Journal of Business Research 90 (2018) 186–195

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

The analytics paradigm in business research

Dursun Delen[a,*], Hamed M. Zolbanin[b]

[a] Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Tulsa, OK 74106, USA
[b] Department of Information Systems and Operations Management, Miller College of Business, Ball State University, Muncie, IN 47306, USA

ARTICLE INFO

Keywords:
Business analytics
Causal-explanatory modeling
Predictive modeling
Big data
Business disciplines
Business research

ABSTRACT

The availability of data in massive collections in recent past not only has enabled data-driven decision-making, but also has created new questions that cannot be addressed effectively with the traditional statistical analysis methods. The traditional scientific research not only has prevented business scholars from working on emerging problems with big and rich data-sets, but also has resulted in irrelevant theory and questionable conclusions; mostly because the traditional method has mainly focused on modeling and analysis/explanation than on the real/practical problem and the data. We believe the lack of due attention to the analytics paradigm can to some extent be attributed to the business scholars' unfamiliarity with the analytics methods/methodologies and the type of questions it can answer. Therefore, our purpose in this paper is to illustrate how analytics, as a complement, rather than a successor, to the traditional research paradigm, can be used to address interesting emerging business research questions.

1. Introduction

The introduction of the commercial Internet and its eventual prevalence over the past two decades has given rise to an influx of data in virtually every domain of the society (Davenport & Kim, 2013). Particularly, the transition from Web 1.0 to Web 2.0 (and to Web 3.0), whereby static pages gave place to user contributed content, inspired organizations all around the globe to invest extensively in infrastructures that improved their ability to collect data throughout and beyond the enterprise. In the business world, this abundance of data has led to increasing interest in almost every industry to develop capabilities and methods for extracting insightful knowledge from data to achieve competitive advantage (Provost & Fawcett, 2013). These new data sources, however, not only are too large and too complex, but also have created new questions that cannot be answered effectively with traditional analysis methods. To overcome these problems, new methodologies and processing techniques were developed that gave birth to a new era in business decision making referred to as the *[business] analytics* (BA) period (Mortenson, Doherty, & Robinson, 2015).

Over the past decade, BA has been regularly reported to be a top priority for many top-level managers (Holsapple, Lee-Post, & Pakath, 2014). Such an interest has not been a fad, but instead a result of compelling evidence corroborating the values of analytics to businesses. For instance, a study by Anderson (2015) showed that every $1.00 spent on analytics applications pays off $13.0. Other studies have also reported complementary benefits and promising contributions of analytics to operations (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011) or productivity of firms via data-driven decision making (Chae, Yang, Olson, & Sheu, 2014; Davenport & Harris, 2007; McAfee & Brynjolfsson, 2012). These findings suggest that developing analytics prowess has become an ineluctable commitment for businesses.

While businesses are at the forefront of employing various facets of analytics, academic research has not fully recognized its potentials. In most business and organizational science journals, research is dominated by certain paradigms that are either traditional and less related to the new analytics approach or adopt a narrow facet of analytics (Holsapple et al., 2014; Putka, Beatty, & Reeder, 2017; Tonidandel, King, & Cortina, 2016). As Shmueli and Koppius (2011) denote, almost all studies in these disciplines have used "*causal-explanatory statistical modeling and statistical inference to test causal hypotheses and to evaluate the explanatory power of underlying causal models*". While these prevalent modeling and problem solving paradigms have generated significant insights over the past decades, they have prevented researchers from working on emerging business problems. Additionally, since the emphasis of academic research has mostly been on modeling and analysis, rather than on the problem and the data, they have resulted in irrelevant theory and questionable conclusions (Breiman, 2001b). For instance, as Shmueli and Koppius (2011, p. 572) denote, several papers published in Information Systems (IS) journals used the discipline's dominant paradigm to make conclusions (e.g., about the predictive

* Corresponding author at: Spears School of Business, Oklahoma State University, 700 North Greenwood Avenue, North Hall 302, Tulsa, OK 74106, USA.
E-mail addresses: dursun.delen@okstate.edu (D. Delen), hmzolbanin@bsu.edu (H.M. Zolbanin).
URL: http://spears.okstate.edu/delen (D. Delen).

https://doi.org/10.1016/j.jbusres.2018.05.013
Received 3 April 2018; Received in revised form 8 May 2018; Accepted 9 May 2018
Available online 25 May 2018
0148-2963/

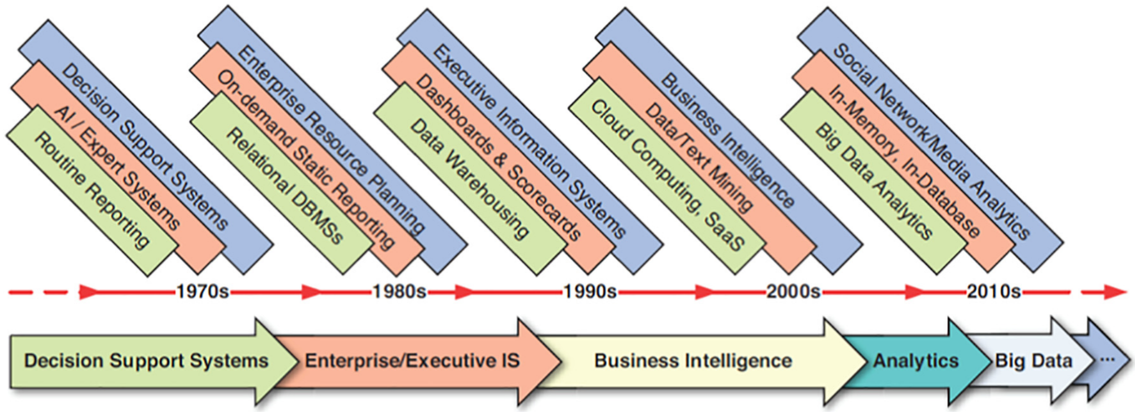

Fig. 1. A longitudinal view of the evolution of analytics.

power of models) that required other analysis approaches to be adequate. With the use of analytics not only can we produce more and more reliable information about the inherent structure of relationships between the focal variables, but also we will be able to generate more relevant research (Breiman, 2001b). Therefore, it is important for business researchers to add other tools besides hammer to their research toolkit, so that not every problem looks like a nail.

Although several reasons have been enumerated for the paucity of analytical studies in business journals, we believe two causes are the most salient. First and foremost, a majority, if not all, of business journals have historically placed a high value on publications that test systems of relationships specified by theory (Aguinis, Pierce, & Culpepper, 2009; Putka et al., 2017). Consequently, researchers have placed greater focus on modeling and analysis than on the problem and the data; leading to an overabundance of structural equation models to the point that other analysis methods are not considered sophisticated – i.e., scientific – enough. Second, and chiefly a ramification of the first cause, most business scholars do not typically receive the training required to understand and apply various business analytics methods during their graduate studies (Putka et al., 2017); and why would they when such methods are given no chance in top-tier business, and particularly management, outlets? Whereas this trend has changed in all industrial sectors and many academic areas, some business disciplines have not yet fully embraced the new analytics paradigm. We believe this trend has to change or those fields will not be able to accurately predict increasingly emerging important outcomes (Breiman, 2001b; Putka & Oswald, 2015), will fail to incorporate into their models some of the key drivers of their phenomena of interest (Putka & Oswald, 2015; Tonidandel et al., 2016), or cannot adequately address model complexity and uncertainty (Breiman, 2001b; Putka & Oswald, 2015).

