees of the selected companies represented in the case study. The participants were located within the United States and were all adults (18 or older) at the time of the interviews. Participants were not part of a vulnerable population. The participants were assumed to be knowledgeable about the subject through their current or former employment by an interstate natural gas pipeline. Participants were selected from all levels of the organization. The researcher ensured the participants were able to opt out of the research at any time, and the responses were kept confidential. The researcher was an insider or partial insider to the participants, so gaining access to the participants was not difficult for this research project. The researcher either worked directly with the participants or had contact with the participants through industry groups. Prior to the commencement of any participant contact, the researcher obtained approval from the Internal Review Board.

Research Design and Method

In Section 1, the problem statement, research questions, literature review, and explanation of research methods lead to the selection of the appropriate methodology for the research design of this study. With the research questions striving to find out how AAML helps an organization, why implementation succeeds or fails, and the factors for success or failure, a case study was the appropriate methodology to answer those research questions. The reasoning behind the choice of a case study is discussed in this section.

Discussion of Method

Qualitative research is utilized when the research study aims to understand the meaning of a human or social problem (Creswell & Creswell, 2018). Qualitative methods are typically used to study complex problems set in the natural world, use multiple methods during the research, and evolve as research is conducted (Marshall, 2016). Qualitative research was conducted in the field at the participant site using the researcher as the primary data collection instrument. The researcher focused on the participants' multiple perspectives and meanings, used multiple methods to collect data, and the design evolved as the research was conducted. Qualitative research strives to present a holistic picture of a complex research topic (Creswell & Poth, 2018). By contrast, a quantitative research project is used to study the relationship between variables using instruments (typically returning numerical results) to collect information and use statistical methods to find relationships between them (Creswell & Creswell, 2018).

The number of participants was limited for this study because of the narrow focus of the interstate natural gas industry. There are approximately 210 interstate natural gas pipelines in the United States and only 30 major interstate natural gas pipelines (EIA, n.d.a, n.d.b, n.d.c). A qualitative method was selected because of the need to understand the participants' complex interactions required to implement an AAML (Creswell & Poth, 2018).

Discussion of Design

Five qualitative research methods were available to conduct this research study, including narrative, grounded theory, phenomenology, and ethnography. Because the study focused on the how and why of real-world events, a case study design was selected (Yin, 2018). The exploratory nature of a case study leans towards answering the how and why questions of research. Since the environment cannot be manipulated, and the events under study were contemporary, a case study was the preferred method for this type of study (Yin, 2018). In addition to interviews, a case study allows the researcher to glean information from documents, physical artifacts, and direct observations (Stake, 1995; Yin, 2018). A case study typically focuses on one case or a small number of cases. Because there are only 30 major interstate natural gas pipelines in the United States, the limited number of possible participants for this research study relates well to a case study (Dul & Hak, 2007; EIA, n.d.c).

Summary of Research Method and Design

The purpose of this study was to discover the reasons for failure to fully utilize AAML and the improvements realized by implementing such a system for interstate natural gas pipelines in the United States. Because this research study was searching for the how and why behind the success and failures of implementing an AAML and the improvements to the organization after implementing an AAML, a case study was chosen as the method to understand these questions. The case study method is well suited to discover the how and why of a problem. The case study does not require control over events and typically focuses on contemporary events (Yin, 2018). Because this study focuses on a contemporary event and seeks to discover the how and why, the researcher deemed the case study method to be the most appropriate qualitative design to fit the research study.

Population and Sampling

The purpose of this research was to discover why AAML installations fail and what qualities help AAML installations succeed. Failures of natural gas pipelines can have severe consequences for the surrounding population, downstream customers, and the environment in general. AAML systems can help predict natural gas utilization along with the pipeline system and help monitor and predict equipment failures on the pipeline system (Farzaneh-Gord & Rahbari, 2016; Shinozuka et al., 2015). Being a qualitative case study, this research project will focus on a limited number of organizations that have implemented an AAML system to monitor and predict natural gas flow along with their pipeline system.

Population

In case study research, the researcher’s population choice is based on the bounds set forth by the proposed case study (Stake, 1995). A case may be one person, group of people, an organization, or a community. A case is also bounded by a system (or object) and a timeframe (Stake, 1995, 2010).

By using the case study methodology, a maximum of two organizations that have implemented an AAML was the focus of the research. The researcher identified an organization that implemented an AAML. There are 30 major natural gas pipelines to explore within the United States (EIA, n.d.c). The researcher has been employed in the natural gas industry for 20 years and has contacts through current employment for an organization that has implemented an AAML.

It is essential to choose people with experiences relative to the phenomenon being studied (Merriam & Tisdell, 2016; Stake, 1995). The population was limited to the organization's leadership, Information Technology professionals, technical personnel, and office personnel who used an AAML system with the organization. All participants were at least 18 years old. Employees under 18 years old and those who did not have experience with the AAML no matter the length of employment (excluding leadership), were excluded from the study. It was important to have participants with direct experience with the AAML to participate in the study.

Sampling

Both quantitative and qualitative methods use sampling to maximize the efficiency and validity of the study. The sampling method chosen must be consistent with the goals and assumptions of the chosen method (Palinkas et al., 2015). In quantitative research, sampling is a method to select participants from a large pool of a population randomly. The population in quantitative sampling is usually chosen at random (Patton, 2015). However, in qualitative research, the population size may be as small as one when performing a single case study (Patton, 2015). The sampling techniques used for this research study began with purposeful sampling and transitioned to snowball sampling after the initial interviews of known participants.

Purposeful Sampling. Purposeful sampling is a popular technique in qualitative research. Purposeful sampling is used to identify information-rich participants to study with the limited resources and time a case study presents (Palinkas et al., 2015). Studying information-rich participants provides an insight into the case that cannot be discovered through the use of generalizations provided by survey results gathered through quantitative research (Patton, 2015). Purposeful sampling also allows for multi-stage sampling to gain deeper insight from the limited population available for a case study (Duan et al., 2015; Merriam & Tisdell, 2016).

Snowball Sampling. With the open-ended nature of qualitative research, the sample population is built as the research is conducted. As interviews are conducted, the researcher inquiries about who else may provide information about the subject. This technique is called snowball sampling (Patton, 2015). As the interviews progress, the sample size grows like a snowball rolling down a hill (Merriam & Tisdell, 2016).

Data Saturation. The sample size with a qualitative research project depends on many factors, including the question researched, the number of available participants, and the uniqueness of the information provided by the participants. Sampling continues until the researcher no longer discovers new information from the participants (Creswell & Creswell, 2018; Merriam & Tisdell, 2016). Morse (1995) defined saturation as “when the domain has been fully sampled – when all data have been collected – then replication of data occurs and, with this replication… the signal of saturation” (p. 148). Morse (1995) also emphasized that the data's quality is more important than the quantity gathered. The rare gem of information may be what puts the rest of the data into focus (Morse, 1995). While some authors like Creswell and Poth (2018) suggested 20-30 interviews may be necessary, Creswell and Creswell (2018) “one stops collecting data when the categories (or themes) are saturated: when gathering fresh data no longer sparks new insights or reveals new properties. This is when you have an adequate sample” (Creswell & Creswell, 2018, p. 186).

Summary of Population and Sampling

The population of a case study is bounded by the parameters outlined in the research study (Creswell & Creswell, 2018; Stake, 1995). It is vital to choose the population that represents the phenomenon within the case study's bounds (Merriam & Tisdell, 2016; Stake, 1995). This case study involved participants 18 years of age or older who were current or former employees of the interstate natural gas pipeline chosen for this study. Purposeful sampling and snowball sampling was used to seek the participants, and interviews were conducted until the data were saturated.

Data Collection

The goal of a case study is to discover the “how” and “why” of a contemporary event that the researcher cannot control (Yin, 2018). The sources for data to be collected for a case study can be quite extensive. Sources include interviews, archival records, documents, and direct observation (Yin, 2018). The primary source of this research project was interviews. This data allowed the researcher insight into why AAML system implementation fails and what aspects contribute to the successful implementation of AAML systems.

Instruments

Because interviews are the primary tool used for this research project, the researcher is the primary research instrument for data collection in this research project (Creswell & Creswell, 2018). In addition to the interviews, the researcher collected documents from the participants, as necessary. The researcher analyzed the data gathered through the interviews and document collection to answer the research study's research questions.

Data Collection Techniques

The interview guide acted as a guide for the interview stage of the research project. The interview guide contained an introduction, which included a statement that the participant was voluntarily participating in the study and could withdraw at any time. The participant was also informed that the responses were anonymous. The interview guide contained questions directly related to the research questions or was demographic in nature (i.e., job title, time with the company, and experience with the AAML system). While there were questions set forth by the interview guide, the researcher did not restrict the participants from speaking freely. The researcher asked follow-up questions, as necessary, based on the path the participants took during the interview. The interview guide is included in Appendix A. This project was completed during a pandemic; therefore, the interview process was slightly different from normal times. The interviews were recorded via audio or video as technology and distance allowed. Other documentation, such as documents and PowerPoint slides, were collected as they were made available by the participant.

Data Organization Techniques

All interviews were recorded during this research project. Each file was named so the researcher knew who the participant was and the date the interview was performed. Any electronic documentation was similarly named to allow for easy retrieval. Any physical documentation was scanned and stored in a similar manner as the data that were originally electronic documentation. All electronic documentation were secured in a password-protected folder structure to organize and secure the data. Any data backup was similarly stored in a secure manner appropriate to the backup method.

Summary of Data Collection

To satisfy the goal of a case study to answer the “how” and “why” of a contemporary event where the researcher does not control the environment, the researcher acted as the research study's primary tool. Interviews were conducted according to the interview guide, documents collected, and other artifacts were secured as available. The data were stored in a secure password-protected location using a file naming system to quickly find the data as needed.

Data Analysis

According to Stake (2010), conducting research requires both constructing and deconstructing when analyzing the data collected from the field. The process of analyzing data does not begin at a particular phase during a case study; rather, analysis happens throughout the entire process (Stake, 1995). When analyzing data, the qualitative researcher uses ordinary methods to break the data down and put it back together to tell the story (Stake, 1995). This analysis includes things known to the researcher and are common themes throughout the study or maybe mentioned only once during the interview process (Stake, 1995). Analysis during a case study is an iterative process (Stake, 2010).

Coding Process

This study used coding of the interviews and any appropriate documentation to discover the themes that emerged from the data. The first step in coding the data was to transcribe the interviews. After transcription, the interviews were coded using the software package Atlas.ti. This software allows the researcher to code the interviews and analyze the content in one package, including audio, video, and documents.

Summary of Data Analysis

Conducting case study research requires deconstructing the data presented to the researcher and synthesizing the information through construction (Stake, 2010). The researcher uses ordinary methods to break down the information by coding the information with the Atlas.ti tool (Stake, 1995). The analysis includes audio, video, and documents provided by the researcher. Using the Atlas.ti software, the common themes were discovered and synthesized to answer the research questions.

Reliability and Validity

A qualitative researcher strives to produce research that is dependable and valid (Krathwohl, 2009). The researcher has many tools available to produce dependable and valid research. Through the use of consistent methods, triangulation, negative evidence, reducing bias, and achieving data saturation, the researcher can produce reliable and valid research (Creswell & Creswell, 2018; Creswell & Poth, 2018; Krathwohl, 2009; Yin, 2018).

Reliability

Reliability can be achieved in qualitative research by using good quality devices to record the interviews, taking detailed field notes during observations, and transcribing recordings to include the timing of non-spoken content such as pauses (Creswell & Poth, 2018). Moreover, using software to assist in recording and analyzing data can increase the reliability of the data collected (Creswell & Poth, 2018). As software has improved over the years, the programs have become able assistants and reliable tools to improve qualitative data analysis (Yin, 2018).

Consistency. To increase reliability, data should be collected in a consistent manner throughout the data collection phase (Krathwohl, 2009). For example, the same interview guide should be used for all participants (Krathwohl, 2009). In addition to using the same interview guide for each participant, the researcher can enhance reliability by checking the transcripts for accuracy (Creswell & Creswell, 2018). When coding the transcripts, the researcher periodically analyzes the coding process to verify that no drift occurs in the coding process. A shift in the meaning of the codes can affect the transcript data's final analysis (Creswell & Creswell, 2018).

Saturation. If a researcher fails to reach saturation, then the study quality and content validity may be compromised (Fusch & Ness, 2015). Creswell and Poth (2018) defined saturation as the point when data collection no longer adds new information. They suggest that 20 to 30 interviews may be necessary for saturation. However, there is no one size fits all solution to saturation. The number of interviews and data collected from other sources for saturation depends on the study design and the phenomena under study (Fusch & Ness, 2015). Enough data has been collected when the amount of time spent interviewing outweighs the value of the information being returned. If a researcher finds additional interviews are not adding to the codebook or returning new information, the researcher may conclude that data saturation has occurred (Guest et al., 2006).

Validity

Creswell and Poth (2018) described the validation as using a “process for assessing the accuracy of the findings as best described by the researcher and the participants” by using multiple qualitative research strategies (Creswell & Poth, 2018; Location 8571). A qualitative researcher is seeking confirmability in their research (Creswell & Poth, 2018). Validation strategies from the researcher's perspective can include triangulation, disconfirming evidence, and clarifying researcher bias (Creswell & Poth, 2018).

Triangulation. To help increase the qualitative data's reliability, the researcher employs triangulation by verifying the data through multiple sources, such as interviews, documentation, researcher notes (Creswell & Creswell, 2018). The triangulation process involves the researcher establishing the data's validity by finding multiple paths through the data. These paths may involve multiple interviews, multiple types of sources representing the data point, or finding the data through multiple methods (Krathwohl, 2009). However, researchers must be cautious when triangulating data in qualitative research. During the process, the researcher may discover data that is inconsistent and contradicts previous data collected (Krathwohl, 2009).

Interviews. The primary source of information for this research project was interviews conducted with participants. The interviews were mostly conducted remotely due to safety considerations using the interview guide in Appendix A. Interviews were conducted with multiple employees within the organization to gain different perspectives into the reason for the failure to utilize an AAML system, to discover why the implementation was successful, and to discover the improvements of service.

Documentation. The researcher gathered documentation, either in paper form or electronically, from the participants. The documentation took the form of presentations and reports produced by participants through the installation and use of the AAML system. This documentation was analyzed and entered into the coding software.

Researcher Notes. The researcher took notes during the participant interviews to record the non-verbal aspects of the interview process. The notes were used to reflect on the participant's mood and attitude towards the AAML system and the AAML system's implementation. Notes were taken during the analysis of documentation provided to the researcher. These notes were either made directly on paper copies and digitized, with e-ink and included with the documents, or directly in the document using the hosting software's notes facilities.

Triangulating the Evidence. The participant interviews were the primary vehicle for gathering information for this research project. The interviews were validated through member-checking (Leedy et al., 2016). The interview transcripts were returned to the participant to review after transcription to allow for any clarifications and corrections. The review process added another level of validity to the study. While the participants worked for the same organization, they were located in different geographic locations and held different roles throughout the company, including leadership and front-line workers. Interviewing participants at different levels of the organization allowed for validating the views expressed by the employees among the different levels of management and non-management throughout the organization. Additionally, interviewing participants in different functional groups allowed the researcher to gain different perspectives on the AAML system throughout the organization.

