



BIROn - Birkbeck Institutional Research Online

Hindle, G. and Kunc, M. and Mortenson, M. and Oztekin, A. and Vidgen, Richard (2019) Business analytics: defining the field and identifying a research agenda. [Editorial/Introduction]

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/29840/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Journal Pre-proof

Business Analytics: Defining the field and identifying a research agenda

Giles Hindle , Martin Kunc , Michael Mortensen , Asil Oztekin ,
Richard Vidgen

PII: S0377-2217(19)30817-3
DOI: <https://doi.org/10.1016/j.ejor.2019.10.001>
Reference: EOR 16087



To appear in: *European Journal of Operational Research*

Please cite this article as: Giles Hindle , Martin Kunc , Michael Mortensen , Asil Oztekin ,
Richard Vidgen , Business Analytics: Defining the field and identifying a research agenda, *European
Journal of Operational Research* (2019), doi: <https://doi.org/10.1016/j.ejor.2019.10.001>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Elsevier B.V. All rights reserved.

Business Analytics: Defining the field and identifying a research agenda

Authors (in alphabetical order)

Giles Hindle

Hull University Business School, UK

Giles.Hindle@hull.ac.uk

Martin Kunc*

University of Southampton, UK

M.H.Kunc@soton.ac.uk

Michael Mortensen

University of Warwick, UK

M.Mortenson@warwick.ac.uk

Asil Oztekin

University of Massachusetts Lowell, USA

Asil_Oztekin@uml.edu

Richard Vidgen

UNSW Business School, University of New South Wales, Australia

and

School of Business, Economics and Informatics, Birkbeck University, UK

r.vidgen@unsw.edu.au

Abstract

The Special Issue on business analytics has been a great endeavor with more than 100 papers received. The call for papers highlighted that business analytics has a clear role to generate competitive advantage in organizations and our focus has been to demonstrate this role through the papers finally selected for the special issue. The editorial aims to provide not only a summary of the papers but also presents our perspective on the current situation of the field through a computational literature review and comparison with the papers in the special issue. Our findings, and discussions on the papers included in the Special Issue, suggest that business analytics is maturing as a field with significant synergies and opportunities for the operational research community.

Keywords: Business Analytics; Practice of OR; Data Science; Topic Models;
Computational Literature Review

1 Introduction

Business analytics is an evolving phenomenon that reflects the increasing significance of data in terms of its growing volume, variety and velocity (Mortenson et al., 2015). There has been growing interest in analytics and data science, as commercial organizations use their growing repositories of data to create value in their businesses, and governments and communities seek to create value of a broader nature (Davenport, 2013, Hindle and Vidgen, 2018). Analytic methods are being used in many and varied ways, for example to predict consumer behavior, to predict the likelihood of medical conditions, to analyze social media, to better manage traffic networks, and so on.

A number of researchers have argued the growing attention and prominence afforded by analytics presents both a challenge and an opportunity for organizations and for the Operational Research (OR) community (Liberatore and Luo, 2010, Ranyard et al. 2015, Mortenson et al. 2015, Vidgen et al. 2017). Many in the community have recognized this growth and sought to align themselves with the ‘analytics’ movement. For instance, OR societies around the world now offer analytics related conferences, certification and publications. However, the volume of analytics-orientated studies published in journals directly associated with OR is still comparatively low (Mortenson et al. 2015).

The current view of analytics is encapsulated by Davenport and Harris’ (2007) succinct and widely adopted definition: “By *analytics* we mean the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.” (p. 7, emphasis in the original). The key aspect of this definition is that analytics ultimately provides findings that are relevant to organizations – it’s not just a scientific exercise. Thus, one of the main concerns in business analytics is related to research into the transformation needed for organizations to become data-driven and evidence-based.

Business analytics can be viewed as the intersection of a variety of disciplines, of which OR, machine learning, and information systems are of particular relevance (Figure 1). As a process it can be characterized by descriptive, predictive, and prescriptive model building using heterogeneous and ‘big’ data sources. These models enable organizations to make quicker, better, and more intelligent decisions in order to create business value in the broadest sense - potentially the difference between survival and extinction in an increasingly competitive world. Thus, business analytics is about the context in which OR and data science are deployed.