This paper seeks to address these issues through a bottom-up approach. In other words, our goal is to raise an awareness among business scholars about the various types of questions that can be answered using the emerging business analytics paradigm. We hope through attending to the second cause of unpopularity of analytics, as indicated above, more interest is formed in this area, which in turn, can lead to a greater support for the promotion of this research paradigm among top-tier business and management journals. It deserves to mention, however, that our goal is not to provide a comprehensive overview of business analytics methods, nor a technical explanation of their mechanics or statistical foundations. Instead, we aspire to introduce some of the more common analytics methods used in prior research and provide examples for the business questions that can be addressed by such techniques. Most of the examples we discuss are taken from the information systems literature, which has traditionally been a pioneer in the application of these methods in business research.

Our focus is on promoting analytics as a complement to the traditional theory-driven hypothesis testing in business disciplines, rather than denigrating this research paradigm. To this goal, we structure the paper as follows. First, we review and integrate the various definitions of analytics to form a common understanding of the concept. Next, we provide an overview of the various types of analytics as classified by practitioners and scholars. Subsequently, we present the potential avenues of employing analytics to augment academic research before concluding the paper with a final discussion.

2. An overview of BA and its components

Despite its ostensible primacy in the past few years, analytics is not an entirely new paradigm and has been employed by corporations for several years, albeit in a narrower sense. It can be considered a continuation of efforts among management science scholars and practitioners in the 1940's when optimization and simulation techniques were developed to maximize output with limited resources (Mortenson et al., 2015). With the development of management information and decision support systems in the late 1960s and 1970s, analytics began to command more attention (Delen, 2015) and eventually evolved into an integration of operational research, machine learning, and information systems (Mortenson et al., 2015). Fig. 1 illustrates the evolution of analytics techniques and related terminology over the last few decades. Most researchers and practitioners in the field believe that the latest names for analytics, such as *big data* and its enabling tools/techniques (such as *deep learning, image processing, text mining*, and *sentiment analysis*) are just new names/labels (i.e., buzz words) for business analytics and its enablers, and the goal is still the same - to convert data into actionable insight for more timely and accurate decision support (Sharda, Delen, & Turban, 2017). That said, there is a specific emphasis in *big data*, which is on the volume, variety and variability of the data. Nowadays, the type of data available for analytics poses a variety of challenges (defined within the context of the three Vs), but at the same time, brings opportunities to organizations that are capable of converting them into real business outcomes (data → information → knowledge → action).

With the popularity of different buzzwords, the use of data and computing power for enhanced decision-making has borne various monikers in recent years. While decision support systems, expert systems, business intelligence, data and text mining, big data, and deep learning have been used to refer to certain techniques and technologies, they all share the same underlying purpose: employing internal or external, structured or unstructured data for actionable insights. Consequently, in this paper we use *analytics* loosely to refer to all such applications of data for better decision making. Therefore, our focus is on the common *process* and *purpose*, rather than different specifics, of such techniques.

The need to make data-driven decisions in a myriad of application areas, multidisciplinary nature, and multidimensionality of analytics has resulted in a multitude of definitions for the term over the past 15 years, leading to a state of confusion and a lack of consensus about the essence and scope of analytics. In one of the earliest accounts of analytics, where the authors used the term interchangeably with data mining, analytics was defined as "the general process of exploration and analysis of data to discover new and meaningful patterns in data" (Kohavi, Rothleder, & Simoudis, 2002). It was not until October 2005 that the term, along with some of its extensions, became popular and started to appear in Google trends (Piatetsky-Shapiro, 2007). Shortly after the publicity of analytics, more scholars and practitioners sought to engage with it, or at least adopt the moniker of analytics (Mortenson et al., 2015). This trend led to more detailed and customized definitions. Appendix A provides a chronological order of some of the most cited descriptions of analytics.

A closer look at the definitions given in Appendix A reveals that although there are different perceptions about the nature and scope of analytics, there is a general agreement that it involves data-driven decision making (Holsapple et al., 2014). Additionally, these definitions suggest that analytics has been viewed as a process that takes various forms of data as input and generates value for the firm as output. Several potential values have been enumerated for the application of analytics, including better decision-making, improved organizational performance, enhanced competitiveness, and realization of organizational objectives. We believe enhanced decisions and improved performance can potentially lead to other outcomes. Therefore, by aggregating the definitions given in Appendix A, we define analytics as a process that employs various techniques to analyze and interpret different forms of data to enable better decisions and improve firm performance.

The key component of analytics is the process in which a set of various techniques transform data into value. As opposed to the dominant (traditional) research paradigm that emphasizes on the use of certain approaches (i.e., pre-specified causal models) to problem solving, proper techniques in analytics are selected regarding the problem and the data. Several dimensions, such as domain, method, and orientation have been proposed for analytics (Holsapple et al., 2014). The domain dimension refers to fields and areas in which analytics is employed, and the method dimension highlights the approaches used to analyze the data. The orientation dimension refers to the line of thought and is not idiosyncratic to one or another business domain (Holsapple et al., 2014). This dimension is the most common taxonomy of analytics techniques and divides it into three dimensions (Lustig, Dietrich, Johnson, & Dziekan, 2010): *descriptive*, *predictive*, *prescriptive*. Recent revisions have further divided analytics with regards to its orientation into four dimensions by adding a *diagnostic* component (Banerjee, Bandyopadhyay, & Acharya, 2013). Fig. 2 depicts a simple taxonomical view of analytics.

Descriptive analytics, also called business reporting or business intelligence (Delen & Demirkan, 2013), is a preliminary stage of data processing that creates a summary of historical data to yield useful information about the business performance and prepare the data for further analysis (Rouse, 2015). It attempts to examine the data content to answer the questions of "what happened?" or "what is happening?" Therefore, descriptive analytics mostly includes the traditional business intelligence (Mortenson et al., 2015) and visualization techniques (Gartner, 2016; Shmueli, Patel, & Bruce, 2016) and helps researchers identify nontrivial patterns and relationships in data.

Diagnostic analytics (as a natural extension of descriptive analytics) examines data or content to answer the question "why did it happen?" It needs exploratory data analysis of the existing data - or additional data if required to be collected - using such tools and techniques as visualization, drill-down, data discovery, and data mining in order to discover the root causes of a problem (Banerjee et al., 2013). In this sense, diagnostic analytics is closely related to the descriptive dimension as well as the traditional BI.

Predictive analytics refers to the building and assessment of algorithmic models that aim at making empirical, rather than theoretical, predictions (Breiman, 2001b; Shmueli & Koppius, 2011). In contrast to explanatory statistical models that are built to test causal hypotheses, predictive models are designed to predict future observations (Shmueli & Koppius, 2011). Therefore, as opposed to explanatory models' retrospective approach to extract knowledge from data, predictive analytics employs a prospective approach to specify the values of new observations based on the structure of the relationship between inputs and outputs (Breiman, 2001b).

Prescriptive analytics follows from descriptive and predictive analytics to find the best course of action under certain circumstances. It involves a set of mathematical techniques that computationally determine the optimal action or decision given a complex set of objectives, requirements, and constraints, with the goal of improving business performance (Lustig et al., 2010). Prescriptive analytics can also suggest decision options for how to take advantage of a future opportunity or mitigate a future risk, and illustrate the implications of each decision option. In practice, prescriptive analytics can continually and automatically process new data to improve the recommendations and provide better decision options (Rouse, 2012).

From this introduction on analytics and its different dimensions, it should be clear that analytics and the scientific method - the dominant paradigm in scientific research for hundreds of years - belong to different epistemologies. Therefore, in the following section, we point out to some of the important distinctions of these research paradigms. While each of these approaches have their protagonists and antagonists, resulting in a debate that has particularly heightened in recent years because of the availability of data in massive amounts, our purpose is to champion their complementarity, rather than disparaging the scientific method.