Disconfirming Evidence. As the researcher is interviewing participants and collecting documentation, some material may not support the hypothesis. The researcher should explore the negative information to provide a more realistic view of the phenomenon under study. Reporting on negative evidence, as well as positive evidence, gives a more realistic assessment of the case study (Creswell & Poth, 2018).

Researcher Bias. The researcher's goal is to produce a report based on real-life experiences in a truthful manner without bias (Amerson, 2011; Yin, 2018). However, all researchers introduce a certain amount of bias into a qualitative study (Creswell & Poth, 2018).

The research can limit or eliminate bias by designing instruments with unbiased questions (Yin, 2018). Also, the researcher should process all the communication paths during an interview by observing the nonverbal communications as well as listing the words the participant is saying to determine the meaning of those words (Yin, 2018).

Summary of Reliability and Validity

As a researcher conducts research for a study, the research must be reliable and valid. Through the use of consistent application of the interview guide and data collection of other sources, the researcher can reduce any bias brought to the study (Yin, 2018). The researcher also brings reliability to the study by reaching data saturation through the interview process (Creswell & Poth, 2018). As the researcher strives for validity, triangulation and negative evidence validate the data collected throughout the research (Creswell & Creswell, 2018; Creswell & Poth, 2018; Krathwohl, 2009).

Transition and Summary of Section 2

This section established the researcher's role, participants, research design and method, population and sampling, data collection methods, data analysis, and reliability and validity of the study. During this qualitative case study, the researcher was the primary instrument of this study. The researcher conducted interviews, took field notes, and analyzed documents provided by participants. The participants of this study were current or former employees of an interstate natural gas pipeline. The participants were sampled through purposeful sampling and snowball sampling. The employees were interviewed by the researcher about their experience with an AAML and the effects the AAML had on transporting natural gas. The primary method of data collection is participant interviews. The population was considered saturated when no new information was realized. The data were coded and entered into the Atlas.ti software for analysis.

All the interviews were conducted using the interview guide, interviews were transcribed and reviewed by the participants, documents were analyzed, and the field notes were coded and entered into the software. The data were triangulated through interviews with multiple participants at different organizational levels and geographic locations, though documentation gathered and coding field notes. The next section will present the discussion of findings, applications to professional practice, recommendations for further study, and reflections on the research study.

Section 3: Application to Professional Practice and Implications for Change

Section 3 provides an overview of the study, a description of the participating organization, and describes the environment before and after the advanced analytic systems were installed. This section then describes the themes discovered throughout the study, the relationship of the themes to the research questions, and an analysis of the findings in relation to the problem and conceptual framework developed for this study. This section then presents the applications to professional practice, recommendations for action, and recommendations for future study. This section also presents reflections of the study by the researcher.

Overview of the Study

A case study is designed to describe the complexities of a single case (Stake, 1995). Implementing AAML systems can be a complex and expensive task for organizations (Giannino et al., 2018; Lechevalier et al., 2019). Therefore, this qualitative single case study added to the body of literature by examining the benefits of applying advanced analytics and machine learning to the interstate natural gas pipeline industry in the United States and why failures occur in fully exploiting advanced analytics and machine learning full potential. The researcher focuses on the following research questions to discover the effect an AAML system has on the organization and why companies fail to utilize AAML systems fully. First, Research Question 1 explores why companies fail to fully utilize the AAML system by discovering the factors for both success and failure. Second, the study explores the effect AAML systems have on the service provided by the organization. Finally, the improvements to the service provided by an interstate natural gas pipeline are explored. The data for this case study were primarily gathered by interviewing employees of the participating organization. Interviews allowed the researcher to have a deeper understanding of the participants' responses through follow-up questions where further clarifications were necessary beyond what a paper survey would allow (Yates & Leggett, 2016).

Presentation of the Findings

This section begins by discussing the demographics of the possible study population, the background of the selected participant organization, and the environment before and after the AAML system was installed. Saturation and triangulation will be discussed. Following saturation and triangulation, a discussion of the findings is presented.

Demographics

The researcher chose a single case study because of the limited number of possible participants for this study located in the United States (EIA, n.d.b). The participant organization was chosen because the researcher had direct connections within the organization, allowing ease of access to the organization's employees. This access allowed the researcher to have more intimate conversations with the participants throughout the time of the study than would have been possible with an organization that was not as well known to the researcher.

Organization Background. The organization has been transporting natural gas within the central United States for over 100 years. Through several thousand miles of pipeline buried underground, the organization supplies gas to local distribution companies, industrial facilities, power plants, and interconnects with other pipelines receiving gas from and delivering gas throughout the greater United States. While the company does not produce or take ownership of the natural gas transported through the system, a network of underground storage fields store gas produced in periods of lower demand to be extracted from the storage fields on behalf of the owning customer during peak demand periods, such as extremely cold weather or when the price of natural gas is higher during these same demand periods.

Environment Before AAML. Before executive leadership introduced advanced analytics and machine learning into the participating organization, the only method of analyzing data was querying data from the hosting applications' databases or extracting data from application-hosted reports to obtain information required to produce an analysis of business conditions. The participating organization was in the early stages of BI&A 1.0. The organization's hosting applications data are stored in RDBMS systems; however, the data were not transferred into OLAP systems for further analysis. The participating organization had both Industry 3.0 and Industry 4.0 technology deployed throughout the organization; however, outside of the hosting applications and Microsoft Excel, there was no analysis of the data generated by the sensors deployed throughout the organization beyond the rudimentary analysis that was allowed by the hosting applications and Microsoft Excel.

AAML Environment. After the retirement of several executive leadership members, the new executive leadership introduced AAML into the organization. Starting with the reorganization at the top levels of management, the organization placed a new focus on analytics to enhance the organization's reporting and decision-making environment. The AAML environment includes an analytics platform for business analytics, an analytics platform for real-time data to a resolution of one minute, and machine learning platforms for predictive decision making.

Participants Groups. The participants were divided into three different categories based on the organizational level within the organization and divided into four different groups based on the employee's job function. The employee categories included (a) Executive Leadership, (b) Non-Executive Leadership, and (c) Non-Leadership. The employee groups were (a) Operations, (b) Measurement and Storage, (c) Information Technology, and (d) Administration and

Customer Service. The categories are described in Table 1, and functional groups are described in Table 2. All participant names referenced in this study are pseudonyms.

Table 1

Employee Categories

Category	Description	Participants
Executive Leadership	Chief Executive Officer (CEO) and Vice Presidents that report directly to the CEO.	3
Non-Executive Leadership	Any management not reporting to the CEO or a Vice President reporting to the CEO.	10
Non-Leadership	All employees not in management.	18

Table 2

Functional Groups

Group	Description	Participants
Operations	Leadership and employees responsible for operating the pipeline who are not in Measurement and Storage.	8
Measurement and Storage	Leadership and employees involved with gas measurement and management of the storage fields. This group includes employees in field operations and customer services.	9
Information Technology	Leadership and employees involved with the management and operations of the information technology assets of the organization	4
Customer Services and Marketing	Leadership and employees responsible for customer service and marketing activities, excluding gas measurement.	4

Group	Description	Participants
Administrative	Leadership and employees responsible for administrative duties within the organization, including executive leadership	4

Saturation

Saturation emerged from the grounded theory methodology (Creswell & Creswell, 2018). Moreover, Creswell and Creswell (2018) recommend 20-30 interviews to reach saturation. However, saturation is reached when enough data has been collected or when the amount of time spent interviewing outweighs the value of the information being returned. If a researcher finds additional interviews are not adding to the codebook or returning new information, the researcher may conclude that data saturation has occurred (Creswell & Creswell, 2018; Guest et al., 2006). For this research project, the researcher was targeting interstate natural gas pipelines in the United States. This target group has limited possible participants. The researcher chose one interstate natural gas pipeline and interviewed 31 current or former employees of the target organization. This research project reached saturation at 31 interviews. The researcher is confident that saturation was achieved given the limited number of possible participants.

Triangulation

Triangulation is the process of corroborating evidence from different sources, perspectives, and methods to validate a theme by locating a code in different data sources to seek convergence of a theme or pattern (Creswell & Creswell, 2018; Creswell & Poth, 2018). The researcher gathered themes generated from the data from across the different employee categories and functional groups. Triangulation was achieved by interviewing participants at all levels of the organization and from different functional areas. When a theme was discovered, the researcher searched for data to triangulate the theme among the participant groups and employment levels. For this research project, the researcher discovered nine themes from the data gathered from the employee categories and functional groups.

Discussion of the Findings

As the participants were interviewed, the themes described in this section began to develop. Each theme is related to a research question. This section brought together the themes discovered, research questions, existing literature, and the conceptual framework. Analysis of the organization was compared and contrasted with the UTAUT and Transformational Leadership. Through this analysis, the organization's environment before AAML was explored, and the results of installing AAML were shared.

Research Question 1: Failure to Fully Utilize AAML. Why do companies fail to fully utilize advanced analytics and machine learning for process monitoring? This research question and the sub-questions explore the causes relating to the failure to take full advantage of the AAML system, including the factors that cause the failures, along with the factors that contribute to the success of the AAML system. The AAML installation can still be a success, depending on what analytical platforms were in place before the AAML system was installed, even if the organization does not fully utilize the full capabilities of the AAML system.

When installing a new information technology system, resources available to the users are an essential part of UTAUT. Both performance expectancy and effort expectancy require resources to be available to the users to increase technology acceptance (Venkatesh et al., 2003). If the resources are not available in the time required, the AAML system's use can be delayed, causing the system to be underutilized. As discovered, sufficient resources were not available to the AAML system users. As Research Question 1a demonstrates, the lack of resources and training are the driving factors for failing to utilize the AAML system fully.

Research Question 1a: Factors Contributing to Failure. What factors contribute to the failure to utilize advanced analytics and machine learning? Implementing an AAML system can be complex and challenging for organizations (Giannino et al., 2018). Therefore, organizations underutilize the volume of data made available by the various sensors placed along the production path to monitor and detect failures of the system (Peres et al., 2018). While analyzing the participants' responses, the researcher considered the factors brought to light by the participants.

Theme 1: Lack of Resources. When asked if the participants had the resources needed to use the AAML system, the majority of the participants (n=23) indicated they had the resources needed. However, the remaining participants (n=8) indicated there was a lack of certain resources. Human resources (n=5) and computing resources (n=5) were the resources mentioned as lacking in the AAML environment.

Contributing factors to the lack of resources were partially budgetary and partially understanding of the new technology required for an AAML system to be fully utilized. Early in the deployment of the AAML system, the Information Technology staff did not have a complete understanding of the required resources; therefore, the allocation of resources was delayed. While the allocation of resources was resolved, the delay caused issues in installing and deploying the AAML systems. From a budgetary standpoint, there is a backlog of demand for developing additional AAML features, which are delayed because there are not enough technical resources to fulfill the requested items.

Human resources to implement the needed features were mentioned by five of the participants. The Information Technology staff's backlog of work is greater than the current staff can process. Linus was concerned with Information Technology human resources available to perform needed tasks to expand the system. He suggested the resources allocated to the AAML were not sufficient to handle the pent-up demand for expansion of the system. As Linus stated,

> One thing I've mentioned to the IT group, mentioned to my supervisor, and I'll mention to anybody, there's the analytics. IT spread so thin. Our guys are taking off and running with what we do. Other departments are taking off and running where they want to run. Even right now, again, IT is spread so thin. So, those two departments are running their way. IT is, maybe, working with another department. Well, we start asking questions that interferes with what they're doing with the other groups. Just start spreading people thin, again. We're in the infancy, and this thing is still growing. I've mentioned to many, can we get somebody, position them in the IT group, position them in that business analytics group, but almost make them the liaison for the separate departments. You would have to add to the staff that they currently have now to make that happen. You could have one person for a couple departments or one person for a department, depending on how big it is from an IT standpoint. It's kind of that liaison and helps helps them drive through that minefield. Or, if it's going stay how it is now, then, to go that further, a bit of background programming and calculations and things we need to happen within the advanced analytics, that starts taken away from the time my people need to spend on what their traditional job is. So, somewhere in there, at some point, nothing's free. Having the people power needs to occur somewhere in there and that's not being supported yet.

The demand may be unknown to the Information Technology department because people are not requesting help due to the perception that the IT staff has more work than they can currently execute with the limited number of IT personnel employed. As Tom stated,

> You know, so many new things, at this moment, so many people are tied up with all these different projects, that many people don't really have the time to show me. There isn't a training class to show me how to create something. And I understand that, and I'm the type that I don't want to burden someone. If what I'm needing right now isn't that important, if it's just something I want to learn for professional development type of thing, I'm not going to burden somebody with that if it's not that important.

In the same vein, Ryder states,

> We can always use more resources, I'll say that. Every manager wants more resources, but, I think, it takes the right type of individuals, and what you're trying to accomplish, for this program to be successful, you've got to have pretty much a good unity with business users; knowing the data, knowing what they need, what type of data they need to see, where it's located, where the primary source of the data is. You also have to have the IT side of building and programming, and changing views, and showing options, being open to change.

The insights provided by Tom, Linus, and Ryder align with Venkatesh et al. (2003) regarding performance expectancy and effort expectancy level's importance for the acceptance of a new information technology system.

Computing resources were another resource type mentioned by five participants as a constraint on the AAML environment. While some of those have been resolved, participants felt that the early struggles of sizing the infrastructure hampered the growth of the system. Because the technology was new to the organization, it was not fully understood how to allocate the AAML system resources. AAML systems produce and store more data and require more compute resources than traditional systems. The Information Technology staff was not familiar with the sizing requirements at the beginning of the project. As the system grows, the ongoing need to allocate additional resources to the AAML system may be challenging from a budgetary standpoint. As Leesa stated,

> We also struggled with infrastructure, understanding our needs, from a resource aspect, in other words, server sizing. The need for some of these systems to have more dedicated resources, as opposed to sharing resources with other systems. The need for us to be able to be more involved with things on the server, as opposed to waiting for someone in the infrastructure to do something for us.

Leesa went on to say other parts of the infrastructure were also a challenge to implement.

> The learning curve for infrastructure, in other words, infrastructure, network security, understanding that this is a whole different animal, it doesn't look like a lot of the other things they've had to support, to a degree. And, were a lot more specialized, and our requirements are different than a lot of the other systems they've been used to.

Kirby also concurred that infrastructure is a challenge to getting the system configured in the early stages.

> For the most part, we've lacked, what I would say, horsepower with the system. So, resources to be able to make the system respond in a timely manner, sometimes, and, some of the knowledge base within the company to be able to implement this with the system with external applications to be able to make us more fruitful with this system has been lacking, as well.

Furthermore, Beth pointed out that the base data are still in a more complex form which hampers non-technical users from easily adopting the more advanced features of an AAML system. As Beth stated, “I think if we get our base data to a point that end users don't need to understand database schema in order to use the tools, then I think that would be easier for them to learn.”

These projects are complex and are different from the traditional SCADA systems that many companies use to control and analyze systems (Lechevalier et al., 2019; Shinozuka et al., 2015). Leesa, Beth, and Kirby highlighted this organization was no different. Learning the new AAML system's requirements was a challenge early in the deployment of the AAML system. Not having the required resources in place slowed the AAML system's deployment and made using the software's analytic capabilities difficult for the participants to utilize fully.

Theme 2: Lack of Training. During the course of the interviews, training was mentioned by six participants as a potential barrier to fully utilizing the AAML system. The participants related there was a lack of training or wanted training on the products. Participants felt they lacked proper training (n=7), and there is a demand for training (n=4) throughout the organization.