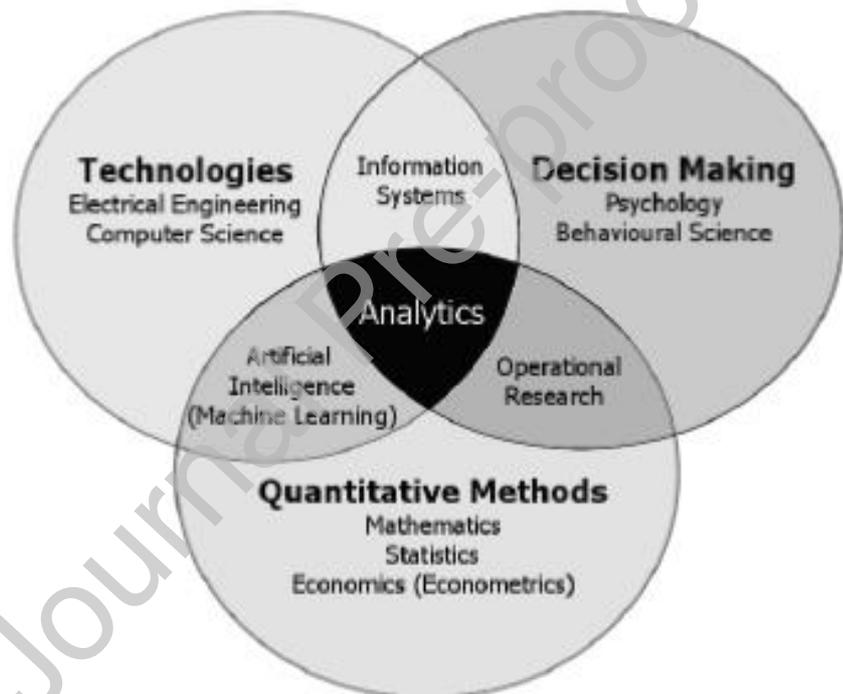


Figure 1: Business analytics (Figure 1, Mortenson et al. 2015)

2 Opportunities for OR arising from Business Analytics

OR grew out of the application of scientific style analysis to military and then industrial operational systems in the 1950s. During the 1960s OR found a home within industry and also in universities. Corbett and Van Wassenhove (1993) argue a “natural drift” happened as the old-style practice-oriented OR (management engineering) remained underdeveloped relative to the academic theory-oriented OR (management

science) and pure practice-oriented OR (management consulting). For many years OR went through a polarization towards its purest forms, leaving management engineering neglected. Pidd and Dunning-Lewis (2001) found a large proportion of published research in OR tended to be disengaged from practice and focused on decision mathematics or optimization, apparently for its own theoretical sake. Moreover, Ormerod (2002) argued OR had reached its maturity in terms of a management-agenda-setting role in the early 1970s and was subsequently decreasingly perceived as important or strategic.

However, the landscape for OR has changed in recent years with the emergence of business analytics and data science. The development of data technology, the internet and the emergence of companies like eBay, Amazon, and Google, that employ algorithms which optimize processes in real time, have created a huge interest and trend in the use of data. The adoption of a range of new technologies has created three main challenges for organizations: high volume of data generated both internally and externally, high velocity in terms of data generation and decision making, and high variety of data, e.g. images, text, video, etc. In this new era, analytics really can drive companies both strategically and operationally since automated data collection, analysis and decision-making can be used to deal with business complexity faster and more reliably than management intuition (Kunc, 2018).

The role of OR is therefore being increasingly challenged by these new developments. Traditional tasks such as crunching data using Excel and generating ad hoc optimizations or statistical evidence are being disrupted by a widespread interest in analytics. Much as OR became engaged in the use of analytical methods to re-engineer organizational processes, the demands from today's organizations have led to the re-engineering of analytical processes. A critical question is, therefore, how does the OR

community need to change? The intention of our special issue is to contribute to this discussion.

3 Aims of the special issue

The objective of this special issue is to publish papers that contribute to both the theory and practice of business analytics. While papers describing new techniques have been published in other journals, we wanted to see how the techniques could be applied in practice and discuss their implications for creating value in organizations. The relationship with data science is direct but distinct: business analytics is an organizational activity that draws on and uses the techniques of data science and OR as appropriate.

Research into analytics should seek to both incorporate the unique aspects of the OR discipline, as well as the innovations, concerns and characteristics of the analytics movement. While not being an exhaustive list of topics, this special issue invited OR scholars and practitioners to look at:

- Ethics and governance issues in business analytics: How should data be obtained? What are the ethical implications of using applications of business analytics to influence behavior? These questions are particularly relevant following the implementation of the General Data Protection Regulation (GDPR) in Europe.
- Big data and business analytics: What are the limitations and applications of optimization and other OR techniques to large datasets? What are the challenges for applications of OR methods within distributed systems? What is the possibility that OR models could in fact be the producers of big data, e.g., large-scale simulation models?

- Business analytics modelling methods: What new methods/models in response to big data, e.g., sentiment mining, can be adopted by OR? What methods can be used for data cleaning and interrogation, hypothesis testing and model validation in large datasets? From the review of the literature, many methods in business analytics, e.g., machine learning, have not been widely recognized or adopted by OR practitioners and scholars.
- Organizational issues in business analytics adoption: What are the issues facing organizations trying to adopt business analytics? What is the role of real-time applications of OR in organizations? This is a critical point of differentiation between analytics and OR.
- Data quality and business analytics: How can semi-structured data be used effectively in traditional OR models? What is the role of multi-methodology in business analytics? What opportunities do open data present for the OR discipline? Many models in OR do not have real data and, further, there is no intention to verify the practicability of collecting the data automatically needed by these models.
- Business analytics and decision support: How should the delivery of OR projects adapt to the evolving needs of decision makers in modern organizations? How might OR applications best capitalize on the developments and the potential of real-time analytics? What role do problem structuring and “soft” OR techniques play in analytics and big data projects? Data visualization, as a method to support decisions in the area of analytics, is an area that has developed intensively in many simulation software environments over the years but structuring the problems being addressed by the application of analytics is still in its infancy.