3. Analytics versus the scientific method

The expected outcome of business analytics, especially within the context of complementing traditional business research, is the discovery of new relationships (e.g., correlations) that emerge from large and feature-rich data, which can then be used to develop new theories for further statistical analysis and testing. In light of the recent developments in tools and technologies used to analyze large collections of data, Peter Norvig, Google's research director, took George Box's (Box, 1976) well-known proclamation, "*all models are wrong, but some are useful*," one step further (Norvig, 2009):

> "*If the model is going to be wrong anyway, why not see if you can get the computer to quickly learn a model from the data, rather than have a human laboriously derive a model from a lot of thought … Having more data, and more ways to process it, means that we can develop different kinds of theories and models. But that does not mean we throw out the scientific method. It is not "The End of Theory." It is an important change (or addition) in the methodology and set of tools that are used by science, and perhaps a change in our stereotype of scientific discovery.*"

What Norvig says is simply a reiteration of what Breiman (2001b) had said a few years earlier: when the goal is to reach conclusions from data, the researcher is not necessarily required to first build a model and a set of hypotheses, and then collect data to confirm it. This could be an important reason behind the ostensible reluctance in many disciplines to embrace analytics as a research method, because accepting that a model may not precede statistical analyses seems to be in contrast to how the scientific method has worked for hundreds of years (Anderson, 2008):

> "*Scientists are trained to recognize that correlation is not causation, that no conclusions should be drawn simply in the basis of correlation between X and Y (it could be just a coincidence). Instead, you must understand the underlying mechanisms that connect the two. Once you have a model, you can connect the data sets with confidence. Data without a*

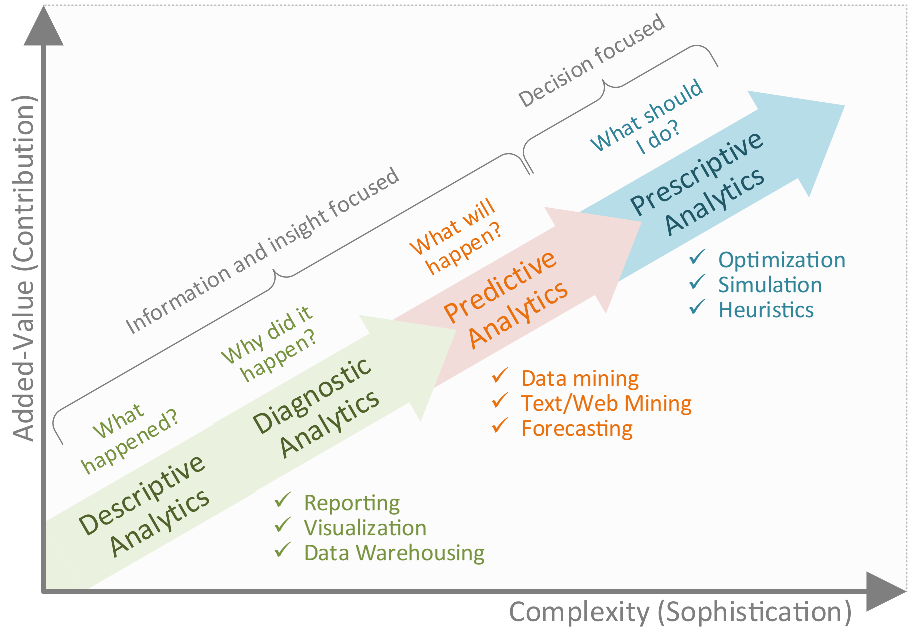

Fig. 2. Different types (sophistication levels) of business analytics.

model is just noise."

The availability of data, which can at times become overwhelmingly large, has introduced a new approach to scientific inquiry. For instance, from the perspective of theory building, analytics methods can offer new ways of observing reality through parsing large sets of potentially criterion-relevant data elements; thus, making series of smaller discoveries that may later be inductively synthesized into new theories (Berk, 2006; Locke, 2007; Norvig, 2009; Putka et al., 2017; Shmueli & Koppius, 2011). Within this paradigm, it is not surprising to see data precede the development of a model (Breiman, 2001b) or generation of a new theory (Shmueli & Koppius, 2011). The new theories would then have the luxury of corroboration by larger, more representative sets of data.

Based on the studies cited in this section, we summarize the major differences between the traditional scientific and the emerging analytical research; differences that should be considered when reviewing analytics research:

- In analytics research, data *may* precede theory or a model.
- Although analytics can lead to the development of new theories, it does so by focusing on the complex relationships and patterns present in the data rather than on hypothesizing.
- Metrics used to evaluate analytic models are different than those used in the traditional research methods.

Additionally, use of analytics as a complement to traditional methods in scientific research not only can prevent the development of irrelevant theories and questionable conclusions (Breiman, 2001b), but also can offer a level of transparency that does not currently exist with regard to post hoc theorizing (Tonidandel et al., 2016), where "numerous published findings and their corresponding theories are likely derived from the data via HARKing – hypothesizing after the results are known." (Kerr et al., 1998; Tonidandel et al., 2016).

The research in analytics is not limited to the complementary role that it plays in supporting traditional statistics and/or theory-development based business research. Generally speaking, business analytics research spans from developing algorithms (new and improved algorithms that are more suitable, faster and/or more accurate than the existing ones) to designing methodologies (for converting a wide variety of data types and sources into actionable insights) and building applications/solutions (innovative and best-practice approaches to the development of solutions to seemingly unsolvable business problems).

With this introduction on the analytics research, we now turn our attention to some of the more common, and relatively newer, analytic methods that have been used in business research over the past two decades. Most of these methods belong to the descriptive and predictive classes.

4. How can analytics be incorporated into business research?

Based on the definitions provided previously, it is clear that descriptive and diagnostic analytics have an exploratory nature, whereas the predictive and prescriptive dimensions involve modeling and mathematical computations. However, predictive and prescriptive dimensions vary in the type of modeling techniques they normally use. In the following sub-sections, we explain how each of the analytics dimensions have been used to complement traditional business research.

4.1. Descriptive (and diagnostic) analytics in business research

Descriptive analytics has long been employed in business research; however, recent developments in computing and machine learning capabilities have enabled its widespread application on large organizational data for enhanced knowledge discovery and improved decision-making. Among the various descriptive methods of analytics, association analysis, sequence analysis, clustering, and link analysis, are the most common in business research. In the following subsections, we review some of the applications of these methods.

4.1.1. Association analysis

Depending on the application domain, association analysis is known under various names: market-basket analysis, co-occurrence grouping, association rule mining, and association discovery are the most common aliases for association analysis. This method attempts to find the relations between entities based on transactions or events that involve them (Provost & Fawcett, 2013). A well-known example for the application of this method is Amazon's cross-selling offers. By mining its transactional data warehouse, Amazon is able to identify other items that appear more frequently with the item of interest in prior customers' shopping carts than would be expected by chance. This approach enhances customer experience and loyalty, while increasing Amazon's revenue.

In academic research, association analysis has been applied to various interesting problems, including law enforcement (Kaza, Wang, & Chen, 2007), outlier detection in categorical data (Pai, Wu, & Hsueh, 2014), comparison of products across online platforms (Huang & Tsai, 2011), and item placement in retailing stores (Chen, Chen, & Tung, 2006). In the context of law enforcement, customs and border protection officers believe that vehicles involved in smuggling drugs and other products from Mexico into the United States operate in groups to increase their chances of successful crossing or minimize the probability of being caught (Kaza, Hu, Hu, & Chen, 2011). Association analysis has been used in such law enforcement scenarios to find out which vehicles crossed the border regularly with a certain vehicle known to the officials to be involved in smuggling (Kaza et al., 2007). In an application similar to Amazon and other e-retailer's use of association analysis, this method was used to construct a multilingual ontology that helped shoppers compare products offered on platforms with different languages (Huang & Tsai, 2011). And as a final example, association analysis has been employed to understand the relationship between the relative spatial distance of displayed products and items' unit sales in a retailer's store, leading to recommendations for retailing and merchandising (Chen et al., 2006).