Many participants felt they could better utilize the AAML resources available if they had training. They also felt training would allow them to be more self-sufficient and reduce their dependence on other employees to fulfill their needs. One participant said they never received any training on using the AAML system. Another participant was self-taught in using the business analytics system, and another participant learns by searching the system for the information needed. These participants desired training and believed it would enhance their use of the AAML systems. For example, Brenda stated.

> I think that comes from my experience prior to coming into it. And, I wish that there were more classes to teach us different tools and tips. I've used YouTube and, of course, people I work with that have taught us. But, I wish there were more classes to teach us about [the systems]. I feel like I could use and sift through the data better if I knew how to navigate.

Mary agreed, “probably just some training. Glenn always offers to help too. He's like, ‘Oh, you go here and go here and go here.’ I'm like, ‘Okay, well, I really just need to understand the whole thing.’ Further, Shannon believes formal training would be helpful to further the use of the system. He states,

> I think that there are people there [that could train us], I probably would have to reach out to them. I don't think there's any training set up or given; but, I could probably ask someone to help me. I think that every time I reach out to anyone in the company to help with something, everyone is super helpful. So, I know that the people would be willing, but there isn't anything, like, where they have all those classes [for the] ERP. There's nothing that they're setting up for [the AAML]. I've never seen anything for that, but I know people would be willing.

Tiffany also is not comfortable without official training in the more advanced analytic capabilities of the system. She commented,

> There's not just been official training. It's just, kind of, here's a link. Go out to [the AAML] and see what you can see, or this is where you need to go to find this. But, as far as pulling in any of the information, looking at the tables and stuff like that, I would not say that I am comfortable with any of that stuff.

Reducing the effort expectancy is important in deploying a new information system (Venkatesh et al., 2003). Training can reduce the effort to use a new information system. The insights shared by Brenda, Mary, Shannon, and Tiffany show the lack of training increased the effort expectancy required for participants to use the system to the fullest extent possible. While participants generally found the reports and screens easy to use once created, creating needed reports without technical assistance was a barrier to the participants.

The participants who do not possess strong technical abilities to extract information from the systems using the previous methods felt that they could use the new system to develop the analytics necessary without involving power users or information technology employees if they had training in the AAML systems capabilities that are necessary to provide the analytics, on an ad hoc basis, required in the performance of their daily duties. While the AAML system is relatively easy to operate from a consumption standpoint, some participants related they felt they could use the system more effectively if they had additional training in the software's operation beyond the usage of the dashboards created by content creators.

Research Question 1b: Success Factors. The factors for the success of the AAML system include (a) ease of use, (b) ease of learning, (c) the ability to use the system most of the time without help, and (d) the participant’s self-direction in creating the vision for the AAML system. Each of these factors contributes to the successful implementation of the AAML system throughout the organization. All four factors contribute to the success of implementing both the business analytics software and the real-time analytics software. When asked if the advanced analytics and machine learning system was a good idea, every participant answered either yes or with a strong yes indication.

Theme 3: Ease of use. The ease of use, or degree of difficulty using the system, is an integral part of the acceptance of a new information technology system (Venkatesh et al., 2003). When deploying a new system, Information Technology managers and developers should place a high focus on the ease of use of a software application (Khechine et al., 2016). Each participant was asked to rate how easy it was to use the AAML system. The participants were also encouraged to discuss why the system was easy to use or was not easy to use. When considering how easy the system was to use, 28 participants indicated the system was easy to use, and three participants indicated the system was not easy to use. The use of the system varied depending upon the participant's job function. With the business analytic system, some users were report creators, some users are report consumers, and some users were a combination of report consumers and report creators. The majority of the report consumers found the system extremely easy to use. Employees who were report creators reported varying degrees of ease of use. The report creators' ease of use depends on whether they also used reports or were just creators of reports. For example, Scott shared, "yes, it is [easy to use]. It does require some bit of training just to get used to it. But yes, it's way, way easier, more intuitive, more user-friendly [than the previous system]. Easier to share information, for sure." Tiffany concurred,

> Yes. I'm figuring it out. It is a lot easier than what we have used in the past. It's just being able to know where to go for what you're looking for, where it's housed under what section, and then maneuvering through it. But it is very helpful.

However, Mary did not find the system easy to use. She stated,

> Well, it's almost like, you can hit a facility, drill down to the unit, then drill down to the PM, then drill down to the task list, and then drill down to the completed, and each one is on a separate page. I can't just print the test list.

The majority (*n*=28) of participants found the system easy to use. The more experienced participants found the system easier to use than those who were newer to the system. Scott, Tiffany, and Mary's statements align with the UTAUT performance expectancy (Venkatesh et al., 2003). The participants who created reports and dashboards relayed the AAML system was easier to use than the previous system but still possessed complexities that hampered the system's ease of use. Overall, however, the majority of the participants felt the system was fairly easy to use, especially if they had had proper training or had the necessary resources available when they needed help in developing reports or dashboards within the new system. Many users felt the system was more intuitive than the previous system, and it was easier to find the information required in their day-to-day tasks.

Theme 4: Ease of Learning. The participants were asked how easy it was to learn how to use the advanced analytics system. Twenty-four participants indicated the system was easy to learn, while six participants indicated the system was not easy to learn. While the ease of learning varied between report users and report creators, the consensus was the system was easy to learn.

The users who use the AAML system's output, typically through dashboards, indicated the system was extremely easy to learn and intuitive. Participants who created dashboards or authored reports using the AAML system reported that the new system was easier to use than the previous methods available to them using Microsoft Excel or Microsoft Access to obtain the data required to accomplish their daily job tasks. Tom stated, “Anybody that is remotely familiar with computers can easily [use the reports]. You click on whatever you want, and you can apply filters. It's pretty simple.” Tucker agreed with Tom’s evaluation of the usability of the dashboarding capabilities of the system. “[The system] was [easy to learn]. Thankfully, I'm kind of part of that era who grew up with technology. So, things like this typically, just as a user, come pretty natural," stated Tucker.

However, participants that created reports indicated there was a learning curve. However, after using the product for a period of time, it became evident how to use the system to produce the output they required for their daily job duties. Leesa pointed out,

> For myself, what I work with, there's a learning curve to it, but overall, it wasn't that bad. For people that I've trained to work with what I work with, which again, is the real-time data analytics type, most of them seem to pick it up fairly quickly, as well.

Ben agreed with Leesa's view,

> I would say the first week or two was a pretty big learning curve, but I think that's just about with anything. After that, it became pretty intuitive. If I thought about something I wanted to do, I had a pretty good idea of how to do it.

One of the critical success factors for the AAML system is the ease of learning how to use the system. Tom and Tucker found learning the system was easy, while Leesa and Ben described the system as more complex to learn from a building standpoint. The findings agreed with Pynoo and van Braak (2014), perceived ease of use is a strong indicator of acceptance.

Venkatesh et al. (2003) proposed that the effort expectancy is higher for women and older individuals when learning a new system. The participants in this research study were in various age groups of adults. Of the participants that stated the system was not easy to learn, the gender was tipped toward males (female = 1, male = 4), and the ages were across the spectrum.

Theme 5: Use Without Help. Participants were asked if they could use the advanced analytics and machine learning system without help most of the time. Twenty-nine participants responded they could use the system without help most of the time, while two of the users expressed they still needed help to use the system. For example, Mark stated,

> Depending on what it is, if it's [the business analytic system] or something like that, yes; but, if it's, [the real-time analytics system] that I'm not in too often, I might need some help, but I can always reach out to somebody. There's always somebody to reach out to if you need help.

Similarly, Jordon commented,

> Now, yes. Most of the time I can get in there and I can pull up the information I need, whether it's [the business analytic system] looking up system balance or line segment balance. I feel really comfortable getting in there, manipulating that information. Then [in the real-time analytic system], I'm still looking for and trying to learn where [the information is] or what I can pull into view. But I'm becoming more comfortable with using it without any help, also.

Allison was more direct. “Yes. I rarely have any issues” she stated.

As Mark, Jordon, and Allison indicated, they were reasonably self-sufficient using the system and only required help on certain portions of the system. This finding agrees with Compeau and Higgins (1995) that the users with a high self-efficacy used the system more successfully. Most of the users (*n*=29) of the AAML system could use the system most of the time without help. As the users gained experience using the system, they required less help. Along with the experience, the ease of use (Theme 3) and ease of learning (Theme 4) played a role in the ability to use the system without help most of the time. All participants indicated they had someone to call if help was required.

The majority of the users expressed they could view reports and filter as necessary without help the majority of the time, while some of the report creators responded that they still needed help, on occasion, to create and design reports. The two users who indicated help were still required when using the system did not specify what type of help was needed. All 31 of the participants indicated there was someone to call if they needed help with the AAML software. Help may be in the form of the organization’s helpdesk or the vendor of the software.

Theme 6: Employee Self Direction. When the executive leadership initiated the AAML projects in the participating organization, they shared a general vision of the future. The AAML system's vision was to enable the business to make quick, efficient, data-driven decisions based on the best available data. Further, the goals included greater information flow within the organization, decisions based on real-time information instead of batch reports looking solely at past data, and introduce real-time predictive decisions, especially for the operating groups. In this light, each group that uses AAML defined a vision based on the guidance given by the executive leadership.

The executive leadership wanted the operating groups to define their specific vision and purpose for the AAML system based on the overall vision and goals for introducing AAML into the organization. As Jeff indicated,

> I didn't want the team to get strapped into one way of doing things. I let the team determine how the dashboards would interact and what the prioritization of what we want to tackle. I set it up and drove it in the beginning, but let the team run with it. Because for me, that's a part of change management. You show them a clear the path, show them the light, if you will, with the tools. What I did was build a team and let them kind of define that [the vision].

Don also stated,

> We challenged the organization. There were some folks that said we were nuts; ‘why are we focusing on this; it's not really going to change things for us.’ We had to overcome some internal barriers. Is this ‘should we just stick with the old systems that we have,’ or do we need to be thinking in a more forward-looking basis.

Linus concurred, “They weren't heavy-handed in saying ‘you got to do this.’ It was pretty much ‘you guys, go do what you need for your department.’ They didn't drive what it had to be.”

As Don, Jeff, and Linus pointed out transformational leaders positively influence the organization through proper funding, allowing adequate time, and personnel to support advanced innovation through the organization. Transformation leaders also encourage creativity throughout the organization, allowing for increased innovation in the products produced by the organization's employees (Gumusluoğlu & Ilsev, 2009). When defining the advanced analytics projects, executive leadership did not provide a specific vision for the system. Executive leadership provided an overview of the desired product. They allowed the employee groups to determine the vision and path to deliver the best product to fit the decision-making executive leadership desired. Many of the visions for the uses of advanced analytics throughout the various departments were conceived and implemented by non-leadership employees and presented to leadership as tools to use for decision-making throughout the organization.

Early on, most leadership and non-leadership did not have enough experience in advanced analytics and machine learning to develop a comprehensive vision for the future with advanced analytics. The leadership allowed the initial groups the time required to experiment and prototype the advanced analytics system to prove the benefit throughout the organization. As the early adopters developed dashboards and reports using advanced analytics and machine learning, the uses became apparent throughout the organization. Demand for the new system began to increase throughout the organization as early adopters began to show their work products through advanced analytics and machine learning.

After operations deployed the real-time analytics system, the measurement and storage group saw the benefits of real-time access to information that they had never been able to gather and analyze prior to the installation of the advanced analytics system at the organization. The storage and measurement group began to develop its vision for the use of advanced analytics. After the development of the vision, measurement and storage began the process of deploying advanced analytics to analyze the equipment for which they were accountable in order to improve the service provided by those groups.

Research Question 2: AAML Effects on Production Process. How do advanced analytics and machine learning affect the production process? By using advanced analytics, the production process can be better monitored for failures, possibly preventing failures from occurring during critical times (Giannino et al., 2018). The AAML system allows the organization to monitor the production process in real-time to a greater detail than allowed by SCADA systems. The monitoring of systems in real-time provides the organization with the ability to detect failures earlier than traditionally allowed through SCADA systems and alarms on PLCs, which typically are not predictive (Shinozuka et al., 2015). By predicting failures in advance, the organization can schedule the repairs to the potentially failing equipment before a severe failure occurs and at a time convenient to the organization and the organization's customers so that service disruptions are minimized or in periods of reduced demand so that existing equipment can be taken out of service when the equipment is not required to transport natural gas.

Theme 7: Prevention of Mechanical Failures. Taking down the equipment in a planned nature before a failure occurs, the repair may cost less than running the equipment to failure or waiting until the PLC shuts down the equipment on alarm which can occur at any time. Suppose the shutdown or failure occurs during extreme cold or extremely hot weather. In that case, service disruption may occur, and customers may not have the volume of natural gas required to serve the end-use customers during extreme weather conditions. Removing equipment from service before failures occur can also have an environmental effect. If a piece of equipment fails in a catastrophic manner, such as a severe mechanical failure in a compressor or a pipeline rupture, natural gas will be released into the atmosphere. Preventing natural gas release into the environment helps reduce the environmental waste potentially created by equipment failure throughout the pipeline system (O'Shea et al., 2019).

Before the installation of the AAML system, failures among the various systems may not have been detected. Failures may include mechanical failures or catastrophic failures of pipeline infrastructure. Any of these failures have the potential to impact customer gas supply and cause downstream devices to malfunction, including catastrophic failures.

Natural gas is an odorless, colorless, and tasteless product transported through underground pipelines by compressing the gas. At the receipt and delivery points, gas is measured with various devices to determine the volumes of gas passing through the pipeline for billing and accounting purposes.

Before the AAML system was installed, a mechanical failure was not always detected in a timely manner to minimize the impact on customers and minimize the repair's financial impact. Because of the short time the AAML system has been installed within the organization and the initial focus on the mechanical area, mechanical failures are the most common failures to be detected. These types of failures can be the most expensive to remedy. Over the AAML system's life, it has been estimated that direct savings in repair costs are approximately two million dollars. These costs do not include potential costs from lost gas, customer disruption, and other downstream costs associated with a mechanical failure. Table 3 shows various mechanical failures, and the consequences had the failures not been detected. Each of the failures discussed by Tucker, Don, Leesa, Steve, and Linus would have created waste in the system. The statements by these participants agree with Giannino et al. (2018), O'Shea et al. (2019), and Shinozuka et al. (2015). Both O'Shea et al. (2019) and Shinozuka et al. (2015) stated that malfunctions along a pipeline system can cause damage to more than just the equipment itself. The damage can result in loss of product, resulting in environmental damages, also. Furthermore, the participants agree with Giannino et al. (2018). Advanced analytic systems help diagnose problems before they occur, allowing operators to shut down and repair the equipment in a controlled and planned manner.