4 Computational literature review of Business Analytics

To further analyze the papers published in this special issue, and also the business analytics field in general, we perform a computational literature review (CLR) of the field (Mortenson & Vidgen, 2016). A CLR provides an automated analysis of a set of research outputs by identifying themes through underlying (latent) topics that are discovered through topic modelling. The CLR includes a wide range of metrics to analyze academic research and is based upon latent Dirichlet allocation (Blei, Ng, & Jordan, 2003), which has “become the ‘de facto’ standard for topic modelling” (Mortenson & Vidgen, 2016). Topic modelling is performed when a predefined number of topics, which are groupings of words based on their co-occurrences representing specific subject matters, are populated algorithmically (Mortenson & Vidgen, 2016). The approach is in the same family of algorithms as principal component analysis (PCA) - indeed it is sometimes referred to as non-linear principal component analysis because it finds latent components (called topics) which can explain variance in the data, which are representations of the subjects discussed across a corpus of text data (Mortenson & Vidgen, 2016).

4.1 Data collection

The data used to perform the CLR is gathered from the academic database Scopus (<https://www.scopus.com/>), using a query for all years available that match the keyword phrases “business analytics” or “data science”, to either title, abstract or keywords of the datastore. In total, 4,597 articles were downloaded: “business analytics” produced 997 results, “data science” 3,632 results, and 32 papers were associated with both searches. The growth in the number of articles in the period 2000 through 2018 can be seen clearly

in Figure 2. The fact that there are relatively few papers that match both queries (just 0.7%) suggests that these two fields are perhaps more distinct than might be imagined.

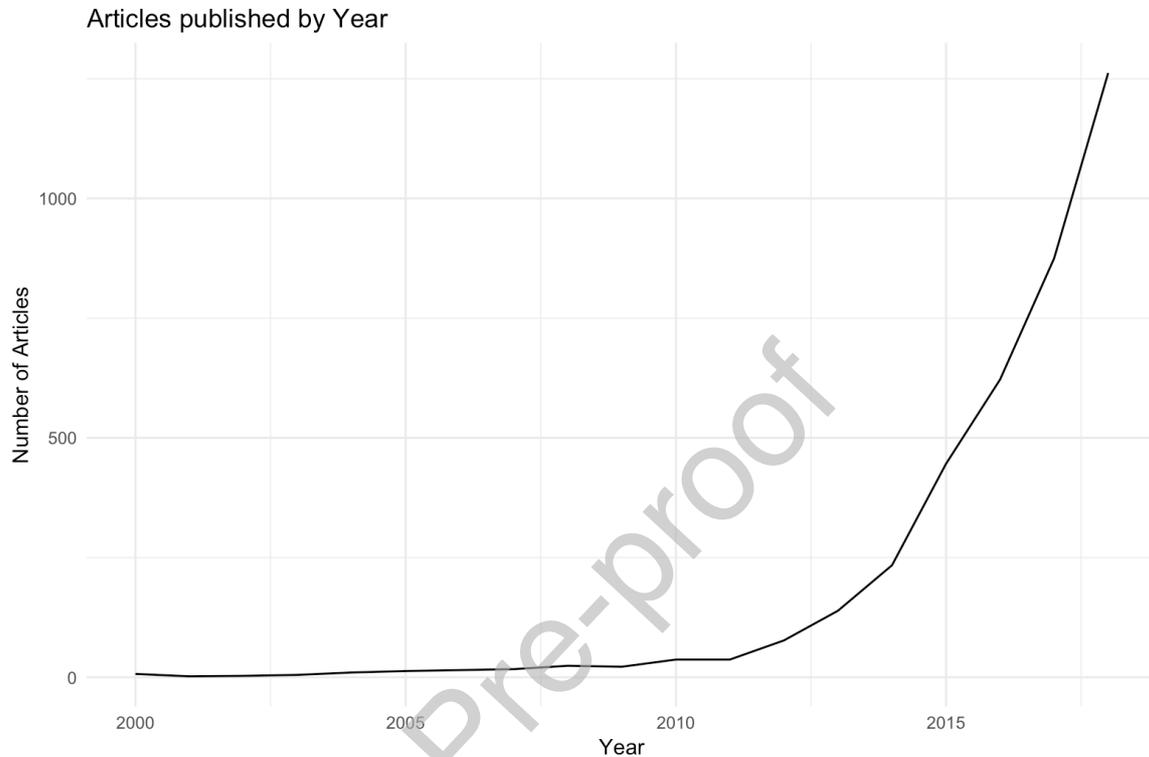


Figure 2: Articles published per annum (2000 – 2018)

To prepare the data for the CLR analysis, many of the standard steps used in natural language processing (NLP) were performed such as:

1. the removal of stopwords (short function words such as “I”, “of” and “and”);
2. the removal of punctuation and non-alphabetical characters;
3. the stemming of words (reducing each word to its shortest stem such that “statistics”, “statistical” and “stats” are all reduced to their shortest stem, “stat”).

This means that all words that share a stem are treated as the same, which in the above case (and most others) is appropriate as the words have highly similar meanings.

With these steps completed, the data was analyzed using the CLR library. The CLR is a package developed in the R programming language and can be downloaded from GitHub (<https://github.com/rvidgen/clar>).

4.2 Impact analysis

The CLR, in addition to its text analytics capabilities, provides access to a range of common metrics for analyzing the impact of papers. For instance, firstly we produce a list of the top 10 papers in the field, as measured by citation count (Table 1).