4.1.2. Sequence analysis

Sequence analysis, also known as sequential pattern mining, discovers patterns of frequent subsequences in a database of events or transactions. A sequence analysis algorithm mines the database records looking for recurrent patterns that occur in a certain order (Mabroukeh & Ezeife, 2010). In this sense, sequence analysis is a special case of association analysis in which the order of events or transactions is also taken into account. Sequence analysis is used to address important business problems with broad applications in analyzing customer behavior, web access and surfing patterns, disease treatment, and natural disasters (Mabroukeh & Ezeife, 2010).

Similar to the business applications of sequence analysis, scholars have applied this method to a variety of domains. For instance, it has been used to study information technology (IT) career histories, mobility patterns, and career success (Joseph, Fong Boh, Ang, & Slaughter, 2012), leading to a better understanding of the concept of *boundaryless* careers for IT professionals. In another study (Prinzie & Van den Poel, 2006), sequence analysis was used as a data preprocessing step in a binary classification model that enabled the use of invariant (with respect to time) and sequential independent variables to improve the prediction accuracy. This application resulted in an enhanced decision support system for customer retention of an international financial services provider. As a final example, sequence analysis, in conjunction with other methods, was used to improve the prediction of the product group of home appliances that customers would purchase next (Prinzie & Van den Poel, 2007).

4.1.3. Clustering

Clustering (also known as cluster analysis) is perhaps the most known descriptive analytics method to traditional researchers. The purpose of clustering is to group a set of objects such that those in the same group are more similar (based on a certain criterion) to each other than to the objects in other groups. Some of the innovative academic applications of clustering include reviewing and grouping of relevant literature (Miaskiewicz & Monarchi, 2008), grouping of smart meter data by utility providers to enable customer specific services (Flath, Nicolay, Conte, Dinther, & Filipova-Neumann, 2012), graph-based clustering of similar questions for better social question answering (John, Goh, Chua, & Wickramasinghe, 2016), segmentation and profiling of Facebook page fans based on their "liking" behavior for enhanced customer relationship management (van Dam & van de Velden, 2015), identification of categories or groups of residents in an elderly nursing home based on their degree of autonomy/disability (Combes & Azema, 2013), improved sales forecasting in the textile industry (Thomassey & Fiordaliso, 2006), and building a decision support tool to identify AACSB peer schools for management education (re)accreditation (Kiang, Fisher, Chen, Fisher, & Chi, 2009).

As an unsupervised method, clustering is inherently based on measures of similarity and distance between objects. If used in supervised learning, these measures may contribute to another method called *similarity matching*, which seeks to recognize similar entities based on the information known about them. It is based on the premise that if two entities (individuals, products, services, organizations) are similar in some way, they should share other characteristics as well. This method is valuable to businesses as it allows them to find companies that are similar to their best business customers (Provost & Fawcett, 2013). In retailing, a combination of similarity matching with other methods, such as association analysis, provides the basis for some of the most popular methods for making product recommendations. Use of this method in business research is less common than other techniques. As an example, we can point to (Hwang & Tang, 2004), where the authors have used similarity matching to determine how a special case in workflow management should be handled by retrieving the solution to an older case whose characteristics are the most similar to the current case.

4.1.4. Link analysis

Link analysis is a data analysis technique used to evaluate or predict connections between entities in a network of objects. Such connections generally point to the existence of some sort of relationship between the network nodes, which can be of various types, including people, organizations, or transactions. Link analysis can also estimate the strength of the connections. This technique has been used in a variety of applications such as movie recommendation, criminal investigation (counterterrorism, kidnapping, narcotics violation, and fraud detection), cybersecurity, web page filtering, medical informatics, and marketing research.

Following the 2001 terrorist attacks, link analysis came to prominence in academic research to contribute possible technological solutions for uncovering terrorist networks and to enhance public safety and national security (Schroeder, Xu, Chen, & Chau, 2007; Xu & Chen, 2004; Xu & Chen, 2005a, 2005b; Xu, Hu, & Chen, 2009). The general idea behind the use of link analysis in these efforts is that if two or more criminals share a common node in their network of relationships, then chances are high that the shared node is also involved in illegal or terrorist activities. Successful applications of this technique led to its widespread use in other domains. In web page filtering, for instance, link analysis was combined with other methods to filter out irrelevant documents from a set of documents collected from the web (Chau & Chen, 2008). Link analysis has also been applied in marketing research to investigate a network of customer preferences to co-purchase pairs of products (Dhar, Geva, Oestreicher-Singer, & Sundararajan, 2014). In this application, data on products' historical demand was aggregated with data on the demand for their immediate neighbors in the network to improve prediction of each product's future demand levels.

4.2. Predictive analytics in business research

Predictive analytics refers to the building and assessment of models that seek to make empirical, as opposed to theoretical, predictions (Shmueli & Koppius, 2011). Although predictive analytics is inherently different from the dominant explanatory modeling paradigm in all its steps, many studies have erroneously used an explanatory approach for empirical predictions. As (Shmueli & Koppius, 2011) denote, simultaneous satisfaction of two predictive modeling criteria is necessary for having predictive power for future samples:

Table 1
Roles of predictive analytics in business research.

Role	Brief description
Generating new theory	Many data sources that are available today are large, rich and detailed, and include multiple types. The patterns and relationships that these data contain are often complex and hard to hypothesize. Predictive analytics can be used to uncover unknown relationships in such data and develop new theoretical models.
Developing measures	Predictive analytics can be used to compare different operationalizations of constructs or different measurement instruments.
Comparing competitive theories	Comparing predictive accuracy of competing theories is possible with predictive analytics.
Improving existing models	Is possible through capturing complex underlying relationships.
Assessing relevance	Predictive analytics is a tool for assessing the distance between theory and practice.
Assessing predictability	Predictive analytics can be used to develop benchmarks of predictive accuracy.

1. Out-of-sample assessment of predictive accuracy (e.g., cross-validation on a hold-out sample)
2. Use of adequate predictive measures (e.g., RMSE, MAPE, PRESS, or overall accuracy rather than *p*-value or R^2).

However, many academic business studies have either not correctly assessed the accuracy of their predictions (i.e., they lack an out-of-sample assessment) or have used such inappropriate measures as R^2 or *p*-value. We refer the readers to (Shmueli & Koppius, 2011) for a complete introduction of predictive analytics and its differences with explanatory modeling. It deserves, however, to point out briefly to the roles of predictive analytics in scientific research, as outlined in that manuscript. As explained in (Shmueli & Koppius, 2011), predictive analytics enables the development and examination of theoretical models through a different lens and therefore, provides a valuable addition (besides explanatory modeling) to scientific research. Table 1 summarizes these roles.

Another justification for the necessity of analytics, especially predictive analytics, in business research is the threat the sheer use of explanatory models poses to scientific research. Most, if not all, explanatory studies confirm their hypotheses and validate their findings by checking data-model fit using goodness-of-fit tests and residual analysis. However, experiments showed that when the relationship between the dependent and independent variables in a regression analysis was nonlinear, goodness-of-fit tests did not reject linearity unless the nonlinearity was extreme. Thus, the omnibus goodness-of-fit tests that test in many directions at the same time, have little power and will not reject until the lack of fit is severe (Breiman, 2001b). This means that an absolute focus on the use of explanatory models without validating their findings with other methods (e.g., predictive analytics) may result in misleading conclusions, even when those findings pass goodness-of-fit tests and residual checks (Breiman, 2001b). This does not mean that we should avoid using explanatory models in scientific research, but rather, it emphasizes on the importance of the problem and data in selecting the appropriate research methods. As we noted before, with the availability of new data in recent years, we should not expect to see theory development or testing and certain analysis methods in every academic business paper. Instead, we need to see if there is a match between the research question and the approach used to answer that question, even if the approach lacks sophistication. From this perspective, "the requirement for a contribution to theory would be replaced with the requirement that any journal paper has a high potential for stimulating research that will [have an] impact on business theory and/or practice" (Avison & Malaurent, 2014).