Table 3

Selected Examples of Mechanical Failure

Participant	Component	Statement
Tucker	Compressor	We have a brand new compressor at a location, and there have been some issues with a differential pressure coming into that station. That's something we typically wouldn't have seen before or if we did see it, it was just kind of value that would alarm once it got outside of the certain threshold. With AAML they're able to set up parameters monitor, not just in gas control. They can see these things coming and we can make proactive measures to take care of the situation before we have a failure. And that actually happened right before the polar vortex. We were able to take the compressor down, fix it, and get it back online a few days before we had this event take place. It would have been horrible if this had happened during that event. Had it failed, most likely, our backup horsepower would not have started due to do the cold temperatures. And, we would have

Participant	Component	Statement
		been losing a lot of flow and pressure going into our market areas, which we could not afford to do during these over two BCF days. And then also it's not just going to impact our market, it would impact our production as well. That compressor is critical for us to get the gas from point A to point B.
Don	Compressor	During the last polar vortex, for instance, the very fact that our network continued to operate for the worst, or the highest, 14 consecutive days, we've ever had on the system. I think the analytics that goes into the reliability and integrity of the system is the only reason we were able to do that. Two and a half times what we've ever experienced, and we really did not have a compressor station failure. We had a few issues, here and there; but, a failure during that time would have been not catastrophic; but would have put a lot of people in the cold. So it was phenomenal.
Leesa	Compressor	For example, in the compression equipment, there's been many, many times where this new system has been able to alert. There was crankcase pressure that was high. There was leakage from around some of the packing of the cylinders. They had one failure of a lifter, to where parts of it actually would have fallen down inside of the cylinder. If that went in there and stayed like that, that could have really mess something up quite a bit.
Steve	Storage	During the last polar vortex, we were able to identify wells that had frozen off, or the meters at least have frozen. And, instead of the techs or operators having to go around each individual wells and check them, we can pinpoint and say, “Hey, well number ABC, check these wells” and able to identify those fairly quickly. Without the AAML system either (a) they would have gone to every single well, and seen what was going on or (b) it just wouldn't have been checked.
Linus	Storage	[For example,] when the dehydration facilities that are starting to have issues that we can address before the system failed. We've seen, in the latest polar vortex, we were able to go out beyond even just the field capabilities. Now, in the system, we actually have our well head measurement where we have that out in the fields; and that data, again, is instantaneous. Able to get it in the same system by clicking just down to the next screen, then being able to trend it. We could look for wells that were frozen off during the polar vortex. We could have a technician go straight to the well and work on that well, and then get that gas deliverability back to the system. Prior to that, if we thought

Participant	Component	Statement
		there might have been wells froze off, the technician is just out driving around the field trying to find that well on his own in whatever weather may be out there. Now we can look at it, and 99 times out of 100 tell which well to go to from here. That gets gas back on, that was feeding customers, keeping people warm during the polar vortex.
James	Measurement	We had a major weather event, the polar vortex, and being able to watch things in as close to real-time as possible, you can see pressures beginning to deviate, see equipment beginning to malfunction or freeze, and those type of tip-offs can see them before they even get to a critical point. So, if you're monitoring it and we have prevented big issues with it. We were watching a town border setting. The pressure started getting low, and it got low enough that it alarmed. The technician was dispatched, and he got out there, and it continued to drop in pressure. His regulator, he suspected it was freezing off, but it wasn't. We determined that the upstream pressure had decreased because the mainline pressure had decreased. So, in the end of the things, it was not feeding the meter enough pressure, so then, the meter was over speeding. That meter would have broken. It would have come apart internally had we would not noticed that and got him out there quick enough and made some adjustments to pressure to slow that meter back down. More than anything, it would have created a lower pressure situation to the customer, which there is always risk there to lose those customers when that happens. But, it would have definitely cost us several thousands of dollars to replace the meter and the man-hours to go with it.
Luke	Measurement	A failure would be a relief valve going off. We were able to see pressures climbing, alarms came in. Technicians were able to respond before an overpressure protection device went off. [If the relief valve activated], we would lose gas, customers could be without gas service. Line pressure could have gone above its MAOP.

Research Question 3: Service Improvements. How do advanced analytics and machine learning improve service in the natural gas industry? There are several possible factors that advanced analytics and machine learning could have on the service proved by interstate natural gas pipelines. Of the 24 questions and follow-up questions about the advanced analytics and machine learning systems in use at the participating organization, five of the questions provided insight into the advances in service realized by the participating organization.

Theme 8: Prevention of Procedural or Manual Errors. Before the AAML system was installed, many reports were generated throughout the organization using manual methods, typically copy and paste from host applications or reports from host applications and placed into Excel spreadsheets for further analysis. This process was a tedious and time-consuming process for the employees of the organization. These reports could take hours or days to produce and did not give the management the information they needed in a real-time fashion. The reports were generated by following a set of steps on a checklist to ensure none of the steps were missed during the production of these reports. While generating these reports in Microsoft Excel, it was possible to create errors within the spreadsheet structure, thereby producing reports containing incorrect information.

Before installing advanced analytics and machine learning systems, the participants had to repeat many manual processes throughout each business cycle to report the pipeline system's conditions. Because of the time required, many of these reports were only provided infrequently, for example, weekly or monthly. When producing these reports, errors could occur when transferring information from a host system report to an Excel spreadsheet for further analysis and management reporting. After installing the advanced analytics system, many of these steps were automated, and the reports were then run on a more frequent basis, such as hourly or at the end of each business cycle. As Tiffany revealed,

> Originally it was a lot of manual [entry, creating] room for error by people. I think for the majority of what I've been dealing with it is a lot easier to review, a lot easier for the information to pull in. And I think it's just made life, in general, easier for us. [The

> previous method] would fall into days of trying to redo, rework an Excel spreadsheet. Sometimes we had to start from scratch to make sure we had all the information correct. Whereas, now it's all housed in the AAML system and easy to look to, and we don't have to start from scratch or revisit it all.

Andy added, “Just taking human error out of it. Whenever it's automated, obviously, there’s times that there are mistakes when it's pulling the wrong data or something doesn't update right on the back end that it's pulling data from.” Billy also highlighted the ability to monitor for procedural failures in the system.

> [The report is a] preventative type report to help prevent us from getting ourselves into trouble because you can visually see where you're going. Look at the summer of 2020. Last year, prices went kind of crazy after the pandemic started. And, crude oil went to minus $40. So, what happened was the gas prices really fell into the dirt. So, we started parking gas like crazy because there were huge spreads, and all of a sudden, we looked up, and we had [redacted] BCF parked, and it was only July. We had a problem. We had to unwind a lot of those parking deals, and we basically had to reimburse the customers for the amount of money. That's why we built those reports because we didn't want to have to go through that again. It was not very pleasant.

As Tiffany, Andy, and Billy show, the Performance Expectancy is increased by using the AAML. The increase in expectancy agrees with Venkatesh et al. (2003), in which the system enhances the user’s job performance. The advanced analytic system has prevented procedural and manual errors from occurring. Now that the AAML system produces these reports automatically and with repeatability, the chance of errors occurring has been dramatically reduced. The dashboards created by the business analytics system allow for the processes to be repeated multiple times a day through automation, eliminating the errors caused by the manual processing of information. These automated reports will also show errors in the host system's data quicker than the old manual reports because of the frequency of their production and review by the analytical staff and management. In addition to copy and paste from host systems or reports, power users could use Microsoft Access to develop queries to provide data that was not allowed through host reports and combined data from database tables that the hosting system reports did not have built into the system. However, the power users may not have the knowledge of the back-end systems required to correctly combine the data to provide the necessary information for business decisions.

Theme 9: Increased Productivity. Before installing the advanced analytics and machine learning system, the participating organization produced many analyses of business conditions manually by copying and pasting data from application-produced reports into Microsoft Excel for further analysis. The process of copying and pasting information from application-generated reports into Excel was time-consuming and tedious for the employees who produced these reports. The manual intervention required to produce an analysis of data was limited because of the manual process involved in creating any analysis. The time it took to produce these reports for management did not allow for the real-time data analysis required to make decisions in a timely fashion.

Productivity increases were reported by many of the participants during the interviews. The productivity gains were realized by eliminating copying and pasting information from various host application reporting systems or manual input of information into Excel for data that are now collected automatically through the real-time analytics system. Participants reported saving many hours of work by implementing the advanced analytics system. The advanced analytics system also reduced the tediousness of employees' jobs by eliminating the repetitive manual work required to produce daily, weekly, and monthly reports that are now automated by the advanced analytics system. The analytics system also allows users to analyze the data in greater depth due to the advanced analytics system's graphing and filtering capabilities.

Of the 31 participants interviewed, 30 participants stated that the advanced analytics and machine learning environment increased their productivity, with only one participant saying it did not increase their productivity. The participant who did not see an increase in productivity was a new user of the system. This participant stated they had not received adequate training at this point to see a productivity improvement in their job. The participants who answered yes to productivity gains through the use of AAML answered with varying degrees of strongness of yes. Seven participants gave a resounding strong yes, while the remaining 22 participants gave a positive yes. Table 4 shows selected examples of statements of productivity gains. As Venkatesh et al. (2003) demonstrated and the participant statements concur, when Performance Expectancy is considered, the AAML system enables users to gain productivity in their daily activities.

Table 4

Selected Examples of Significant Statements for Productivity Gains

Participant	Answer	Statement
Mark	Yes	Spending a bunch of time just running the same report every single week and doing the same calculations on the same thing. Once you get it set up, it's easy. It just refreshes and goes unless something breaks.
Tom	Yes	I was productive in creating the report. However, I am more productive in other areas; because I'm not having to spend time creating the reports that are now automatic.

Participant	Answer	Statement
Don	Strong Yes	Absolutely. It took a little bit of time to build the processes that we have, but we have seen amazing success from this. The reliability of the system has gone up dramatically because of what we've been able to discern using this. What we've been able to discern has gone up dramatically for us. Our ability to understand what's coming from a market perspective is significantly better than what we had ever experienced as a company. And so, we’re able to extract greater value because we have a better idea of what's going to happen as a result of weather, or how much wind there is. We didn't even factor wind into our initial discussions, and now we're able to measure that impact on our business, as well.
Ryder	Strong Yes	In my department, some of these reports it cuts because maybe a two, three-hour job that they were doing, and now, they've got time to focus on something else—may be helping a customer or some self-development on learning the analytics system. They got time back that they can apply and add value to the company that they were tied up, producing data.
Mary	No	Well, I don't know how to use the system, so no. I wish I could. I wish I could use it. I mean, there's so many times I go in there and, I know this data is in here somewhere, and I try to figure it out, and I just get lost and confused.

The AAML system also allows the participants to perform their job quicker. All participants except for one indicated the AAML system allowed them to perform their job quicker than before. Of the 30 that responded yes, seven of them indicated a strong yes. Strong yes indications included absolutely, definitely, and without a doubt. Participants indicated they were able to cut hours of work from their daily tasks, and it also increased communication among the employees. One participant even equated the new AAML system to life before a cell phone. When looking back, the participant does not know how the job was performed without the AAML system.

Similarly, the participant does not know how life would be without a cell phone. Table 5 shows significant statements relating to quicker job performance. As Venkatesh et al. (2003) also demonstrated, and the participant statements concur, when Performance Expectancy is considered, the AAML system enables users to perform their job quicker overall than the previous system allowed.

Table 5

Selected Examples of Significant Statements for Quicker Job Performance

Participant	Statement
Tom	Absolutely. It has saved me at least eight to ten hours a month, just in report creation.
Tucker	Yes, I am able to devote a lot more time to things that need it, instead of spending half a day trying to get data.
Billy	Oh, yeah, because I can get more information out to people a lot easier.
Jordon	I think so, yes. I think because we're able to use all three systems and get a better picture of what is going on with the meters and on the system. And we're able to help our techs get the information they need and ourselves be able to understand better. And, I think, the communication is 10 times, 100 times better with being able to have these new systems or these new tools available to us in order to perform our jobs.
Linus	Absolutely. I've made this comment to a couple other people before; I look at the system right now, kind of like, cell phones. If you took somebody's cell phone away, it would be like how [did I] ever live without having cell phone before now? And with what we're able to do in [the real-time analytic system] right now. It's almost like we've forgotten how we did it before, because this does it so well.

Analysis of Conceptual Framework

The conceptual framework analyzes the effect of implementing an AAML package on an organization by comparing the organization's previous state with the state of the organization post-implementation. This study examined the concepts of customer satisfaction and the potential for increased profits. The result was customer satisfaction, and profits were achieved by installing an advanced analytics and machine learning system into this organization.

Customer Satisfaction. Increasing customer satisfaction with the use of advanced analytics and machine learning was achieved through a multi-faceted approach when deploying the advanced analytics and machine learning system. While most of the benefits are transparent to the customer, the customers realize increased satisfaction through less unplanned downtime through predictive analytics.

The advanced analytics system allows the organization to take advantage of the vast number of sensors distributed throughout the transportation system and monitor the sensors in a real-time fashion that was not allowed before the advanced analytics system was installed. Before the advanced analytics system was installed, many of these centers were either not monitored, only used to produce alarms on PLCs, or monitored at a much lower frequency than is possible with the current advanced analytics system. Before the installation of the advanced analytics system, the portion of the data from these sensors that were stored was rarely used unless a problem occurred. Even after the problem had occurred, the data were difficult to obtain and analyze quickly.

Using predictive analytics, the organization can plan downtime of equipment when that piece of equipment is not required for natural gas movement, for example, during periods of temperate weather in the spring and fall of the year. Also, predictive analytics allows the company to perform maintenance before a catastrophic failure occurs, thereby reducing unexpected downtime enabling both the organization and the customer to plan for the outage at a time convenient for both parties.

Profit. Advanced analytics and machine learning allow the organization to realize profits both monetary and non-monetary from the system's usage. From a monetary standpoint, the advanced analytics system was estimated to save at least $2,000,000 in repair expenses that would have been incurred had the AAML system not been in place. While these savings may not have gone directly to the bottom line, the cost avoidance enabled this money to be used elsewhere throughout the organization.

From a non-monetary aspect, the advanced analytics system allowed the organization to avoid costs relating to equipment repair and environmental costs by reducing natural gas vented to the atmosphere through incidents that were avoided with the use of predictive analytics. For example, in at least one incident, the analytics system prevented a pressure relief device from activating, which would have vented natural gas to the atmosphere until the pressure on the system segment was under the pressure relief device's setpoint. Not only was an environmental cost avoided, but there were also costs avoided in other potential equipment failures downstream from the relief valve and the direct cost of natural gas vented to the atmosphere.

Improved Service. By reducing unplanned downtime of equipment, service has improved throughout the operation of the pipeline system. For example, during the recent polar vortex, the pipeline system did not experience any major equipment disruptions, which could potentially have caused widespread customer disruption through the loss of service of a major piece of equipment. Depending on the piece of equipment that may have failed, the outage caused by the failure could have been spread across multiple states and multiple large metropolitan areas. While the outages may not have been costly in terms of life, the repair cost could have been extreme and time-consuming, in addition to customers losing natural gas service during a dangerously cold time of the year.

Also, with the reduction of manual processes throughout the organization, employees are freed up to perform other tasks that require greater cognitive abilities beyond manual data movement processes. The employees can now provide a deeper analysis of the data through their job functions and provide management and customers with answers quicker than before the advanced analytic system was installed. The AAML system allows the organization to potentially avoid hiring additional employees to provide the same level of service that they can provide today by using advanced analytics to assist in compiling and analyzing the organization's data.

The advanced analytics system has also removed much of the tedious work employees had to perform day-to-day and allowed them to use their analytical skills to provide a deeper analysis of the data and provide real-time decision-making to the organization's leadership. By reducing the tedious work, employees see a greater value in the work they produce on a day-to-day basis in support of the organization's mission.

Summary of the Findings

This study sought to find the reasons why organizations failed to fully utilize AAML systems, what factors lead to these failures, what factors lead to the success of implanting an AAML, what effects an AAML system has on the productions process, and how advanced analytics improve the service provided by interstate natural gas pipelines in the United States. This study found that organizations fail to fully utilize AAML systems in their processes due to a lack of training and lack of resources. The resources include both human and computing resources.