Perhaps unsurprisingly, given the greater volume of publications associated with “data science” in comparison to “business analytics”, the top 10 articles comprise entirely papers associated with the former search term. However, an analysis of the average citation count for each term has “business analytics” at 6.68 cites per paper and “data science” at 4.69.

The H-index (Hirsch, 2005) is commonly used to assess a researcher’s impact in a standardized form. For example, a researcher with an index of 10 has published 10 papers each of which has been cited at least 10 times. The H-index can also be applied to publication sources, such as journals and conference proceedings. To gain a sense of the impact of the various publication sources we calculate the H-index for each of the sources in the dataset (Table 2).

EJOR comes in at number 3 on the basis of H-index. In total citation count EJOR ranks 19th and in terms of citations per paper ranks 170th. This is largely due to EJOR only being active in publishing papers with the relevant keywords since 2014; given more time, the articles published in EJOR will likely gather more citations and the citations per article rise accordingly. It is also worth noting that papers associated with “business analytics” in EJOR compare favorably to the overall trend in respect to “cites

per paper”, with the local score of 13.8 from Table 2 compared to an average of 3.4 for the journal as whole.

4.3 Topics in Business Analytics

The second part of the analysis is to identify the topics present in the papers. As with most statistical models of its type, Latent Dirichlet Analysis (LDA) requires the researcher to specify the number of topics (K) *a priori*. Following K -finding diagnostics and inspection we fitted a topic model with $K=36$. The topics were labelled by three of the editors independently and then reviewed and labels agreed (Table 3). The labelling was supported by a report generated in R markup that shows the wordcloud for each topic, a list of the ten most likely words for each topic, and a list of the titles of the ten articles most likely to load on each topic (this report is available as an online appendix). A sample wordcloud is shown in Figure 3, for a topic that we labelled “smart cities and geospatial”.

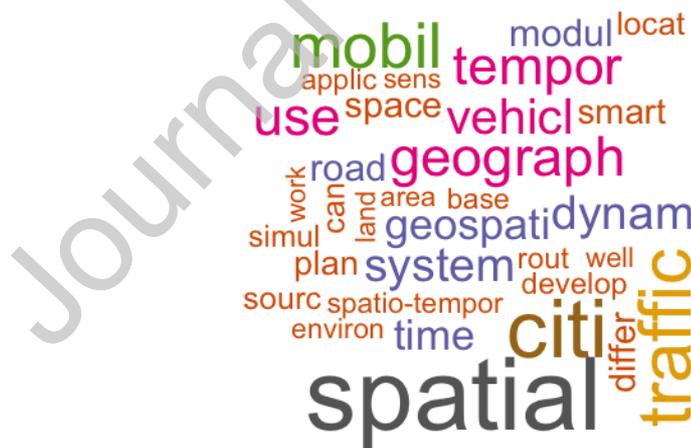


Figure 3: Word cloud of topic 2 (labelled “smart cities and geospatial”)

5 The Special Issue papers

We received 103 papers for the special issue, of which 14 were accepted. These 14 papers have also been fed into the topic model built from the business analytics and data science literature in order to estimate their topic proportions. Figure 4 shows the average proportions of the top 10 topics across these 14 papers.