After reviewing the necessity and role of predictive analytics in scientific research, we now turn our attention to some of the more common methods used in predictive analytics.[1]

[1] Here, we focus on methods and techniques that are mainly used in analytics studies and refrain from describing such well-known methods as linear regression that are also common in explanatory studies.

4.2.1. Decision tree

Decision trees can be used for regression and classification purposes. As a classification algorithm, each non-leaf node of a decision tree indicates a test on an attribute of the input cases; each branch corresponds to an outcome of the test; and each leaf node indicates a class prediction. Classification accuracy and size of a decision tree are used to determine its quality (Lee, 2010). Decision trees recursively separate observations into branches to construct a tree for improving prediction accuracy. In doing so, they use mathematical algorithms to identify a variable and a corresponding threshold for that variable that splits the input observation into two or more subgroups. This process is repeated at each leaf node until the complete tree is constructed. The splitting algorithm seeks to find a variable-threshold pair that maximizes the homogeneity (order) of the resulting subgroups of samples. The most commonly used mathematical algorithms for splitting include entropy-based information gain, Gini index, and Chi-squared test.

Examples for the use of decision trees in prior studies include developing theoretical profiles of the decision rationale applied to IS profile prioritization (Karhade, Shaw, & Subramanyam, 2015) and employing decision trees to build an automated data-driven methodology for addressing self-selection in observational impact studies in management research (Yahav, Shmueli, & Mani, 2015).

4.2.2. Artificial neural networks

Artificial neural networks (ANNs) are defined as "massively parallel processors, which tend to preserve experimental knowledge and enable their further use" (Hájek, 2011). Modeled after the processes of learning in the cognitive system and the neurological functions of the brain, ANNs are capable of modeling very complex non-linear functions and predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a so-called process of learning from existing data.

Standard ANNs comprise many connected neurons that act as processors to produce a sequence of real-valued activations. Input neurons get activated through sensors that perceive the environment, and other neurons get activated through weighted connections from previously active neurons. Learning in such networks involves obtaining weights that make the ANN exhibit desired behavior, which may require long causal chains of computational stages depending on the problem and how the network is structured (Schmidhuber, 2014). While each of these stages more often than not employs a non-linear function to transform the aggregate activation of the network, few such stages have traditionally been arranged in ANNs (Schmidhuber, 2014). A new subset of machine learning methods that primarily leverage ANNs (Trask, 2016) run through many stages to transform the representation at one level (starting with the raw input) into a representation at higher, more abstract levels (LeCun, Bengio, & Hinton, 2015). These methods, which are called deep neural networks (DNNs), are capable of learning extremely complex functions (LeCun et al., 2015) and have attracted wide-spread attraction in the past decade by winning many official international pattern recognition competitions (Schmidhuber, 2014).

Some of the novel applications of DNNs in business research include predicting the next event in a business process (Evermann, Rehse, &

Fettke, 2017); financial decision support (Kraus & Feuerriegel, 2017), and insurance fraud (Y. Wang & Xu, 2018). In the first paper, the authors used data on past process instances to make predictions about the current ones. Predicting the next event in business processes can be of extreme value to firms as they would be able to provide better customer support when confronted with inquiries about the remaining time until an issue is resolved; they can predict the completion time of a production process for better planning and higher utilization; and they would be able to identify likely compliance violations to mitigate business risk (Evermann et al., 2017). In the second example, the authors use DNNs to mine a large corpus of financial disclosures to predict how stock prices move in response to the release of such documents (Kraus & Feuerriegel, 2017). And in the third example, the authors go beyond the traditional variables used in detecting automobile insurance fraud (for instance, time of the claim and brand of the insured car) to build a deep learning model that uses text analytics to also incorporate the textual information in the claims (Y. Wang & Xu, 2018).

4.2.3. Partial least squares (PLS) regression

PLS regression is a method that predicts Y (output variable) based on X (input variables) and explains their common structure. It generalizes and integrates features from multiple regression and principal component analysis. PLS regression extracts a set of components from both X and Y, such that these components explain as much as possible of the covariance between X and Y. This method is specifically useful when the number of predictors is large. It surpasses the standard regression when multicollinearity exists among the predictors (Abdi, 2003).

Previous research has used PLS for predicting bankruptcy (Serrano-Cinca & Gutiérrez-Nieto, 2013), segmentation and behavioral characterization of auction bidders (Mancha, Leung, Clark, & Sun, 2014), and for mining churning behavior and developing retention strategies (Lee, Lee, Cho, Im, & Kim, 2011).

4.2.4. Least angle regression (LARS)

LARS is a regression algorithm for high-dimensional data that informs the variable selection process by generating estimations of variables to include, along with their coefficients. In that sense, it relates to the classical model selection method known as "forward stepwise regression". The LARS algorithm, however, works differently from the forward selection method. The procedure starts by setting all predictor coefficients equal to zero. Next, it identifies the predictor mostly correlated with the target variable (P_1). This predictor is then used to build the regression model until some other predictor (P_2) has as much correlation with the current residual. At this point, LARS departs from the forward selection method and proceeds in a new direction that is equiangular between P_1 and P_2. The algorithm continues in a similar way to consider all the predictor variables. At each step, a new predictor (P_i) will find its way into the "most correlated" set. Following the inclusion of P_i, LARS proceeds equiangularly between P_1, P_2, …, P_i, i.e., along the "least angle direction," until all predictors are considered (Efron, Hastie, Johnstone, & Tibshirani, 2004).

Examples for the application of LARS in business research include its use for near infrared spectra analysis to determine the internal quality of navel oranges (Liu, Yang, & Deng, 2015) and applying LARS to select a sparse and relatively stable set of indicators for predicting stock return (Z. Wang & Tan, 2009).

4.2.5. Random forest

A random forest grows multiple decision trees. To classify a new observation from an input vector, the observation is sent as input to each of the trees in the forest. Each tree specifies a classification, or "votes" for that class. At the end, the forest chooses the classification that has obtained the most votes among all trees (Breiman, 2001a). A random forest is grown in three steps:

1. A random sample (with replacement) is drawn from the original data. The sample size is equal to the number of cases in the training set.
2. Assuming there are *M* input variables, a number *m*, which is relatively much smaller than *M* and whose value is held constant during the growth of the forest, is specified. Then at each node, *m* variables are randomly selected out of the *M* input variables. The best split on these *m* selected variables is used to split the node.
3. Nodes are split using the selected variables to grow the trees to the largest extent possible, without any pruning.

Random forests are computationally efficient and robust to noise (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011). Examples of previous studies that have used random forests include improving the predictive ability of tax avoidance models by constructing a network of firms connected through shared board membership (Lismont et al., 2018), developing a decision support system for predicting diabetic retinopathy (Piri, Delen, Liu, & Zolbanin, 2017), and identifying freshmen's profiles likely to face major difficulties to complete their first academic year (Hoffait & Schyns, 2017).