The study found the organization did not allocate enough computing resources to the AAML system, especially during the program's early stages. The lack of resources slowed the progress of implementing the system to the fullest extent. Additionally, human resources are lacking in certain areas of the organization. There is a demand for further analytics throughout the organization, especially in the data-intensive areas where analysis is critical for business functions. Furthermore, the organization lacks formal training for the advanced use of the analytical systems, preventing the organization from fully utilizing the AAML system. Participants indicated they could better utilize the AAML system with additional training. On the flip side, several factors have influenced the AAML system's success, even though it is not fully utilized. Most participants found the system easy to use, not difficult to learn, could use the system most of the time without help, and enabled participants to perform their job quicker than before the AAML system was available.

The AAML had a positive effect on the production process by reducing the maintenance expense required by identifying problems with mechanical systems before a serious failure occurred. By catching the failures sooner, the organization could repair the machinery at a convenient time and at a lower cost. On the gas flow side of the production process, inaccuracies of measuring equipment were found sooner. Finding these issues sooner allowed for more accurate measurement of natural gas and potentially reduced loss of gas from the system.

Improvements to the service provided by the organization were realized by a reduction in tedious, repetitive manual processes, which allowed for employees to focus on the analytical duties of their jobs. The participants indicated the automatic processing reduced errors incurred while performing manual data movement from reports to Excel spreadsheets. Furthermore, having information available to answer inquiries from management and customers allowed the participants to provide better service to their customers.

The leadership positively influenced the move to AAML systems from manual, reactive systems through transformational leadership. Because of the lack of experience had by many of the organization's employees, leadership drove the early stages of the implementation of the AAML system. As employees gained knowledge of and experience with the AAML system and realized their informational needs, leadership provided an environment that fostered creativity and provided the funding necessary to implement the AAML system.

Application to Professional Practice

As the findings above indicate, this study shows there are gaps in business practices when implementing advanced analytics and machine learning systems in natural gas pipelines in the United States. This section will explore the recommendations for general business practice and potential application strategies to resolve the gaps in business practices.

This study explored the use of advanced analytics and machine learning by an interstate natural gas pipeline located in the United States. Before installing the advanced analytics and machine learning system, the organization produced reports manually by copying and pasting information from host system reports, direct queries from the back-end database, or manual entry of information from paper forms. The organization also had automation that could provide information at a much greater resolution than was being collected by the electronic systems installed in the organization. The data collected through automation were rarely used for predictive analytics and typically only used to study, after the fact, what caused a piece of equipment to fail.

After installing the advanced analytics and machine learning system, the organization saw an increase in real-time data usage to make timely decisions. However, the advanced analytics and machine learning system was not fully utilized throughout the organization. The organization failed to fully utilize the advanced analytics and machine learning system for a variety of reasons. In the beginning, the organization did not have the resources needed to fully implement the advanced analytics system throughout the entire organization. Moreover, while it is still the case that there is a lack of resources to meet the demand for advanced analytics that has been generated through the use of the AAML system, many resource constraints have been addressed. Further, training on the advanced analytics and machine learning system has been limited and has not been conducted throughout the entire organization. Most of the training in the advanced analytics software packages has been informal in nature, with no formal class being offered to the organization's employees.

Improvements realized throughout the organization have included (a) increased productivity of employees who utilized advanced analytics, (b) decreased maintenance costs realized by performing the maintenance in a proactive manner instead of a reactive manner, and (c) increased customer satisfaction using predictive analytics to plan maintenance of the equipment during times that are convenient for both the customer and the organization. The advanced analytic system has prevented failures from occurring during high-demand usage periods, increasing customer satisfaction for the organization.

Improving General Business Practice

This study demonstrates how AAML systems can improve a business's performance through various aspects of process improvement. From a business analytics standpoint, an AAML system can help the business analyze data in a real-time fashion, allowing quicker decisions. From a real-time analytics standpoint, an AAML system can help reduce maintenance costs and increase uptime for mechanical equipment through predictive maintenance.

Business Analytics. This study demonstrates that the business analytics portion of AAML can improve business practice by reducing repeated manual processing of data. By eliminating manual processing of data, the business reduces the chances of errors when producing reports and analyses caused by missed steps when processing the data. Many times, the data required to produce an analysis of business conditions may come from many different sources and host systems. It can be challenging to gather the data from each host system to produce the necessary report.

Business analytics allows for the automation of processes that enable consistent processing of data each time the analytic process occurs. The analytics system can pull the data from multiple systems and combine the data into one view of business operations. Business analytics also allows the presentation of data through the use of graphics, filters, and slicers to quickly view the data in different manners to improve the analysis. Furthermore, by automating the data ingestion process to create analytic reports, the business can process the data more frequently. For example, a monthly report may take several hours or days to produce through the traditional manual processes but may only take minutes after the automation is set up. Because the report is now automated, it can be produced on-demand or scheduled at any time throughout the month. The automation allows for information to be more visible to the business, allowing decisions to be made based on the data in a more real-time fashion. Moreover, the employees who produced the data have more time to analyze the data through the reports produced by the automation gathered from distributed sources and combined through the business analytics process.

Real-time Analytics. Real-time analytics allows an organization to monitor the production process and evaluate events as they occur. The real-time analytics system monitors the data as it is ingested into the system, trend data, and alert operators should the process fall out of normal operating ranges. Furthermore, the data can be analyzed to find equipment that will soon fail, allowing the organization to schedule maintenance before a major failure occurs, reducing maintenance costs and production downtime.

This study shows the real-time analytics system also allows analysts to monitor and analyze the data, directing the maintenance personnel to a failing piece of equipment. Knowing which piece of equipment may fail optimizes the maintenance personnel's time by eliminating the need to physically search for the failing piece of equipment. The advanced analytics system also prevents unnecessary maintenance from occurring by eliminating the guesswork typically needed without advanced analytics to find the failing piece of equipment.

Employee Considerations. This study also demonstrated that many of the analyses performed by organizations before advanced analytics were manual in nature. Advanced analytics allow the employees to reduce the manual and tedious work of producing analytics, thereby increasing their productivity by reducing the amount of time it takes to produce a report. The reduction in reporting time allowed the employees to deliver more and deeper analysis versus focusing on data movement from one report to another to produce an analysis. The reduction in manual processes allowed the employees to provide information to management and customers more timely than before the analytic system was available.

Potential Application Strategies

Based on the study’s findings, the following recommendations for action have been developed to improve the utilization of AAML systems for organizations. The recommendations proposed by the researcher should help organizations improve the utilization of the AAML systems throughout their organizations, thereby improving business outcomes. These recommendations should not be construed to be the only areas that need improvement. However, the researcher found these areas to have the greatest impact on impending an AAML system.

Employee Training. Training of employees should begin before choosing an AAML system. Employees involved in selecting an AAML should be trained in the basics of an AAML system and that system's purpose. After due diligence in selecting an AAML system has been completed, the initial group of employees who will be configuring and using the AAML system should be trained in the areas that they will be using the AAML system. Employees who will be using the AAML system should be trained in the use of the dashboards and information produced by the AAML system. Ongoing training should occur for the additional employees who will be using and configuring the AAML system. This training will ensure that employees will have the knowledge and skills to utilize the AAML system while performing their job duties fully.

Simplify Data Access. When building reports, it is important to have data stored in the most straightforward manner possible. The data schema's simplicity will help report creators who are not technical find the data they need in the format they need. To accomplish this simplification, the organization's data should be stored in a data warehouse. The data warehouse will allow the technical users to simplify the layout of the data for non-technical users. The various application data can be stored in one central location and be cleaned during the ETL process while loading the data into the data warehouse. The data warehouse technology could be as simple as another relational database or purpose-built data warehousing software.

Right Size Resources. Rightsizing the AAML system's resources will help increase the system's utilization by ensuring the system is sized appropriately and the staffing is in place to support the AAML system. Participants in this study stated that resources were one of the components that prohibited the AAML system's full utilization. The resources necessary for fully utilizing an AAML system include human resources and computing resources.

Human Resources. Having the appropriate human resources in place to implement an AAML system is vital to the full utilization of an AAML system. Not only are human resources necessary to implement the AAML system, but human resources are also necessary for the ongoing maintenance of the AAML system. During the early implementation phases, resources are needed to install software, configure the hardware, and train users. After the AAML system is implemented, appropriate staffing with business knowledge and technical knowledge is necessary to ensure the proper usage of the AAML system. As different groups are brought into the AAML ecosystem, additional training will be necessary. Furthermore, as additional groups utilized the AAML system, additional technical resources to assist in building and configuring dashboards and analytics are necessary to utilize the AAML system fully.

Computing Resources. Ensuring the proper computing resources are available to the AAML system is essential in the full utilization of an AAML system. Ensuring the proper computing resources are available will allow the complex analytics to process in the time necessary to make business decisions, allow for expansion of the analytics system, and ensure the AAML system's availability. When sizing the AAML system, the Information Technology department should ensure adequate resources to house the data and process the analytics. The configuration of the computing resources should be such that expanding the resources as necessary will be as simple as possible. The expansion could be achieved through using virtual computing and storage or hosting the AAML systems resources at a cloud provider.

Summary

This study demonstrates how business analytics and real-time analytics can improve an organization’s performance through quicker decision making, reduced maintenance costs, and increase productivity by eliminating manual and repetitive processes. These improvements are achieved by forward-thinking leadership by transforming an organization from a manual information processing organization into an analytic organization through advanced analytics and machine learning. With proper human and computing resources, employee training, and vision from leadership, employees can deploy advanced analytics and simplified data access for the consumers of the data.

Recommendations for Further Study

This qualitative single case study explored the reasons interstate natural gas pipelines in the United States fail to fully utilize their AAML systems. There was sufficient literature describing UTAUT and Transformational Leadership in the general business context. However, in the context of advanced analytics and machine learning, little information was found for the researcher to draw upon. Further research is recommended in these areas. Additionally, the researcher found little information targeted explicitly at interstate natural gas pipelines in the United States relating to UTAUT and Transformational Leadership. Further research relating to UTAUT and Transformation Leadership would be beneficial.

The research showed the failure to fully utilize an AAML system was partly due to insufficient training. The lack of training slowed the full use since employees did not have the knowledge to use the more advanced features of the system. There was an unspoken demand for training of which the organization was unaware. Further research to discover why employees are reluctant to request training is needed.

The study also discovered that sufficient resources were not allocated to the AAML system, especially in the early stages of the AAML system's deployment. Both human and computing resources were lacking during these stages. While some of the resourcing issues were due to former management’s reluctance to embrace the newer technology that was not the only cause. Research is recommended in the proper allocation of human resources and the allocation of computing resources relating to an AAML system deployment, expansion, and ongoing operations. Research into resource allocation to provide recommendations for the proper allocation would guide organizations to the methods and resource count necessary for the quicker realization of full utilization.

Reflections

The doctoral journey has provided the researcher with growth in both professional and personal aspects. This section will discuss the professional and personal growth gained during the journey. Also, the business functions of an AAML system will be discussed in relation to the Christian worldview.

Personal and Professional Growth

Professional Growth. The doctoral journey and this research study allowed the researcher to enhance professional skills. The doctoral program exposed the researcher to broader areas than typical in an information technology position. Many of these areas of study were last visited in the degree programs leading up to the beginning of the doctoral program. The skills gained from the course of study will allow the researcher to continue to expand in the area of professional growth.

Personal Growth. The doctoral program forced the researcher off “the couch of life” by giving a renewed focus to every day. While the doctoral program's pace may have been intense for the last three years, it was the boost the researcher needed to get out of the daily rut of work to home and back again. While the last year of the journey has only been a 90 degree turn from work to homework brought about by the COVID-19 pandemic, at least the researcher finally moved into an office with a window.

Doctoral Journey. Upon considering the journey to conducting this study, the university's course work prepared the researcher to conduct this study. The researcher found the agile dissertation process to be in alignment with the style of work the researcher typically works within. While the researcher found the literature review challenging, that section provided a solid foundation for the study findings. Even when challenging issues arose, the researcher kept in mind the end goals and the true Source of Power, which guided the completion of this research study.

Now that this journey is finally complete, this begs the question, “what’s next?” God has a plan, or he would not have led the researcher to this university and this program when it was most needed. The researcher hopes the plan is more than mowing 10 acres instead of six, however.

Biblical Perspective

When examining an AAML system from a Christian worldview, an AAML system brings together better use of resources, application of new knowledge, and profit through the smarter use of resources through knowledge. These factors allow the AAML system to contribute to the world in general and to the Christian worldview specifically by making wise use of the resources God has provided to humankind.

Resources. When God created the Earth and placed man on this planet, He expected us to be good stewards of His planet (New International Version Bible, 1978/2011, Genesis 1). Any production process should be monitored to ensure that the process works as efficiently as possible. Natural gas, being a natural resource, is no different. Natural gas transporters should take the necessary precautions to ensure the gas is transported as efficiently as possible with minimal loss.

As Genesis 2:15 states, “the Lord God took the man and put him in the Garden of Eden to work it and take care of it.” (*New International Version Bible*, 1978/2011, Genesis 2:15).

Traditional tools utilized for monitoring the production process, such as those used in controlling the transportation of natural gas, are not typically advanced enough to monitor the system at the granularity level that a modern sensory network can provide (Giannino et al., 2018; Krumeich et al., 2016; Shinozuka et al., 2015). With the charge to take care of the garden and, by extension, the Earth, humans should make use of every tool to maximize the resources given to us. Advanced analytics and machine learning help maximize the resources given to us by God by allowing the organization to monitor the process to reduce waste generated by failures and inefficient processes. For example, running a machine to failure wastes material and money, which could be put toward another purpose. By repairing the machinery before a complete failure occurs, less material and money are used to put the equipment back into service. Additionally, using the AAML system to pinpoint failures allows the maintenance personnel to go directly to the problem. The use of the AAML system saves time and other resources that would be spent trying to find a piece of equipment with an issue that was not otherwise known when using previous methods to monitor the processes involved with natural gas transportation.

Knowledge. Proverbs 3:13-15 states, “Blessed are those who find wisdom, those who gain understanding, for she is more profitable than silver and yields better returns than gold. She is more precious than rubies; nothing you desire can compare with her.” (New International Version Bible, 1978/2011, Proverbs 3:13-15). The knowledge brought forth by the AAML system helps the organization operate in a more efficient and effective manner. The organization profits from this knowledge and can make better use of the resources God provides for us.

The AAML system gives the organization the knowledge to reduce the waste in the process of moving natural gas. This waste reduction allows the organization to reduce the methane emissions that would otherwise go into the atmosphere instead, moving the methane down the pipeline to consumers of the product. The reduction of waste not only removes the methane emissions from the atmosphere but also allows a reduced cost of service to the customer.

Profit. During the course of conducting business, God gives man profits from the operations. Proverbs 14:3 states, “All hard work brings a profit, but mere talk leads only to poverty” (New International Version Bible, 1978/2011, Proverbs 14:3). However, hard work alone does not bring profit. Work combined with knowledge allows the organization to work smarter. The AAML system allows the organization to reduce unnecessary trips to physical facilities to search for problems by pinpointing the problem and focus resources where they are needed the most. The focus brought about by the knowledge the AAML system brings about, more intelligent use of resources is possible. The AAML system also reduces the cost of individual repairs, thereby giving a better allocation of resources.