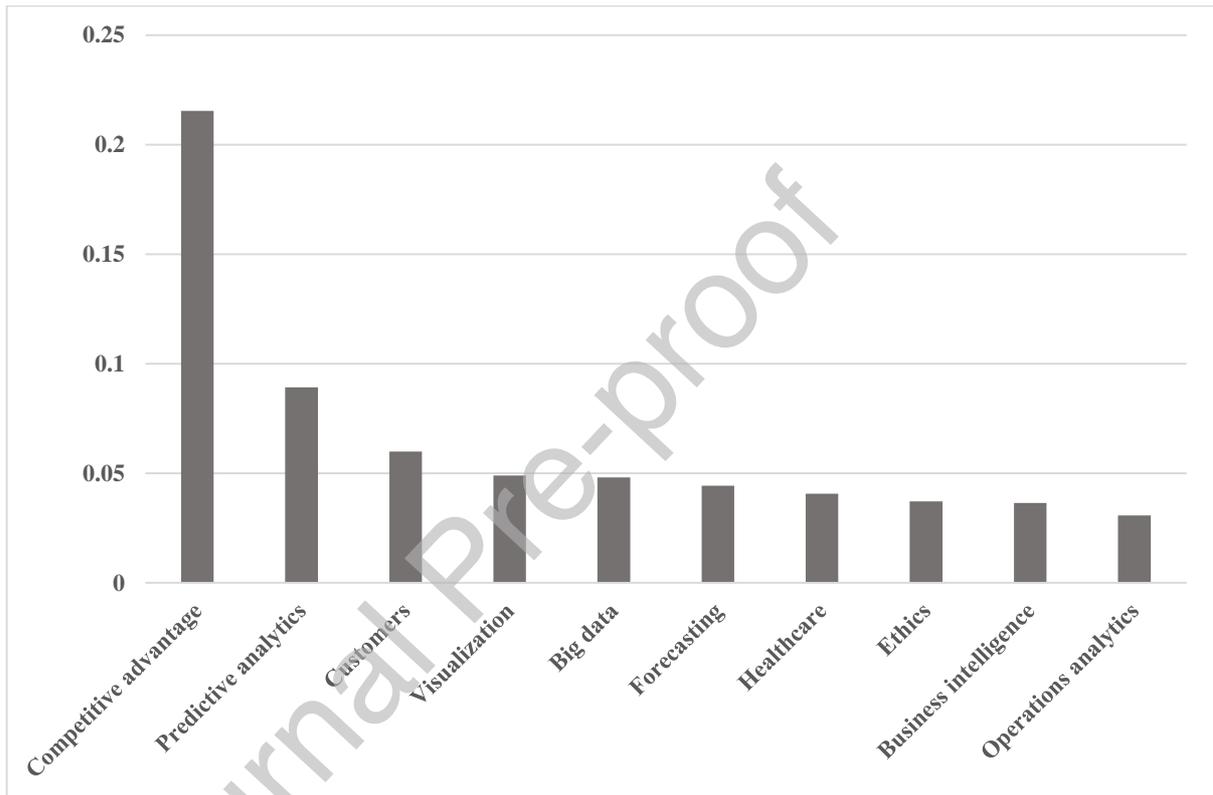


Figure 4: Average topic proportions for the top 10 topics in the papers of the special issue

It is particularly noteworthy the overall focus on “competitive advantage” as by far the most frequent topic in the special issue papers. This is not completely unexpected given the content of the call for papers and further reflects the increasing importance of business analytics driving companies both strategically and operationally.

To further categorize the papers into themes we use the framework proposed by Vidgen et al. (2019). Figure 5 shows that data can come from many sources and while this data may be classified as ‘big’, it is not a requirement that it be so. Indeed, many

organizations can create significant business value from relatively small volumes of data, which may have not been exploited previously, to give insight into the organization's customers, processes and competitive environment. The data must be captured, stored and managed, and its quality assured. Analytics methods can then be applied to the data in order to support better decision-making through descriptive, predictive, and prescriptive applications of analytics, ultimately leading to the creation of business value. All of this analytics activity takes place in some organizational and wider environmental context that is typified by cultural, social, ethical, political and economic dimensions.

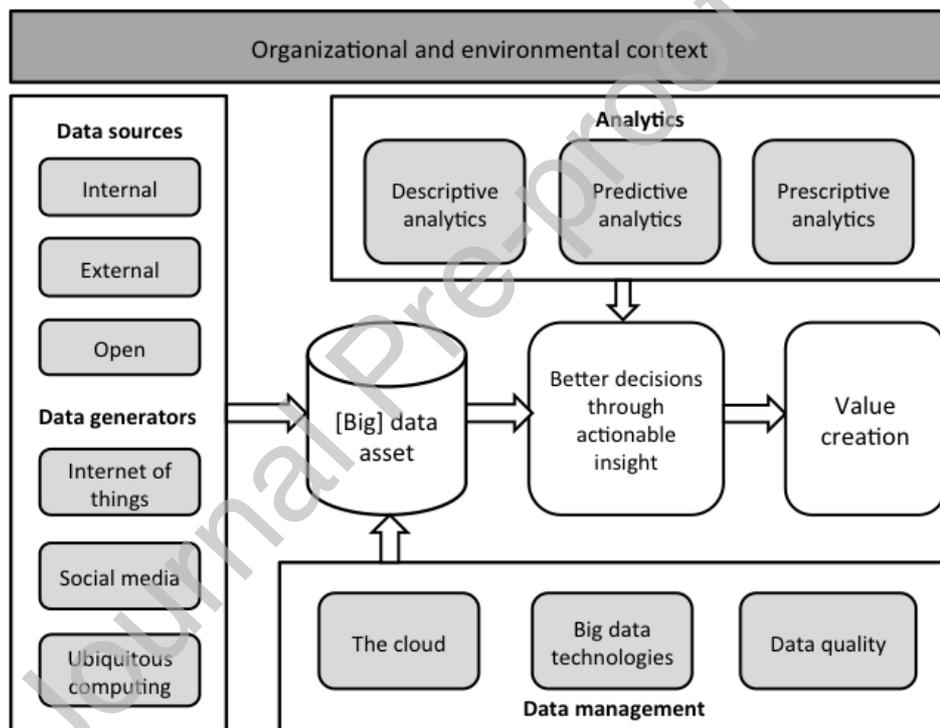


Figure 5: Business analytics in context (Vidgen et al. 2019)

Using the framework in Figure 5 we can see where the articles included in this special issue fall, and, using our topic model we can see which of the topics in Table 3 are most pertinent for each of the articles:

Data sources and generators. Li et al. address an important element of new technologies, sustained use over time, through the analysis of the adoption of Augmented Reality technologies. The paper's main two topics, as predicted in our topic model, are