4.2.6. Gradient boosting trees

Gradient boosting is a technique that generates a prediction model in the form of an ensemble of weak prediction models.[2] However, it is different from common ensemble techniques, such as random forests, that simply use the average of models in the ensemble. The family of boosting methods employs a different, constructive strategy of ensemble formation (Natekin & Knoll, 2013). It constructs additive regression models by successively fitting a simple parametrized function (base learner) to current "pseudo" residuals by least squares at each iteration (Friedman, 2002). In other terms, a new weak, base learner model is trained at each iteration with respect to the error of the whole ensemble learned so far. The learning procedure continues by sequentially fitting new models to improve the accuracy with which the target variable is estimated (Natekin & Knoll, 2013).

Many researchers embracing the analytics paradigm have used gradient boosting models to build decision support systems. Some of the novel applications of these models include loss cost modeling and prediction for automobile insurance (Guelman, 2012), creating an ensemble of boosting trees and other data mining methods to predict the number of software faults in a typical software development project (Rathore & Kumar, 2017), and combining gradient boosting trees with other binary classification models to predict early user attrition in computer gaming (Milošević, Živić, & Andjelković, 2017).

4.2.7. Support vector machine (SVM)

SVM nonlinearly maps input vectors (variables) to a very high-dimension feature space in which a linear decision surface is constructed to classify the objects into one of two categories (Cortes & Vapnik, 1995). Therefore, given labeled training data, the algorithm creates an optimal hyperplane that can be used to categorize new observations. While there are numerous linear hyperplanes that can separate the two classes of the response variable, the optimal hyperplane is the one that lies in the middle of the fattest bar (i.e., the margin) between the two classes (Provost & Fawcett, 2013).

Examples for the application of SVM in academic research are (Piri, Delen, & Liu, 2018) and (Khan, Schmidt-Thieme, & Nanopoulos, 2017). In the first paper, the authors develop a synthetic minority oversampling technique that uses SVM to identify which minority cases are better (i.e., more informative) candidates for oversampling. The second article builds an SVM-based collaborative classification method in scale-free peer-to-peer networks that not only improves local classification accuracy, but also keeps the communication cost low throughout the network.

[2] Gradient Boosting. (*n.d.*) *In Wikipedia*. Retrieved July 3, 2017, from http://en.wikipedia.org/wiki/Gradient_boosting

While this list of analytical methods is in no ways comprehensive, we believe it provides a useful introduction to some of the commonly used approaches to address business problems through academic research. Interested readers can refer to (Putka et al., 2017) or other cited references for detailed discussions on various analytics techniques.

5. Discussion and conclusion

Availability of data in large volumes and varieties, and the advances in analytics methods, especially in those employing data mining and machine learning techniques, furnish an unprecedented opportunity for scholars to address numerous, interesting questions (Tonidandel et al., 2016). While many of these modern-day business problems cannot be effectively answered using the traditional research methods, there still is a shadow of doubt in some disciplines over the use of analytics methods for addressing such problems. Especially, founding the arguments on an established theory, developing hypotheses guided by theory, and collecting data to test those hypotheses appear to be the sine qua non of an authentic research in some business disciplines. Such sheer reliance on theories in the traditional paradigm not only has prevented investigating emerging phenomena, but also has led to "questionable scientific conclusions." (Breiman, 2001b).

In this manuscript, we provided brief introductions on various types of analytic methods and listed some of the more common methods researchers have used to address a variety of business questions. Our focus was on the complementarity relationship between the traditional and analytics paradigms, rather than advocating for absolute supremacy of one over the other. By referring to the extant literature, we summarized how the emergent analytics paradigm can supplement scientific inquiry. These aspects include: generating new theories, developing measures, comparing competing theories, improving existing models, assessing relevance, and assessing predictability. In addition, we pointed out to some of the known issues in the dominant research paradigm, such as HARKing, that can potentially be avoided if both approaches are accepted in business journals.

We hope this summary helps in embracing analytic efforts in top business journals and paves the way for generating not only rigorous, but also relevant research in these disciplines.

Appendix A. Definitions of [Business] analytics

Definition of [Business] analytics	Driving forces	Reference
"The general process of exploration and analysis of data to discover new and meaningful patterns in data."	Business problems	(Kohavi et al., 2002)
"Collection, storage, analysis, and interpretation of data in order to make better decisions and improve organizational performance."	Data, the need for improved decisions	(Davenport, 2006)
"The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions."	Data, the need for improved decisions	(Davenport & Harris, 2007)
"A group of tools that are used in combination with one another to gain information, analyze that information, and predict outcomes of the problem solutions."	Data integration, data mining	(Bose, 2009)
"A process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving."	Data, Process, Software, People	(Liberatore & Luo, 2010)
A set of tools and techniques that enable organizations to "know what is happening now, what is likely to happen next and what actions should be taken to get the optimal result."	Environmental complexity, data deluge, tough competition	(LaValle, Hopkins, Lesser, Shockley, & Kruschwitz, 2010)
"Interpreting organizational data to improve decision-making and to optimize business processes."	Organizational data	(Shanks, Sharma, Seddon, & Reynolds, 2010)
"Structured data analytics includes three categories of increasing complexity and impact: descriptive, predictive, and prescriptive."	Structured organizational data	(Lustig et al., 2010)
"The scientific process of transforming data into insight for making better decisions."	Robust data	(INFORMS, 2012)
Using such methods as data mining to "constantly drive meaningful information from plethora of data and make critical business decisions."	Good quality and integrated data	(Nemati & Udiavar, 2012)
"… the process that transforms raw data into valuable information about capabilities, market positions, activities, and goals that organizations could pursue in order to stay competitive."	Business data residing in data warehouses	(Anand, Sharma, & Kohli, 2013)
"The process of developing actionable decisions or recommendation for actions based upon insights generated from historical data."	Historical data	(Sharda, Asamoah, & Ponna, 2013)
"Use of data to make sounder, more evidence-based business decisions."	Data	(Tamm, Seddon, & Shanks, 2013)
"Analytics facilitates realization of business objectives through reporting of data to analyze trends, creating predictive models to foresee future problems and opportunities and analyzing/optimizing business processes to enhance organizational performance."	Data	(Delen & Demirkan, 2013)
"… BA systems are marked by their increasing focus on pattern recognition and prediction, rather than historical reporting."	Big data, availability of powerful techniques	(Gillon, Aral, Lin, Mithas, & Zozulia, 2014)
"Evidence-based problem recognition and solving that happen within the context of business situations."	Evidence (Data)	(Holsapple et al., 2014)
"The use of data ("big" or "small") and data storage, retrieval, and analysis tools for gaining efficient and effective insights for decision making."	Data	(Turel & Kapoor, 2016)
"BA is the practice and art of bringing quantitative data to bear on decision making."	Quantitative data	(Shmueli et al., 2016)