Summary

The doctoral journey has allowed this researcher to grow in personal and professional ways not imagined before. The professional knowledge gained through this program is invaluable. From a personal perspective, the program has re-engaged the researcher to “get off the couch of life.”

The AAML system integrates the Christian worldview with the business worldview through wise resource use, knowledge, and profit. God expects us to use the resources provided to use wisely. The wise utilization of resources is obtained through advanced knowledge of the environment humans operate within. Wise use of the resources God gives us through knowledge leads to profits. The profits can be returned to the business or to the investors to be wisely used to continue the best use of resources granted to humans.

Natural Gas, COVID-19, and This Study

The foundation for this study began during the early days of the COVID-19 outbreak of 2020 and was conducted throughout the pandemic. The virus was first detected in China and spread throughout the world, being declared a pandemic by the World Health Organization in March of 2020 (Rabadan, 2020). Once the disease began to spread throughout the world, the landscape of how business was conducted changed almost overnight (Hussain et al., 2021). As governments began to institute public health rules, businesses began to realize the consequences (Hussain et al., 2021). As social distancing rules were put into place, employees who were able began to work from home on a regular basis, even when they were not able to before the pandemic (Jenkins & Smith, 2021). The participating organization was not different.

All employees who were able either worked from home or began their workday at home, instead of at an office. Many of the employees who operated the pipeline worked alone even before the pandemic. For these employees, the pre-pandemic workday began at one of the operating facilities, with assignments reviewed before the beginning of the workday. After the social restrictions were put into place, the work assignments were given via the work management system, and the employees proceeded to the job directly from their homes. Office employees began working from home regularly. The participating organization had been moving towards a more remote workforce before the pandemic began. However, the full remote workforce plan was forced into action with little notice and on a grander scale than imagined. Fortunately, because the participating organization was on the road to more remote work, the disruption was not as severe as it could have been if the plan were not already in motion.

Because of the social distancing requirements put into place, this study was conducted entirely remotely using Microsoft Teams and the conferencing abilities within the product. Pre-pandemic, many of the interviews would have been conducted in-person, while the participants located at a distance would have been conducted remotely.

Summary and Study Conclusions

This qualitative single case study involved 31 participants from one interstate natural gas pipeline in the United States. By interviewing 31 participants in different departments and at different levels within the organization, triangulation was achieved through the participants' different viewpoints. The findings discovered in this study can be applied to any interstate natural gas pipeline and businesses in general when implementing an AAML system.

This research study recommends that organizations provide training at all stages of AAML implementation and operation to fully utilize the system. Additionally, simplifying data access through the use of a data warehouse will help non-technical users further utilize the AAML system in their day-to-day activities. Finally, an organization should continually monitor the resources allocated to the AAML system to verify the right people are in place, and the correct amount of computing resources are available for the operation of the AAML system. These three recommendations can be brought forth through a transformational leadership team that encourages creativity, spurring innovation while providing the proper funding, time, and personnel to support the AAML system.

References

Abawajy, J. (2015). Comprehensive analysis of big data variety landscape. *International Journal of Parallel, Emergent and Distributed Systems*, *30*(1), 5–14. https://doi.org/10.1080/17445760.2014.925548

Akpinar, M., & Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies*, *9*(9). https://doi.org/10.3390/en9090727

Al-Jepoori, M., & Al-Khanjari, Z. A. (2018). Framework for handling data veracity in big data. *International Journal of Computer Science and Software Engineering*, *7*(6), 138–141. http://ijcsse.org/published/volume7/issue6/p2-V7I16.pdf

Al-Salim, A. M., El-Gorashi, T. E. H., Lawey, A. Q., & Elmirghani, J. M. H. (2018). Greening big data networks: Velocity impact. *IET Optoelectronics*, *12*(3), 126–135. https://doi.org/10.1049/iet-opt.2016.0165

Alvesson, M., & Kärreman, D. (2015). Intellectual failure and ideological success in organization studies: The case of transformational leadership. *Journal of Management Inquiry*, *25*(2), 139–152. https://doi.org/10.1177/1056492615589974

Alzyadat, W. J., AlHroob, A., Almukahel, I. H., & Atan, R. (2019). Fuzzy map approach for accruing velocity of big data. *Compusoft*, *8*(4), 3112–3116. https://www.proquest.com/openview/9acc91cce16fd069d580f8664cd55db7/1?pq-origsite=gscholar&cbl=2032622

Amerson, R. (2011). Making a case for the case study method. *Journal of Nursing Education*, *50*(8), 427–428. https://doi.org/10.3928/01484834-20110719-01

Andersen, L. B., Bjørnholt, B., Bro, L. L., & Holm-Petersen, C. (2017). Achieving high quality through transformational leadership: A qualitative multilevel analysis of transformational leadership and perceived professional quality. *Public Personnel Management*, *47*(1), 51–72. https://doi.org/10.1177/0091026017747270

Arnaboldi, M. (2018). The missing variable in big data for social sciences: The decision-maker. *Sustainability*, *10*(10), 3415. https://doi.org/10.3390/su10103415

Bacchelli, A., Mocci, A., Cleve, A., & Lanza, M. (2017). Mining structured data in natural language artifacts with island parsing. *Science of Computer Programming*, *150*, 31–55. https://doi.org/10.1016/j.scico.2017.06.009

Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., & Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, *80*, 937–949. https://doi.org/10.1016/j.eneco.2019.03.006

Bhattacharjya, J., Ellison Adrian, B., Pang, V., & Gezdur, A. (2018). Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter. *The International Journal of Logistics Management*, *29*(2), 723–738. https://doi.org/10.1108/IJLM-06-2017-0157

Boeren, E. (2017). The methodological underdog: A review of quantitative research in the key adult education journals. *Adult Education Quarterly*, *68*(1), 63–79. https://doi.org/10.1177/0741713617739347

Bottomley, P., Mostafa, A. M. S., Gould-Williams, J. S., & León-Cázares, F. (2016). The impact of transformational leadership on organizational citizenship behaviours: The contingent role of public service motivation. *British Journal of Management*, *27*(2), 390–405. https://doi.org/10.1111/1467-8551.12108

Bottomley, S. (2014). Patenting in England, Scotland and Ireland during the Industrial Revolution, 1700–1852. *Explorations in Economic History*, *54*, 48–63. https://doi.org/10.1016/j.eeh.2014.08.002

Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in Industry*, *101*, 1–12. https://doi.org/10.1016/j.compind.2018.04.015

Brinch, M. (2018). Understanding the value of big data in supply chain management and its business processes: Towards a conceptual framework. *International Journal of Operations & Production Management*, *38*(7), 1589–1614. https://doi.org/10.1108/IJOPM-05-2017-0268

Bryant, A., & Charmaz, K. (2019). *The sage handbook of current developments in grounded theory*. https://doi.org/10.4135/9781526485656

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of significance: Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077

Castellani, F., Astolfi, D., Sdringola, P., Proietti, S., & Terzi, L. (2017). Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment. *Applied Energy*, *185*, 1076–1086. https://doi.org/10.1016/j.apenergy.2015.12.049

Chamberlin, D. D. (2012). Early history of SQL. *IEEE Annals of the History of Computing*, *34*(4), 78–82. https://doi.org/10.1109/MAHC.2012.61

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology [Article]. *Communications of the ACM*, *54*(8), 88–98. https://doi.org/10.1145/1978542.1978562

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188. https://doi.org/10.2307/41703503

Chen, Y.-T., & Sun, E. W. (2018). Automated business analytics for artificial intelligence in Big Data@X 4.0 era. In M. Dehmer & F. Emmert-Streib (Eds.), *Frontiers in Data Science* (pp. 223–251). CRC Press.

Chertkov, M., Backhaus, S., & Lebedev, V. (2015). Cascading of fluctuations in interdependent energy infrastructures: Gas-grid coupling. *Applied Energy*, *160*, 541–551. https://doi.org/10.1016/j.apenergy.2015.09.085

Clandinin, D. J., & Caine, V. (2008). The SAGE Encyclopedia of Qualitative Research Methods. In (pp. 542–544). Sage. https://doi.org/10.4135/9781412963909

Codd, E. F. (1970). A relational model of data for large shared data banks. *Commun. ACM*, *13*(6), 377–387. https://doi.org/10.1145/362384.362685

Čolaković, A., & Hadžialić, M. (2018). Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, *144*, 17–39. https://doi.org/10.1016/j.comnet.2018.07.017

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189–211. https://doi.org/10.2307/249688

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches*. Sage.

Dai, D., Xu, T., Wei, X., Ding, G., Xu, Y., Zhang, J., & Zhang, H. (2020). Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Computational Materials Science*, *175*, 109618. https://doi.org/10.1016/j.commatsci.2020.109618

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Doctoral dissertation, Massachusetts Institute of Technology]. Cambridge, MA. http://hdl.handle.net/1721.1/15192

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/MNSC.35.8.982

De, G., & Gao, W. (2018). Forecasting China’s natural gas consumption based on adaboost-particle swarm optimization-extreme learning machine integrated learning method. *Energies*, *11*(11). https://doi.org/10.3390/en11112938

Deslonde, V., & Becerra, M. (2018). The technology acceptance model (TAM): Exploring school counselors' acceptance and use of naviance. *The Professional Counselor*, *8*(4), 369–382. https://doi.org/10.15241/vd.8.4.369

Desmet, K., Greif, A., & Parente, S. L. (2020). Spatial competition, innovation and institutions: the Industrial Revolution and the Great Divergence. *Journal of Economic Growth*, *25*(1), 1–35. https://doi.org/10.1007/s10887-019-09173-3

Djiroun, R., Boukhalfa, K., & Alimazighi, Z. (2019). Designing data cubes in olap systems: A decision makers' requirements-based approach. *Cluster Computing*, *22*(3), 783–803. https://doi.org/10.1007/s10586-018-2883-7

Dong, Y., & Qin, S. J. (2018). Dynamic latent variable analytics for process operations and control. *Computers & Chemical Engineering*, *114*, 69–80. https://doi.org/10.1016/j.compchemeng.2017.10.029

Drosio, S., & Stanek, S. (2016). The Big Data concept as a contributor of added value to crisis decision support systems. *Journal of Decision Systems*, *25*(sup1), 228–239. https://doi.org/10.1080/12460125.2016.1187404

Duan, N., Bhaumik, D. K., Palinkas, L. A., & Hoagwood, K. (2015). Optimal design and purposeful sampling: Complementary methodologies for implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 524–532. https://doi.org/10.1007/s10488-014-0596-7

Dul, J., & Hak, T. (2007). *Case study methodology in business research*. Taylor & Francis Group.

EIA. (n.d.a). *About U.S. natural gas pipelines - transporting natural gas*. https://www.eia.gov/naturalgas/archive/analysis_publications/ngpipeline/index.html

EIA. (n.d.b). *Interstate natural gas pipeline segment*. https://www.eia.gov/naturalgas/archive/analysis_publications/ngpipeline/interstate.html

EIA. (n.d.c). *Thirty largest U.S. interstate natural gas pipeline systems, 2008*. https://www.eia.gov/naturalgas/archive/analysis_publications/ngpipeline/MajorInterstatesTable.html

Elahe Kordi, G. (2018). Explaining brand switching behavior using pull–push–mooring theory and the theory of reasoned action. *Journal of Brand Management*, *25*(4), 293–304. https://doi.org/10.1057/s41262-017-0080-2

Elforjani, M., & Shanbr, S. (2018). Prognosis of bearing acoustic emission signals using supervised machine learning. *IEEE Transactions on Industrial Electronics*, *65*(7), 5864–5871. https://doi.org/10.1109/TIE.2017.2767551

Fagan, M. H., Neill, S., & Wooldridge, B. R. (2008). Exploring the intention to use computers: An empirical investigation of the role of intrinsic motivation, extrinsic motivation, and perceived ease of use. *The Journal of Computer Information Systems*, *48*(3), 31–37. https://doi.org/10.1080/08874417.2008.11646019

Farhan Zeb, K., & Imamuddin, K. (2018). Do ‘employee training programs’ affect employee performance? *Journal of Business Strategies*, *12*(1), 71.

Farzaneh-Gord, M., & Rahbari, H. R. (2016). Unsteady natural gas flow within pipeline network, an analytical approach. *Journal of Natural Gas Science and Engineering*, *28*, 397–409. https://doi.org/10.1016/j.jngse.2015.12.017

Fissore, D., Pisano, R., & Barresi, A. A. (2018). Process analytical technology for monitoring pharmaceuticals freeze-drying – A comprehensive review. *Drying Technology*, *36*(15), 1839–1865. https://doi.org/10.1080/07373937.2018.1440590

Flouris, I., Giatrakos, N., Deligiannakis, A., Garofalakis, M., Kamp, M., & Mock, M. (2017). Issues in complex event processing: Status and prospects in the Big Data era. *Journal of Systems and Software*, *127*, 217–236. https://doi.org/10.1016/j.jss.2016.06.011

Fu, X., & Juan, Z. (2017). Understanding public transit use behavior: Integration of the theory of planned behavior and the customer satisfaction theory. *Transportation*, *44*(5), 1021–1042. https://doi.org/10.1007/s11116-016-9692-8

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*, *20*(9), 1408–1416. https://doi.org/10.46743/2160-3715/2015.2281

Gallinucci, E., Golfarelli, M., Rizzi, S., Abelló, A., & Romero, O. (2018). Interactive multidimensional modeling of linked data for exploratory OLAP. *Information Systems*, *77*, 86–104. https://doi.org/10.1016/j.is.2018.06.004

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, *27*(1), 101–113. https://doi.org/10.1016/j.jsis.2017.10.001

Giannino, F., Esposito, S., Diano, M., Cuomo, S., & Toraldo, G. (2018). A predictive decision support system (DSS) for a microalgae production plant based on Internet of Things paradigm. *Concurrency and Computation: Practice and Experience*, *30*(15), e4476. https://doi.org/10.1002/cpe.4476

Gil, D., Johnsson, M., Mora, H., & Szymański, J. (2019). Review of the complexity of managing big data of the Internet of Things. *Complexity*, *2019*, 4592902. https://doi.org/10.1155/2019/4592902

Goh, T. T., & Sun, P.-C. (2015). Teaching social media analytics: An assessment based on natural disaster postings. *Journal of Information Systems Education*, *26*(1), 27–36. https://aisel.aisnet.org/jise/vol26/iss1/3/

Goodarzi, F., Araban, M., Eslami, A. A., & Zamani-Alavijeh, F. (2019). Development and psychometric evaluation of the diabetic men's dietary behaviors inventory based on the theory of reasoned action. *Archives of Public Health*, *77*(1), 3. https://doi.org/10.1186/s13690-018-0328-7

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Gumusluoğlu, L., & Ilsev, A. (2009). Transformational leadership and organizational innovation: The roles of internal and external support for innovation. *Journal of Product Innovation Management*, *26*(3), 264–277. https://doi.org/10.1111/j.1540-5885.2009.00657.x

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence [Article]. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Hagger, M. S., & Chatzisarantis, N. L. D. (2009). Assumptions in research in sport and exercise psychology. *Psychology of Sport and Exercise*, *10*(5), 511–519. https://doi.org/10.1016/j.psychsport.2009.01.004

Hancock, D. R., & Algozzine, B. (2017). *Doing case study research: A practical guide for beginning researchers* (3rd ed.). Teachers College Press.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105–120. https://doi.org/10.1016/j.techfore.2015.12.014

Herbrich, R., & Graepel, T. (2015). *Machine Learning An Algorithmic Perspective*. CRC Press.