“competitive advantage” (15%) and “customers” (13%). Augmented reality can become a critical complement to business analytics technologies to provide realistic contextualization to decision making. The authors use expectation confirmation theory as their theoretical framework. Then, they use text-analytics (tweets) to generate insights about user sentiments that is employed to evaluate the benefits/risk associated with the technology. Huang et al. focus on predicting hospital readmission using machine learning on an atypical healthcare dataset: doctors’ notes, from 16 regional hospitals in the US. Perhaps unsurprisingly, the topic model assigns “predictive analytics” and “healthcare” as the top two topics (24% and 17% respectively). One of the key challenges for the research reported in the paper is to overcome the lack of adequate data in healthcare records, which hampers the use of methods to predict hospital events. The authors used natural language processing techniques, Latent Topic Ensemble Learning, with a library of medical vocabularies to tease out information from semi-structured healthcare records. With this approach, their model can use doctors’ notes as a primary data source. Their study also evaluates the robustness of the predictive model using costs of misclassification instead of area under curve (AUC). Ni et al. provide an example of the integration of non-traditional data sources, such as wearable sensors, with analytics to support individual decision making. Again, they are categorized against the same two topics, “healthcare” (15%) and “predictive analytics” (11%). They focus on personalized advice towards healthy lifestyles using lifelogging data and a two-stage hybrid model to predict human physical activity status. Zhou et al. use a dataset from Amazon to develop a demand prediction model of remanufactured products. This type of product has a high degree of uncertainty in terms of demand and customers’ returns. The authors compare three machine learning models (CART, M5 and RF) with linear regression and neural networks. They found that social interaction, brand equity, seller reputation, product popularity are strong predictors of remanufactured products.

Data Management. Zhan et al. address one of the biggest problems in business analytics (creating seamless data platforms) in one of the critical functions for companies (supply chain). Supply chains can generate rich big data from multiple sources: RFID, GPS locations, Point of Sale, social media, sensor data, etc., but the data is in ‘information silos’. Their abstract was predicted as 23% on “competitive advantage” and 17% on “big data”. The authors employ competence set analysis as their method for structuring the data into useful information for strategic decision-making. They further present a case study to show the practical aspects of their proposition.

Analytics. Delen et al. develop a predictive model to evaluate student attrition. This area of application has received increasing attention in the field recently due to its impact on education institutions, e.g. attrition is a key indicator used in college evaluations/rankings as well as funding from the government. The model matches 28% of the abstract to the “predictive analytics” topic, and 12% to “education”. The authors use Bayesian Belief Network to generate probabilities for first-year student attrition using a dataset containing 10 years of information. They find that imbalanced datasets are better for their model in terms of specificity and accuracy and the feature selection process reduces the number of predictor variables by half to achieve similar results to a full model. Simpler models help decision makers to focus only on important factors. Martinez et al. offer an application of machine learning to predict customer purchasing behavior in manufacturing contexts when customers are not subject to contracts, a domain in which there is high uncertainty. As this description may suggest, the top two topics found are “customers” (29%) and “predictive analytics” (20%). The application can have critical impact on allocation of resources for a manufacturing company. One of the salient aspects of this paper is the use of a richer set of customer characteristics with a shorter time frame. The authors compare logistic lasso regression, single-hidden layer feedforward neural network (SLFN) and gradient tree boosting using AUC as a

performance measure with a dataset taken from more than 10,000 different customers. Jiang et al. offer an application of predictive analytics in appointment scheduling, which is one of the largest areas of concern for the healthcare industry due to its impact on resources and uncertainty on patients' behavior. The most frequent topics are "forecasting" (23%) and "healthcare" (15%). They perform descriptive analytics of a large dataset related with MRI services covering more than 3.7 million data records from 74 hospitals. Then, they model the problem as a dynamic scheduling process using multi-class and multi-priority patients and test their application with a subset of data and two scheduling policies to reduce waiting time. Cui et al. address an important area of concern for manufacturers and retailers, especially online: customers' product return volume. Again, the highest topic proportions are assigned to "predictive analytics" (26%) and "customers" (20%). They work with a manufacturing company that has a large variety of products and a rich data set related with each item. They test four high-dimensional machine learning methods: LASSO, LARS-OLS hybrid, SCAR and Elastic Net as well as two tree-based machine learning methods: random forest and gradient boosting. They find that the LASSO model has the most robust performance in their case.. Kraus et al. offer an evaluation of the use of deep learning in business analytics through a review of models and applications. Their top two topics are "competitive advantage" (24%) and "machine learning" (19%). The authors claim deep learning has much better performance than machine learning algorithms and argue that it will become the industry standard for predictive analytics.

Better decisions and value creation. As noted in Figure 4, "competitive advantage" represented the most frequent topic in the SI. Tim et al. present a case study evaluating one of the critical aspects of business analytics: how to translate analytics use into organizational value. Unsurprisingly, "competitive advantage" is the most frequent topic in the abstract (47%), with visualization second (18%). Using the case study of a mobile

game company, they apply a technology affordances perspective, which focuses not only on the technological aspects but also on the actors' perceptions and use of the technology, to uncover the process of implementing an analytics-driven transformation. This study offers an in-depth view on the components for implementing business analytics in order to create value. Critically, it offers a four-stage guidance for practitioners to generate value from business analytics. Conboy et al. offer a perspective from the practice of Operational Research into the world of business analytics through eight case studies. They use the concept of dynamic capabilities, which are the activities responsible for organizational development in terms of sensing opportunities and threats, seizing those opportunities and transforming resources, to evaluate the case studies. The most frequent topic presented in the abstract is "competitive advantage" (67%). Their main contribution is the identification of 14 foundations of dynamic capabilities supported by OR techniques, e.g., scheduling, forecasting, etc., that are adapted to business analytics practices. Duan et al. take the brave step of demonstrating how business analytics contributes to innovation success. The abstract presented is classed as 80% on the topic of "competitive advantage". They use absorptive capacity theory to explain how business analytics helps companies using data that was captured through a questionnaire with responses from 218 businesses. They find that business analytics impacts on the capacity to scan the environment to identify trends for new products but this capacity has to be enabled by a data-driven organizational culture. Another interesting insight is the positive correlation between business analytics and competitive advantage in the perception of the respondents of the survey. Chen et al. offer a procedure for segmenting customers based on their data usage/demand, which is a critical aspect for the development of mobile applications. Their main topics are predicted as "sustainability" (13%) and "classification and clustering" (10%). Their approach is based on a modified expectation maximization (EM) algorithm for the Multivariate Gaussian