References

Abdi, H. (2003). *Partial least squares (PLS) regression. Encyclopedia of Social Sciences, Research Methods. 2003*. Thousand Oaks (CA): Sage.
Anand, A., Sharma, R., & Kohli, R. (2013). *Routines, reconfiguration and the contribution of business analytics to organisational performance. Australasian Conference on Information Systems*Australia: RMIT University1–10.
Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete|WIRED. *Wired*. Retrieved from https://www.wired.com/2008/06/pb-theory/.
Anderson, C. (2015). *Creating a data-driven organization: Practical advice from the trenches*. O'Reilly Media, Inc.
Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods, 12*, 623–652.
Avison, D., & Malaurent, J. (2014). Is theory king?: questioning the theory fetish in information systems. *Journal of Information Technology, 29*(4), 327–336. http://dx.doi.org/10.1057/jit.2014.8.
Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data Analytics: Hyped Up Aspirations or True Potential? *Vikalpa, 38*(4), 1–12. http://dx.doi.org/10.1177/0256090920130401.
Berk, R. A. (2006). An introduction to ensemble methods for data analysis. *Sociological Methods & Research, 34*(3), 263–295. http://dx.doi.org/10.1177/0049124105283119.
Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: a comparative study. *Decision Support Systems, 50*(3), 602–613.
Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems, 109*(2), 155–172. http://dx.doi.org/10.1108/02635570910930073.
Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799. http://dx.doi.org/10.1080/01621459.1976.10480949.
Breiman, L. (2001a). Random forests. *Machine Learning, 45*(1), 5–32.
Breiman, L. (2001b). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199–231.
Chae, B. K., Yang, C., Olson, D., & Sheu, C. (2014). The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. *Decision Support Systems, 59*, 119–126.
Chau, M., & Chen, H. (2008). A machine learning approach to web page filtering using content and structure analysis. *Decision Support Systems, 44*(2), 482–494.
Chen, Y.-L., Chen, J.-M., & Tung, C.-W. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. *Decision Support Systems, 42*(3), 1503–1520.
Combes, C., & Azema, J. (2013). Clustering using principal component analysis applied to autonomy–disability of elderly people. *Decision Support Systems, 55*(2), 578–586.
Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine Learning, 20*(3), 273–297.
Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review, 84*(1), 98.
Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
Davenport, T. H., & Kim, J. (2013). *Keeping up with the quants: Your guide to understanding and using analytics*. Harvard Business Review Press.
Delen, D. (2015). *Real-world data mining: Applied business analytics and decision making*. Upper Saddle River, New Jersey: FT Press (a Pearson company).
Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems, 55*, 359–363.
Dhar, V., Geva, T., Oestreicher-Singer, G., & Sundararajan, A. (2014). Prediction in economic networks. *Information Systems Research, 25*(2), 264–284.
Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*(2), 407–499.
Evermann, J., Rehse, J.-R., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems, 100*, 129–140. http://dx.doi.org/10.1016/j.dss.2017.04.003.
Flath, C., Nicolay, D., Conte, T., Dinther, C., & Filipova-Neumann, L. (2012). Cluster analysis of smart metering data. *Business information systems engineering. 4. Business information systems engineering* (pp. 31–39). In.
Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal. 38*(4), 367–378.
Gartner (2016). IT glossary: Descriptive analytics. Retrieved from http://www.gartner.com/it-glossary/descriptive-analytics.
Gillon, K., Aral, S., Lin, C.-Y., Mithas, S., & Zozulia, M. (2014). Business analytics: Radical shift or incremental change? *Communications of the Association for Information Systems, 34*(1).
Guelman, L. (2012). Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications, 39*(3), 3659–3667. http://dx.doi.org/10.1016/j.eswa.2011.09.058.
Hájek, P. (2011). Municipal credit rating modelling by neural networks. *Decision Support Systems, 51*(1), 108–118.
Hoffait, A.-S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems, 101*, 1–11. http://dx.doi.org/10.1016/j.dss.2017.05.003.
Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems, 64*, 130–141.
Huang, S.-L., & Tsai, Y.-H. (2011). Designing a cross-language comparison-shopping agent. *Decision Support Systems, 50*(2), 428–438.
Hwang, S.-Y., & Tang, J. (2004). Consulting past exceptions to facilitate workflow exception handling. *Decision Support Systems, 37*(1), 49–69.
INFORMS (2012). Operations research & analytics. Retrieved from https://www.informs.org/Explore/Operations-Research-Analytics.
John, B. M., Goh, D. H. L., Chua, A. Y. K., & Wickramasinghe, N. (2016). Graph-based cluster analysis to identify similar questions: A design science approach. *Journal of the Association for Information Systems, 17*(9), 590.
Joseph, D., Fong Boh, W., Ang, S., & Slaughter, S. A. (2012). The career paths less (or more) traveled: A sequence analysis of IT career histories, mobility patterns, and career success. *MIS Quarterly, 36*(2).
Karhade, P., Shaw, M. J., & Subramanyam, R. (2015). Patterns in information systems portfolio prioritization: Evidence from decision tree induction 1. *MIS Quarterly, 39*(2), 413–433.
Kaza, S., Hu, P. J.-H., Hu, H.-F., & Chen, H. (2011). Designing, implementing, and evaluating information systems for law enforcement-A long-term design-science research project. *CAIS, 29*, 28.
Kaza, S., Wang, Y., & Chen, H. (2007). Enhancing border security: Mutual information analysis to identify suspect vehicles. *Decision Support Systems, 43*(1), 199–210.
Kerr, N. L., Adamopoulos, J., Fuller, S., Greenwald, T., Kiesler, S., Laughlin, P., ... Wit, A. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. Retrieved from https://pdfs.semanticscholar.org/27e0/30d584791528f97a842eaca16f4c59078a1c.pdf.
Khan, U., Schmidt-Thieme, L., & Nanopoulos, A. (2017). Collaborative SVM classification in scale-free peer-to-peer networks. *Expert Systems with Applications, 69*, 74–86. http://dx.doi.org/10.1016/j.eswa.2016.10.008.
Kiang, M. Y., Fisher, D. M., Chen, J.-C. V., Fisher, S. A., & Chi, R. T. (2009). The application of SOM as a decision support tool to identify AACSB peer schools. *Decision Support Systems, 47*(1), 51–59.
Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. *Communications of the ACM, 45*(8), 45–48.
Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems, 104*, 38–48. http://dx.doi.org/10.1016/j.dss.2017.10.001.
LaValle, S., Hopkins, M. S., Lesser, E., Shockley, R., & Kruschwitz, N. (2010). Analytics: The new path to value. *MIT Sloan Management Review, 52*(1), 1–25.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. http://dx.doi.org/10.1038/nature14539.
Lee, S. (2010). *Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls. 49*, 486–497. http://dx.doi.org/10.1016/j.dss.2010.06.002.
Lee, H., Lee, Y., Cho, H., Im, K., & Kim, Y. S. (2011). Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decision Support Systems, 52*(1), 207–216. http://dx.doi.org/10.1016/j.dss.2011.07.005.
Liberatore, M. J., & Luo, W. (2010). The analytics movement: Implications for operations research. *Interfaces, 40*(4), 313–324.
Lismont, J., Cardinaels, E., Bruynseels, L., De Groote, S., Baesens, B., Lemahieu, W., & Vanthienen, J. (2018). Predicting tax avoidance by means of social network analytics. *Decision Support Systems.*. http://dx.doi.org/10.1016/j.dss.2018.02.001.
Liu, C., Yang, S. X., & Deng, L. (2015). A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of navel oranges. *Expert Systems with Applications, 42*(22), 8497–8503. http://dx.doi.org/10.1016/j.eswa.2015.07.005.
Locke, E. A. (2007). The case for inductive theory building†. *Journal of Management, 33*(6), 867–890. http://dx.doi.org/10.1177/0149206307307636.
Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). The analytics journey. *Analytics Magazine, 3*(6), 11–13.
Mabroukeh, N. R., & Ezeife, C. I. (2010). A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys (CSUR), 43*(1), 3.
Mancha, R., Leung, M. T., Clark, J., & Sun, M. (2014). Finite mixture partial least squares for segmentation and behavioral characterization of auction bidders. *Decision Support Systems, 57*, 200–211. http://dx.doi.org/10.1016/j.dss.2013.09.003.
McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68.
Miaskiewicz, T., & Monarchi, D. E. (2008). A review of the literature on the empathy construct using cluster analysis. *Communications of the Association for Information Systems, 22*(1), 7.
Milošević, M., Živić, N., & Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games. *Expert Systems with Applications, 83*, 326–332. http://dx.doi.org/10.1016/j.eswa.2017.04.056.
Mortenson, M. J., Doherty, N. F., & Robinson, S. (2015). Operational research from Taylorism to terabytes: A research agenda for the analytic sage. *European Journal of Operational Research, 241*(3), 583–595. http://dx.doi.org/10.1016/j.ejor.2014.08.029.
Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics, 7*.
Nemati, H., & Udiavar, A. (2012). *Organizational readiness for implementation of supply chain analytics*.
Norvig, P. (2009). All we want are the facts, ma'am. Retrieved December 23, 2017, from http://norvig.com/fact-check.html.
Pai, H.-T, Wu, F., & Hsueh, P.-Y. S.(Sabrina) (2014). A relative patterns discovery for enhancing outlier detection in categorical data. *Decision Support Systems, 67*, 90–99. http://dx.doi.org/10.1016/j.dss.2014.08.006.
Piatetsky-Shapiro, G. (2007). Data mining and knowledge discovery 1996 to 2005:

overcoming the hype and moving from "university" to "business" and "analytics". *Data Mining and Knowledge Discovery, 15*(1), 99–105. http://dx.doi.org/10.1007/s10618-006-0058-2.