Higashino, W. A., Capretz, M. A. M., & Bittencourt, L. F. (2016). CEPSim: Modelling and simulation of complex event processing systems in cloud environments. *Future Generation Computer Systems*, *65*, 122–139. https://doi.org/10.1016/j.future.2015.10.023

Hildenbrand, K., Sacramento, C. A., & Binnewies, C. (2018). Transformational leadership and burnout: The role of thriving and followers' openness to experience. *Journal of Occupational Health Psychology*, *23*(1), 31–43. https://doi.org/10.1037/ocp0000051

Hoffmeister, K., Gibbons, A. M., Johnson, S. K., Cigularov, K. P., Chen, P. Y., & Rosecrance, J. C. (2014). The differential effects of transformational leadership facets on employee safety. *Safety Science*, *62*, 68–78. https://doi.org/10.1016/j.ssci.2013.07.004

Hofmann, E. (2017). Big data and supply chain decisions: The impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *55*(17), 5108-5126. https://doi.org/10.1080/00207543.2015.1061222

Hopkins, J., & Hawking, P. (2018). Big data analytics and IoT in logistics: A case study. *The International Journal of Logistics Management*, *29*(2), 575-591. https://doi.org/10.1108/IJLM-05-2017-0109

Huang, J. (2019). *Achieving performance predictability in transactional databases* (Publication No. 27815085) [Doctoral dissertation, University of Michigan]. ProQuest Dissertations & Theses Global. Ann Arbor.

Hussain, S., Xuetong, W., Hussain, T., Khoja, A. H., & Zia, M. Z. (2021). Assessing the impact of COVID-19 and safety parameters on energy project performance with an analytical hierarchy process. *Utilities Policy*, *70*, 101210. https://doi.org/10.1016/j.jup.2021.101210

Kalbandi, I., & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and Hadoop technology. *Procedia Computer Science*, *48*, 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Izadyar, N., Ong, H. C., Shamshirband, S., Ghadamian, H., & Tong, C. W. (2015). Intelligent forecasting of residential heating demand for the District Heating System based on the monthly overall natural gas consumption. *Energy and Buildings*, *104*, 208-214. https://doi.org/10.1016/j.enbuild.2015.07.006

Jamil, N. B. C. E., Ishak, I. B., Sidi, F., Affendey, L. S., & Mamat, A. (2015). A systematic review on the profiling of digital news portal for big data veracity. *Procedia Computer Science*, *72*, 390-397. https://doi.org/10.1016/j.procs.2015.12.154

Jenkins, F., & Smith, J. (2021). Work-from-home during COVID-19: Accounting for the care economy to build back better. *The Economic and Labour Relations Review*, *32*(1), 22-38. https://doi.org/10.1177/1035304620983608

Jha, S., & Topol, E. J. (2016). Adapting to artificial intelligence: Radiologists and Pathologists as information specialists. *JAMA*, *316*(22), 2353-2354. https://doi.org/10.1001/jama.2016.17438

Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process [Article]. *Journal of Product Innovation Management*, *34*(5), 640–658. https://doi.org/10.1111/jpim.12397

Karagiannis, A., Vassiliadis, P., & Simitsis, A. (2013). Scheduling strategies for efficient ETL execution. *Information Systems*, *38*(6), 927–945. https://doi.org/10.1016/j.is.2012.12.001

Kassner, L., Gröger, C., Mitschang, B., & Westkämper, E. (2015). Product life cycle analytics – next generation data analytics on structured and unstructured data. *Procedia CIRP*, *33,* 35–40. https://doi.org/10.1016/j.procir.2015.06.008

Kelle, E. V., Visser, J., Plaat, A., & Wijst, P. V. d. (2015, May 18). An empirical study into social success factors for agile software development. 2015 IEEE/ACM 8th international workshop on cooperative and human aspects of software engineering. https://doi.org/10.1109/CHASE.2015.24

Keskes, I., Sallan Jose, M., Simo, P., & Fernandez, V. (2018). Transformational leadership and organizational commitment: Mediating role of leader-member exchange. *Journal of Management Development*, *37*(3), 271–284. https://doi.org/10.1108/JMD-04-2017-0132

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, W. K. M., Alam, M., Shiraz, M., & Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *TheScientificWorldJournal*, *2014*, 712826–712826. https://doi.org/10.1155/2014/712826

Khechine, H., Lakhal, S., & Ndjambou, P. (2016). A meta-analysis of the UTAUT model: Eleven years later. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, *33*(2), 138–152. https://doi.org/10.1002/cjas.1381

Kim, B.-J., & Jung, S.-Y. (2019). The mediating role of job strain in the transformational leadership⁻safety behavior link: The buffering effect of self-efficacy on safety. *International Journal of Environmental Research and Public Health*, *16*(8), 1425. https://doi.org/10.3390/ijerph16081425

Kiran, D. R. (2019). Chapter 31 - Manufacturing resource planning (MRP II). In D. R. Kiran (Ed.), *Production Planning and Control* (pp. 441–455). Butterworth-Heinemann. https://doi.org/https://doi.org/10.1016/B978-0-12-818364-9.00031-7

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4), 303–312. https://doi.org/10.1177/0340035216672238

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, *116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037

Kornecki, M., & Strube, J. (2018). Process Analytical Technology for Advanced Process Control in Biologics Manufacturing with the Aid of Macroscopic Kinetic Modeling. *Bioengineering*, *5*(1). https://doi.org/10.3390/bioengineering5010025

Krathwohl, D. R. (2009). *Methods of educational and social science research: The logic of methods*. Waveland Press.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657–2664. https://doi.org/10.1016/j.jacc.2017.03.571

Krumeich, J., Werth, D., & Loos, P. (2016). Prescriptive control of business processes: New potentials through predictive analytics of big data in the process manufacturing industry. *Business & Information Systems Engineering*, *58*(4), 261–280. https://doi.org/10.1007/s12599-015-0412-2

Kurt, R. (2019). Industry 4.0 in terms of industrial relations and its impacts on labour life. *Procedia Computer Science*, *158*, 590-601. https://doi.org/10.1016/j.procs.2019.09.093

Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, *56*(1-2), 508–517. https://doi.org/10.1080/00207543.2017.1351644

Lechevalier, D., Shin, S.-J., Rachuri, S., Foufou, S., Lee, Y. T., & Bouras, A. (2019). Simulating a virtual machining model in an agent-based model for advanced analytics. *Journal of Intelligent Manufacturing*, *30*(4), 1937–1955. https://doi.org/10.1007/s10845-017-1363-x

Leedy, P. D., Ormrod, J. E., & Johnson, L. R. (2016). *Practical research planning and design* (12th ed.). Pearson Education.

Lever, J., Krzywinski, M., & Altman, N. (2016). Points of significance: Model selection and overfitting [Report]. *Nature Methods*, *13*, 703–704. https://doi.org/10.1038/nmeth.3968

Li, J., & Tao, D. (2013). Simple exponential family PCA. *IEEE Transactions on Neural Networks and Learning Systems*, *24*(3), 485–497. https://doi.org/10.1109/TNNLS.2012.2234134

Li, Z., Ma, X., & Xin, H. (2017). Feature engineering of machine-learning chemisorption models for catalyst design. *Catalysis Today*, *280*, 232–238. https://doi.org/10.1016/j.cattod.2016.04.013

Liu, Y., Blome, C., Sanderson, J., & Paulraj, A. (2018). Supply chain integration capabilities, green design strategy and performance: a comparative study in the auto industry. *Supply Chain Management: An International Journal*, *23*(5), 431–443. https://doi.org/10.1108/SCM-03-2018-0095

Louridas, P., & Ebert, C. (2016). Machine learning. *IEEE Software*, *33*(5), 110–115. https://doi.org/10.1109/MS.2016.114

Ma-Lin, S., Fisher, R., Jian-Lin, W., & Lian-Biao, C. (2018). Environmental performance evaluation with big data: theories and methods. *Annals of Operations Research*, *270*(1–2), 459–472. https://doi.org/10.1007/s10479-016-2158-8

Machado, G. V., Cunha, Í., Pereira, A. C. M., & Oliveira, L. B. (2019). DOD-ETL: distributed on-demand ETL for near real-time business intelligence. *Journal of Internet Services and Applications*, *10*(1), 1–15. https://doi.org/10.1186/s13174-019-0121-z

Marshall, C. (2016). *Designing qualitative research* (6th ed.). Sage.

Maslach, C., & Leiter, M. P. (2016). *Burnout* (3rd ed.). Gale Cengage Learning.

Merkel, G., Povinelli, R., & Brown, R. (2018). Short-term load forecasting of natural gas with deep neural network regression. *Energies, 11*(8). https://doi.org/10.3390/en11082008

Merriam, S. B., & Tisdell, E. (2016). *Qualitative research: A guide to design and implementation*. John Wiley & Sons, Inc.

Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2017). Smart manufacturing: Characteristics, technologies and enabling factors. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *233*(5), 1342–1361. https://doi.org/10.1177/0954405417736547

Mittleman, D., & Druckenmiller, D. (2019). *Annual editions: Technologies, social media, and society*. McGraw-Hill/Dushkin.

Montano, D. E., & Kasprzyk, D. (2008). *Theory of reasoned action, theory of planned behavior, and the integrated behavioral model.* In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education: Theory, research, and practice* (pp. 67–96). Jossey-Bass.

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, *5*(2), 147–149. https://doi.org/10.1177/104973239500500201

Moustakas, C. (1994). *Phenomenological research methods*. Sage. https://doi.org/10.4135/9781412995658

Muthamia, S., Lewa, P., & Ndwiga, M. (2015). Effects of transformational leadership on work stress among top managers in state owned enterprises (SOEs) in Kenya. *European Journal of Business and Management*, *7*(36), 185–194. https://core.ac.uk/download/pdf/234627015.pdf

Muthee, V., Bochner, A. F., Osterman, A., Liku, N., Akhwale, W., Kwach, J., Prachi, M., Wamicwe, J., Odhiambo, J., Onyango, F., & Puttkammer, N. (2018). The impact of routine data quality assessments on electronic medical record data quality in Kenya. *PLoS ONE*, *13*, 4. https://doi.org/10.1371/journal.pone.0195362

Naganathan, H., Chong, W. O., & Chen, X. (2016). Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches. *Automation in Construction*, *72*, 187–194. https://doi.org/10.1016/j.autcon.2016.08.002

Nagy, B., Petra, D., Galata, D. L., Démuth, B., Borbás, E., Marosi, G., Nagy, Z. K., & Farkas, A. (2019). Application of artificial neural networks for Process Analytical Technology-based dissolution testing. *International Journal of Pharmaceutics*, *567*, 118464. https://doi.org/10.1016/j.ijpharm.2019.118464

Nakano, Y., Katakuse, Y., & Azechi, Y. (2018). An Application of X-ray Fluorescence as Process Analytical Technology (PAT) to Monitor Particle Coating Processes. *Chemical and Pharmaceutical Bulletin, advpub*. https://doi.org/10.1248/cpb.c17-00906

Nastasoiu, M. A., Bendle, N., & Vandenbosch, M. (2019). Improving measurement with big data: Variety-seeking and survival. *Applied Marketing Analytics*, *4*(3), 253–263. https://www.ingentaconnect.com/content/hsp/ama/2019/00000004/00000003/art00008

New International Version Bible. (1978/2011). Zondervan. https://www.biblegateway.com/ (Original work published 1978)

Nkwake, A. M. (2013). *Working with assumptions in international development program evaluation*. Springer.

O'Grady, G., Clandinin, D. J., & O'Toole, J. (2018). Engaging in educational narrative inquiry: making visible alternative knowledge. *Irish Educational Studies: Engaging in Narrative Inquiry: Making Visible Alternative Knowledge*, *37*(2), 153–157. https://doi.org/10.1080/03323315.2018.1475149

O'Shea, N., O'Callaghan, T. F., & Tobin, J. T. (2019). The application of process analytical technologies (PAT) to the dairy industry for real time product characterization - process viscometry. *Innovative Food Science & Emerging Technologies*, *55*, 48–56. https://doi.org/10.1016/j.ifset.2019.05.003

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Ord, K. (2020). Data adjustments, overfitting and representativeness. *International Journal of Forecasting*, *36*(1), 195–196. https://doi.org/10.1016/j.ijforecast.2019.02.014

Östlund, U., Kidd, L., Wengström, Y., & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: A methodological review. *International Journal of Nursing Studies*, *48*(3), 369–383. https://doi.org/10.1016/j.ijnurstu.2010.10.005

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Papageorgiou, K., I., Papageorgiou, E., Poczeta, K., Bochtis, D., & Stamoulis, G. (2020). Forecasting of day-ahead natural gas consumption demand in Greece using adaptive neuro-fuzzy inference system. *Energies*, *13*(9). https://doi.org/10.3390/en13092317

Pathak, M., & Rathore, A. S. (2017). Implementation of a fluorescence based PAT control for fouling of protein A chromatography resin. *Journal of Chemical Technology & Biotechnology*, *92*(11), 2799–2807. https://doi.org/10.1002/jctb.5358

Patton, M. Q. (2015). *Qualitative research & evaluation methods*. Sage.

Peres, R. S., Dionisio Rocha, A., Leitao, P., & Barata, J. (2018). IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0. *Computers in Industry*, *101*, 138–146. https://doi.org/10.1016/j.compind.2018.07.004

Perez-Castillo, R., Carretero, A. G., Caballero, I., Rodriguez, M., Piattini, M., Mate, A., Kim, S., & Lee, D. (2018). DAQUA-MASS: An ISO 8000-61 based data quality management methodology for sensor data. *Sensors*, *18*(3105). https://doi.org/10.3390/s18093105

Phelps, R. I. (1986). Artificial intelligence-an overview of similarities with O.R. *The Journal of the Operational Research Society*, *37*(1), 13–20. https://doi.org/10.2307/2582542

Pynoo, B., & van Braak, J. (2014). Predicting teachers' generative and receptive use of an educational portal by intention, attitude and self-reported use. *Computers in Human Behavior*, *34*, 315–322. https://doi.org/10.1016/j.chb.2013.12.024

Rabadan, R. (2020). *Understanding Coronavirus*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/9781108920254

Rao, T. R., Mitra, P., Bhatt, R., & Goswami, A. (2019). The big data system, components, tools, and technologies: A survey. *Knowledge and Information Systems*, *60*(3), 1165–1245. https://doi.org/10.1007/s10115-018-1248-0

Reeves, S., Kuper, A., & Hodges, B. D. (2008). Qualitative research methodologies: Ethnography. *BMJ: British Medical Journal (Online)*, *337*. https://doi.org/10.1136/bmj.a1020

Rehman, M. H. u., Liew, C. S., Wah, T. Y., & Khan, M. K. (2017). Towards next-generation heterogeneous mobile data stream mining applications: Opportunities, challenges, and future research directions. *Journal of Network and Computer Applications*, *79*, 1–24. https://doi.org/10.1016/j.jnca.2016.11.031

Richter, A., Ulrica von Thiele, S., Lornudd, C., Lundmark, R., Mosson, R., & Hasson, H. (2016). iLead-a transformational leadership intervention to train healthcare managers implementation leadership. *Implementation Science*, *11*. https://doi.org/10.1186/s13012-016-0475-6

Robson, C., & McCartan, K. (2016). *Real world research: A resource for users of social research methods in applied settings* (4th ed.). Wiley.