Mixture (MGM) model. The proposed framework starts with describing mobile data usage and clustering using the MGM model to develop a predictive model that is transformed into a prescriptive model (based on machine learning and simulation) for developing strategic options. Vidgen et al. discuss the ethical aspects of implementing business analytics. They propose a business ethics canvas based on five ethical principles (utility, rights, justice, common good and virtue) to evaluate the ethical issues around the use of business analytics. The main topics predicted are “ethics” (29%) and “visualization” (21%). More importantly, they argue that ethics should not be separated from value creation, as happened with the Cambridge Analytica case.

6 Summary

A key driver of this special issue is to explore how OR can embrace business analytics. In managing the special issue, and through the CLR analysis of the literature, we suggest that OR practitioners and scholars will need to demonstrate to managers that they know how to design and implement algorithm-based solutions to complex problems that are driven by data that is characterized by volume, variety and velocity. A further finding of significance for scholars and practitioners is that they must be willing to build and test multiple models and then compare them to identify their strengths and limitations. The approach of tackling every problem with one’s favorite modelling tool is being replaced by a ‘horses for courses’ approach in which the OR practitioner will need to be comfortable with a ‘bricolage’ approach to analytics development.

While we find that business analytics has multiple specialisms, all of them show a clear pattern: analytics is a multidisciplinary endeavor with a focus on implementation. The articles in this special issue can be differentiated from data science due to an emphasis on organizational value and impact, pluralistic approaches, and interdisciplinary methods. However, we also see the inclusion of many of the ‘new’ methods of machine learning

being incorporated in this special issue, and, certainly, more so than would likely be included in a 'standard' OR journal issue.

We argue that the field of business analytics is gaining maturity and is building and strengthening its identity in the fast-moving world of data technologies and data analyses. Further, there is a clear focus on the application of these methods to real business problems alongside the 'traditional' approaches that may be more familiar to OR researchers and practitioners. In short, we believe this special issue presents a showcase for the OR community of where business analytics is today, and where it may go in the future. This vision of business analytics is of an applied science that combines the latest developments in modern computing (e.g., visualization and deep learning), with the latest analytical methods and approaches (including predictive analytics and natural language processing), but in the context of real organizational problems and the complexities they bring (requiring a problem structuring capability). Although the technologies and algorithms may change, this is a space in which operational research should find a natural home. It is an opportunity that we hope that the OR community will wholeheartedly embrace.

References:

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Corbett, C. J., & Van Wassenhove, L. N. (1993). The natural drift: What happened to operations research?. *Operations Research*, 41(4), 625-640.
- Davenport, T.H. and Harris, J.G., (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- Davenport T.H., (2013). Analytics 3.0. *Harvard Business Review*, 91(12), 64-72.
- Department for Business Innovation and Skills, (2013). *Seizing the data opportunity: A strategy for UK data capability*.
- Hindle, G., and Vidgen, R., (2018). Developing a business analytics methodology: a case study in the food bank sector. *European Journal of Operational Research*, 268(3): 836-851.

- Hirsch, J. E., (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Kunc, M., (2018). *Strategic Analytics: Integrating Management Science with Strategic Management*. Wiley & Sons.
- Liberatore, M.J. and Luo, W., (2010). The analytics movement: Implications for operations research. *Interfaces*, 40(4), pp.313-324.
- Mortenson, M.J., Doherty, N.F. and Robinson, S., (2015). Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3), pp.583-595.
- Mortenson, M., and Vidgen, R., (2016). A computational literature review of the technology acceptance model. *International Journal of Information Management*. 36: 1248 – 1259.
- Ormerod, R. J., (2002). On the nature of OR: Taking stock. *Journal of the Operational Research Society*, 53(5), 475-491.
- Pidd, M., & Dunning-Lewis, P., (2001). Innovative research in OR/MS? *European Journal of Operational Research*, 128(1), 1-13.
- Ranyard, J.C., Fildes, R. and Hu, T.I., (2015). Reassessing the scope of OR practice: the influences of problem structuring methods and the analytics movement. *European Journal of Operational Research*, 245(1), pp.1-13.
- Vidgen, R., Shaw, S. and Grant, D. B., (2017). Management challenges in creating value from business analytics, *European Journal of Operational Research*, Elsevier, vol. 261(2), pages 626-639.
- Vidgen, R., Kirshner, and S., Tan, F., (2019). *Business Analytics: a management approach*. Red Globe Press.