Piri, S., Delen, D., & Liu, T. (2018). A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets. *Decision Support Systems, 106*, 15–29. http://dx.doi.org/10.1016/j.dss.2017.11.006.

Piri, S., Delen, D., Liu, T., & Zolbanin, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decision Support Systems, 101*, 12–27. http://dx.doi.org/10.1016/J.DSS.2017.05.012.

Prinzie, A., & Van den Poel, D. (2006). Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM. *Decision Support Systems, 42*(2), 508–526. http://dx.doi.org/10.1016/J.DSS.2005.02.004.

Prinzie, A., & Van den Poel, D. (2007). Predicting home-appliance acquisition sequences: Markov/Markov for Discrimination and survival analysis for modeling sequential information in NPTB models. *Decision Support Systems, 44*(1), 28–45. http://dx.doi.org/10.1016/J.DSS.2007.02.008.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.* O'Reilly Media, Inc.

Putka, D. J., & Oswald, F. (2015). Implications of the Big Data movement for the advancement I-O science and practice. In S. Tonidandel, E. King, & J. Cortina (Eds.). *Big data at work: The data science revolution and organizational psychology* (pp. 181–212). New York, NY: Taylor & Francis.

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2017). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods*. http://dx.doi.org/10.1177/1094428117697041.

Rathore, S. S., & Kumar, S. (2017). Towards an ensemble based system for predicting the number of software faults. *Expert Systems with Applications, 82*, 357–382. http://dx.doi.org/10.1016/j.eswa.2017.04.014.

Rouse, M. (2012). Prescriptive analytics. Retrieved from http://searchcio.techtarget.com/definition/Prescriptive-analytics.

Rouse, M. (2015). Descriptive analytics. Retrieved from http://whatis.techtarget.com/definition/descriptive-analytics.

Schmidhuber, J. (2014). Deep learning in neural networks: An overview. Retrieved from http://www.idsia.ch.

Schroeder, J., Xu, J., Chen, H., & Chau, M. (2007). Automated criminal link analysis based on domain knowledge. *Journal of the Association for Information Science and Technology, 58*(6), 842–855.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2013). Partial Least Square discriminant analysis for bankruptcy prediction. *Decision Support Systems, 54*(3), 1245–1255. http://dx.doi.org/10.1016/j.dss.2012.11.015.

Shanks, G., Sharma, R., Seddon, P., & Reynolds, P. (2010). *The impact of strategy and maturity on business analytics and firm performance: A review and research agenda*.

Sharda, R., Asamoah, D. A., & Ponna, N. (2013). *Business analytics: Research and teaching perspectives*.

Sharda, R., Delen, D., & Turban, E. (2017). *Business intelligence analytics, and data science: A managerial perspective.* Upper Saddle River, New Jersey: Pearson.

Shmueli, G., & Koppius, O. R. (2011). *Predictive analytics in information systems research.* MIS Quarterly: Management Information Systems Research Center, University of Minnesotahttp://dx.doi.org/10.2307/23042796.

Shmueli, G., Patel, N. R., & Bruce, P. C. (2016). *Data mining for business analytics: concepts, techniques, and applications with XLMiner.* John Wiley & Sons.

Tamm, T., Seddon, P., & Shanks, G. (2013). *Pathways to value from business analytics*.

Thomassey, S., & Fiordaliso, A. (2006). A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems, 42*(1), 408–421.

Tonidandel, S., King, E. B., & Cortina, J. M. (2016). Big data methods. *Organizational Research Methods, 109442811667729*. http://dx.doi.org/10.1177/1094428116677299.

Trask, A. W. (2016). *Grokking deep learning.* Manning Publications.

Turel, O., & Kapoor, B. (2016). A business analytics maturity perspective on the gap between business schools and presumed industry needs. *Communications of the Association for Information Systems, 39*(1), 6.

van Dam, J.-W., & van de Velden, M. (2015). Online profiling and clustering of Facebook users. *Decision Support Systems, 70*, 60–72. http://dx.doi.org/10.1016/J.DSS.2014.12.001.

Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems, 105*, 87–95. http://dx.doi.org/10.1016/j.dss.2017.11.001.

Wang, Z., & Tan, S. (2009). Identifying idiosyncratic stock return indicators from large financial factor set via least angle regression. *Expert Systems with Applications, 36*(4), 8350–8355. http://dx.doi.org/10.1016/j.eswa.2008.10.018.

Xu, J., & Chen, H. (2005a). Criminal network analysis and visualization. *Communications of the ACM, 48*(6), 100–107.

Xu, J., Hu, D., & Chen, H. (2009). The dynamics of terrorist networks: Understanding the survival mechanisms of Global Salafi Jihad. *Journal of Homeland Security and Emergency Management, 6*(1).

Xu, J. J., & Chen, H. (2004). Fighting organized crimes: Using shortest-path algorithms to identify associations in criminal networks. *Decision Support Systems, 38*(3), 473–487.

Xu, J. J., & Chen, H. (2005b). CrimeNet explorer: A framework for criminal network knowledge discovery. *ACM Transactions on Information Systems (TOIS), 23*(2), 201–226.

Yahav, I., Shmueli, G., & Mani, D. (2015). A tree-based approach for addressing self-selection in impact studies with big data. *MIS Quarterly, 40*(4), 819–848.

Dursun Delen is the holder of William S. Spears and Neal Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). He received his Ph.D. in Industrial Engineering and Management from OSU in 1997. Prior to his appointment as an Assistant Professor at OSU in 2001, he worked for a privately-owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for five years, during which he led a number of decision support, information systems and advanced analytics related research projects funded by federal agencies, including DoD, NASA, NIST and DOE. His research has appeared in major journals including Decision Support Systems, Communications of the ACM, Computers and Operations Research, Computers in Industry, Journal of Production Operations Management, Artificial Intelligence in Medicine, Expert Systems with Applications, among others. He recently published seven books/textbooks in the broader are of Business Analytics. He is often invited to national and international conferences for keynote addresses on topics related to Healthcare Analytics, Data/Text Mining, Business Intelligence, Decision Support Systems, Business Analytics and Knowledge Management. He regularly serves and chairs tracks and mini-tracks at various information systems and analytics conferences, and serves on several academic journals as editor-in-chief, senior editor, associate editor and editorial board member. His research and teaching interests are in data and text mining, decision support systems, knowledge management, business intelligence and enterprise modeling.

Hamed M. Zolbanin is an Assistant Professor and the Director of the Business Analytics program at Ball State University. He earned his doctorate in Management Science and Information Systems from Oklahoma State University. He has presented his research in several national and international conferences. He has published a healthcare analytics research article in decision support systems journal, and has several articles under review at reputable journals in the area of analytics and information systems. His research interests include healthcare analytics, medical informatics, business intelligence, decision support systems, machine learning, predictive modeling and business analytics.