Roßmann, B., Canzaniello, A., von der Gracht, H., & Hartmann, E. (2018). The future and social impact of big data analytics in supply chain management: Results from a delphi study. *Technological Forecasting & Social Change*, *130*, 135–149. https://doi.org/10.1016/j.techfore.2017.10.005

Sánchez-Camargo, A. d. P., Parada-Alonso, F., Ibáñez, E., & Cifuentes, A. (2019). Recent applications of on-line supercritical fluid extraction coupled to advanced analytical techniques for compounds extraction and identification. *Journal of Separation Science*, *42*(1), 243–257. https://doi.org/10.1002/jssc.201800729

Savage, N. (2015). Mobile data: Made to measure. *Nature*, *527*(7576), S12–S13. https://doi.org/10.1038/527S12a

Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, *14*(3), 204–220. https://doi.org/10.1108/13620430910966406

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & van Knippenberg, D. (2008). The role of transformational leadership in enhancing team reflexivity. *Human Relations*, *61*(11), 1593–1616. https://doi.org/10.1177/0018726708096639

Sendawula, K., Saadat Nakyejwe, K., Juma, B., & Muganga, G. N. (2018). Training, employee engagement and employee performance: Evidence from Uganda's health sector. *Cogent Business & Management*, *5*(1). https://doi.org/10.1080/23311975.2018.1470891

Shahzad, M., & Liu, A. X. (2015). Probabilistic optimal tree hopping for RFID identification. *IEEE/ACM Transactions on Networking*, *23*(3), 796–809. https://doi.org/10.1109/TNET.2014.2308873

Shen, Y., Ju, C., Koh, T. Y., Rowlinson, S., & Bridge, A. J. (2017). The impact of transformational leadership on safety climate and individual safety behavior on construction sites. *International Journal of Environmental Research and Public Health*, *14*(1), 45. https://doi.org/10.3390/ijerph14010045

Shi, H., Wang, S., & Zhao, D. (2017). Exploring urban resident's vehicular PM2.5 reduction behavior intention: An application of the extended theory of planned behavior. *Journal of Cleaner Production*, *147*, 603–613. https://doi.org/10.1016/j.jclepro.2017.01.108

Shinozuka, M., Papakonstantinou, K. G., Torbol, M., & Kim, S. (2015). Real-time remote monitoring: the DuraMote platform and experiments towards future, advanced, large-scale SCADA systems. *Structure and Infrastructure Engineering*, *11*(4), 588–603. https://doi.org/10.1080/15732479.2014.951861

Simon, M., & Goes, J. (2018). *Dissertation and scholarly research: Recipes for success*. Dissertation Success, LLC.

Spear, B. (2014). Coal – Parent of the Industrial Revolution in Great Britain: The early patent history. *World Patent Information*, *39*, 85–88. https://doi.org/10.1016/j.wpi.2014.06.002

Stake, R. E. (1995). *The art of case study research*. Sage.

Stake, R. E. (2010). *Qualitative research: Studying how things work*. Guilford Publications.

Stark, L. (2018). Algorithmic psychometrics and the scalable subject. *Social Studies of Science*, *48*(2), 204–231. https://doi.org/10.1177/0306312718772094

Su, H., Zio, E., Zhang, J., Yang, Z., Li, X., & Zhang, Z. (2018). A systematic hybrid method for real-time prediction of system conditions in natural gas pipeline networks. *Journal of Natural Gas Science and Engineering*, *57*, 31–44. https://doi.org/10.1016/j.jngse.2018.06.033

Sun, D., Yan, H., Gao, S., Liu, X., & Buyya, R. (2018). Rethinking elastic online scheduling of big data streaming applications over high-velocity continuous data streams. *The Journal of Supercomputing*, *74*(2), 615–636. https://doi.org/10.1007/s11227-017-2151-2

Sun, S., Cegielski, C. G., Jia, L., & Hall, D. J. (2018). Understanding the factors affecting the organizational adoption of big data. *Journal of Computer Information Systems*, *58*(3), 193–203. https://doi.org/10.1080/08874417.2016.1222891

Tasca, L. C., Freitas, E. P. d., & Wagner, F. R. (2018). Enhanced architecture for programmable logic controllers targeting performance improvements. *Microprocessors and Microsystems*, *61*, 306–315. https://doi.org/10.1016/j.micpro.2018.06.007

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly*, *15*(1), 125–143. https://doi.org/10.2307/249443

Thomson, R. (2011). *Did the telegraph lead electrification? Industry and science in American innovation.* Wilmington.

Tourish, D., & Pinnington, A. (2002). Transformational leadership, corporate cultism and the spirituality paradigm: An unholy trinity in the workplace? *Human Relations*, *55*(2), 147–172. https://doi.org/10.1177/0018726702055002181

Unluer, S. (2012). Being an insider researcher while conducting case study research. *The Qualitative Report*, *17*(29), 1–14. https://doi.org/10.46743/2160-3715/2012.1752

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89. https://doi.org/10.1016/0378-8733(95)00256-1

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 29, pp. 271–360). Academic Press. https://doi.org/10.1016/S0065-2601(08)60019-2

van Assen, M. F. (2021). Training, employee involvement and continuous improvement – the moderating effect of a common improvement method. *Production Planning & Control*, *32*(2), 132–144. https://doi.org/10.1080/09537287.2020.1716405

Vandepitte, L., Bosch, S., Tyberghein, L., Waumans, F., Vanhoorne, B., Hernandez, F., De Clerck, O., & Mees, J. (2015). Fishing for data and sorting the catch: Assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. *Database*, *2015*. https://doi.org/10.1093/database/bau125

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Wang, H., Wang, Z., Li, N., & Kong, X. (2019). Efficient OLAP algorithms on GPU-accelerated Hadoop clusters. *Distributed and Parallel Databases*, *37*(4), 507–542. https://doi.org/10.1007/s10619-018-7239-z

Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factory of industrie 4.0: An outlook. *International Journal of Distributed Sensor Networks*, *12*(1), 3159805. https://doi.org/10.1155/2016/3159805

Watson, D. S., Kerchner, K. R., Gant, S. S., Pedersen, J. W., Hamburger, J. B., Ortigosa, A. D., & Potgieter, T. I. (2016). At-line process analytical technology (PAT) for more efficient scale up of biopharmaceutical microfiltration unit operations. *Biotechnology Progress*, *32*(1), 108–115. https://doi.org/10.1002/btpr.2193

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, *40*(9), 96–99. https://doi.org/10.1109/MC.2007.331

Wei, N., Li, C., Duan, J., Liu, J., & Zeng, F. (2019). Daily natural gas load forecasting based on a hybrid deep learning model. *Energies*, *12*(2). https://doi.org/10.3390/en12020218

Wilson, J. M. (2016). The origin of material requirements planning in Frederick W. Taylor's planning office. *International Journal of Production Research*, *54*(5), 1535–1553. https://doi.org/10.1080/00207543.2015.1092616

Wojnar, D. M., & Swanson, K. M. (2007). Phenomenology: An exploration. *Journal of Holistic Nursing*, *25*(3), 172–180. https://doi.org/10.1177/0898010106295172

Wortmann, F., & Flüchter, K. (2015). Internet of things: Technology and value added. *Business & Information Systems Engineering*, *57*(3), 221–224. https://doi.org/10.1007/s12599-015-0383-3

Wren, D. A. (2011). The centennial of Frederick W. Taylor's the principles of Scientific Management: A retrospective commentary. *Journal of Business and Management*, *17*(1), 11–22. http://jbm.johogo.com/pdf/volume/1701/JBM-1701-02-full.pdf

Wu, X., Zhu, X., Wu, G., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Xiao, F., Wang, Z., Ye, N., Wang, R., & Li, X. (2018). One more tag enables fine-grained RFID localization and tracking. *IEEE/ACM Transactions on Networking*, *26*(1), 161–174. https://doi.org/10.1109/TNET.2017.2766526

Xie, X., Zhang, A., Wen, L., & Bin, P. (2019). How horizontal integration affects transaction costs of rural collective construction land market? An empirical analysis in Nanhai District, Guangdong Province, China. *Land Use Policy*, *82*, 138–146. https://doi.org/10.1016/j.landusepol.2018.11.029

Yates, J., & Leggett, T. (2016). Qualitative research: An introduction. *Radiologic Technology*, *88*(2), 225–231. http://www.radiologictechnology.org/content/88/2/225.extract

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Yin, Y., Stecke, K. E., & Li, D. (2018). The evolution of production systems from Industry 2.0 through Industry 4.0. *International Journal of Production Research*, *56*(1–2), 848–861. https://doi.org/10.1080/00207543.2017.1403664

Yizhong, X., Baranchenko, Y., Lin, Z., Lau Chi, K., & Ma, J. (2019). The influences of transformational leadership on employee employability: Evidence from China. *Employee Relations*, *41*(1), 101–118. https://doi.org/10.1108/ER-02-2018-0052

Zhang, X., Wu, C., & Zuo, L. (2016). Minimizing fuel consumption of a gas pipeline in transient states by dynamic programming. *Journal of Natural Gas Science and Engineering*, *28*, 193–203. https://doi.org/10.1016/j.jngse.2015.11.035

Zhou, W., Piramuthu, S., Chu, F., & Chu, C. (2017). RFID-enabled flexible warehousing. *Decision Support Systems*, *98*, 99–112. https://doi.org/10.1016/j.dss.2017.05.002

Zhuang, Y., Wang, Y., Shao, J., Chen, L., Lu, W., Sun, J., Wei, B., & Wu, J. (2016). D-ocean: An unstructured data management system for data ocean environment. *Frontiers of Computer Science*, *10*(2), 353–369. https://doi.org/10.1007/s11704-015-5045-6

Appendix

Interview Guide

START RECORDING

This interview is with (NAME) (POSITION). Please note, that your name will be redacted from the final transcript. Thank you once again for taking the time to participate in the interviews for this research project. This interview is being recorded. As a reminder, my name is BJ Stigall, and I am a doctoral candidate at Liberty University. I am conducting this research project as part of the requirements for the Doctor of Business Administration degree. The purpose of this research project is to discover the improvements realized by implementing an advanced analytics and machine learning (AAML) and the reasons for failure to fully utilize such a system for interstate natural gas pipelines in the United States. The AAML system may not be one unified system for the purposes of this research project.

I would like to remind you that your participation in this research study is voluntary, and your decision to participate in this study will not affect your relationship with Liberty University. You may withdraw from the study at any time. Your decision to withdraw will not affect your relationship with Liberty University.

The record of this interview and any documentation you provide will be kept in a secure location, and any publication of this interview or documents provided will have all identifiable components removed.

This interview is about your experience with AAML systems at ________________ (company name). This interview will explore the environment before the AAML system was implemented, including any previous system or method used, what made this implementation successful, and how the AAML system has helped improve the information available to you and your job performance.

You have already signed the consent form (or given electronic consent). Do you have any questions before we proceed?

Note: verbs will be changed as appropriate for current/former employees.

Demographic Questions (All Participants)

1. Please describe your position and how long you have been/were with [the company under study].

2. Please disclose the decade of your age. For example, if you are 49, you would indicate you are in your 40s.

AML User Questions (All Participants)

1. How long have you/did worked with the Advanced Analytics or Machine Learning System or been the beneficiary of the AAML system?
2. Before coming to this organization, did you have any experience with an Advanced Analytics or Machine Learning System?
   a. Please describe that system.
3. Please describe the way you retrieved information about the [researcher to choose based on the job] (compressors/gas flow) before the AAML system was installed.
   a. How easy was this information to gather?
   b. Was the information available in one location, or did you have to search multiple locations?
   c. Was the system fully implemented?
      i. If not, why?
4. Is/was using this new AAML system part of your job duties? Do you think installing the system was a good idea?
5. Did management make this a requirement of performing your job?
6. Did management give a compelling vision of the future in using AAML?
7. Did management motivate you to use the system?

8. Was management clear in setting goals for an AAML system?
9. Has/was management been helpful in the use of the system?
10. Do/did you find the new system easy to use?
11. Do/did you think it will be easy to master?
12. Was learning how to use the system easy?
13. Do/did you have the resources needed to successfully use the system?
    a. If no, what do you think you need/needed to help with success?
14. Are/were you able to use the system without help most of the time?
15. Is/was there someone to call if you need help?
16. Do/did you consider the new system to be easier to use than the previous system?
17. Do/did you find the current system useful in your job?
18. Are/were you able to perform your job quicker than before?
19. Does/did the system increase your productivity?
20. How has/did the current AAML system improved the analytic capabilities in your job?
    a. Are there any specific examples you can share?
21. Has the AAML system helped to prevent failures from occurring?
    a. Can you provide any specific examples?
    b. If the device had failed, what would have been the results?

Leadership Questions (Executive Leadership)

1. What motivated you to have the AAML system installed at this organization?

a. What were the goals for the AAML project(s)?

2. Please describe the methods used to perform analytics before the AAML system was deployed at this organization.
a. Did these tools give you the results in the time you required?
3. Did you clearly define the purpose for implementing the AAML to your direct reports?
a. What is the purpose?
4. What is the vision for the AAML system?
5. When defining the AAML project(s), did you encourage different viewpoints?
6. Were the rewards clearly defined if performance was met when implementing the AAML system?
7. Do/did you focus on mistakes, exceptions, or variances from the company's standard policies and practices?
8. Do you actively look for exceptions or do you allow them to be brought to your attention before acting?
9. Do/did you support exploring new ways to achieve a task, or do you believe that procedures should not be changed until something goes wrong?
10. Do/did you respond to urgent matters in a timely fashion?

Leadership Questions (Non-Executive Leadership)

1. What is your vision for the AAML system?
2. When defining the AAML project(s), did you encourage different viewpoints?
3. Were the rewards clearly defined if performance was met when implementing the AAML system?

4. Do/did you focus on mistakes, exceptions, or variances from the company's standard policies and practices?
5. Do you actively look for exceptions or do you allow them to be brought to your attention before acting?
6. Do/did you support exploring new ways to achieve a task, or do you believe that procedures should not be changed until something goes wrong?
7. Do/did you respond to urgent matters in a timely fashion?

Leadership Questions (Non-Leadership, Non-Executive Leadership)

1. Does/Did your leadership encourage a sense of purpose for implementing the AAML?

a. What is the purpose?

2. Did your leadership provide a compelling vision for the AAML system?

a. What is the vision for the AAML system?

3. Did your leadership encourage different viewpoints when implementing the AAML system?
4. Does/did your leadership have a clear explanation of what the reward will be if performance goals are met?
5. Does/did the leadership focus on mistakes, exceptions, or variances from the company's standard policies and practices?
6. Does/did the leadership support exploring new ways to achieve a task, or do they believe that procedures should not be changed until something goes wrong?
7. Does/did your leadership respond to urgent requests in a timely manner?

Wrap-up Question (All)

1. And finally, is there anyone else you can recommend to interview for this research project?

Once again, thank you for taking the time to participate in the interview process. My next step will be transcribing the interview. Once I have completed the transcription process, I will share the transcript with you to verify the accuracy of the document. If you feel there are any clarifications, please let me know. If you can think of anything else that would be appropriate to add to this interview, please contact me at [phone number redacted] or e-mail me at [redacted]. Do you have any questions at this time?