Papers from the Special Issue

- Baechle, C., Huang, C. D., Agarwal, A., Behara, R. S., & Goo, J. (2019). Latent topic ensemble learning for hospital readmission cost optimization. *European Journal of Operational Research*.
- Chen, Y. T., Sun, E. W., & Lin, Y. B. (2019). Merging anomalous data usage in wireless mobile telecommunications: Business analytics with a strategy-focused data-driven approach for sustainability. *European Journal of Operational Research*.
- Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2019). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*.
- Cui, H., Rajagopalan, S., & Ward, A. R. (2019). Predicting Product Return Volume Using Machine Learning Methods. *European Journal of Operational Research*.
- Delen, D., Topuz, K., & Eryarsoy, E. (2019). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research*.
- Duan, Y., Cao, G., & Edwards, J. S. (2018). Understanding the impact of business analytics on innovation. *European Journal of Operational Research*.

- Jiang, Y., Abouee-Mehrizi, H., & Diao, Y. (2018). Data-driven analytics to support scheduling of multi-priority multi-class patients with wait time targets. *European Journal of Operational Research*.
- Kraus, M., Feuerriegel, S., & Oztekin, A. (2018). Deep learning in business analytics and operations research: Models, applications and managerial implications. *arXiv preprint arXiv:1806.10897*.
- Li, H., Gupta, A., Zhang, J., & Flor, N. (2018). Who will use augmented reality? An integrated approach based on text analytics and field survey. *European Journal of Operational Research*.
- Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2018). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*.
- Ni, J., Chen, B., Allinson, N. M., & Ye, X. (2019). A hybrid model for predicting human physical activity status from lifelogging data. *European Journal of Operational Research*.
- Van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2019). Predicting customer demand for remanufactured products: A data-mining approach. *European Journal of Operational Research*.
- Tim, Y., Hallikainen, P., Pan, S. L., & Tamm, T. (2018). Actualizing business analytics for organizational transformation: A case study of Rovio Entertainment. *European Journal of Operational Research*.
- Vidgen, R., Hindle, G., & Randolph, I. (2019). Exploring the ethical implications of business analytics with a business ethics canvas. *European Journal of Operational Research*.
- Zhan, Y., & Tan, K. H. (2018). An analytic infrastructure for harvesting big data to enhance supply chain performance. *European Journal of Operational Research*.

Table 1: Top 10 articles by citation count (n=4,597)

Source title	No. of articles	Citations	Start year	End year	Citations per article	H-index
Decision Support Systems	12	590	2007	2019	49.2	9
Proceedings of the National Academy of Sciences of the United States of America	14	251	2014	2019	17.9	8
European Journal of Operational Research	13	179	2014	2019	13.8	8
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	173	288	2009	2019	1.7	8
Management Science	8	326	2014	2017	40.8	7
Future Generation Computer Systems	17	288	2014	2019	16.9	7
Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	27	208	2010	2018	7.7	7
Procedia Computer Science	37	107	2014	2018	2.9	7
Big Data	19	699	2013	2019	36.8	6
Royal Society Open Science	7	151	2014	2016	21.6	6

Table 2: Top 10 sources of articles (n=4,597)

Authors	Article title	Year	Source title	Citations
Jordan M.I., Mitchell T.M.	Machine learning: Trends, perspectives, and prospects	2015	Science	473
Van der Aalst W.	Process mining: Data science in action	2016	Process Mining: Data Science in Action	337
Stephens Z.D., Lee S.Y., Faghri F., Campbell R.H., Zhai C., Efron M.J., Iyer R., Schatz M.C., Sinha S., Robinson G.E.	Big data: Astronomical or genetical?	2015	PLoS Biology	331
Swan M.	The quantified self: Fundamental disruption in big data science and biological discovery	2013	Big Data	329
Najafabadi M.M., Villanustre F., Khoshgoftaar T.M., Seliya N., Wald R., Muharemagic E.	Deep learning applications and challenges in big data analytics	2015	Journal of Big Data	320
Waller M.A., Fawcett S.E.	Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management	2013	Journal of Business Logistics	318
Provost F., Fawcett T.	Data Science and its Relationship to Big Data and Data-Driven Decision Making	2013	Big Data	260
Hazen B.T., Boone C.A., Ezell J.D., Jones-Farmer L.A.	Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and	2014	International Journal of Production Economics	244

	applications			
McGibbon R.T., Beauchamp K.A., Harrigan M.P., Klein C., Swails J.M., Hernandez C.X., Schwantes C.R., Wang L.-P., Lane T.J., Pande V.S.	MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories	2015	Biophysical Journal	236
Dhar V.	Data science and prediction	2013	Communicati ons of the ACM	230

Journal Pre-proof

Table 3: Topic model analysis (K=36)

T opic no	Topic label	T opic no	Topic label
1	Healthcare	1 9	Social media
2	Smart cities and geospatial	2 0	Process analysis
3	Food and wine	2 1	Statistics and computational statistics
4	Predictive analytics	2 2	Medical trials
5	Model building and scientific method – analytics methodologies	2 3	Materials science
6	Education (teaching)	2 4	Algorithm design and optimization
7	Collaboration and community	2 5	Sustainability
8	Medical diagnosis	2 6	The environment – analytics for public good
9	Genomics	2 7	Forecasting
1 0	Operations analytics	2 8	Ethics
1 1	Service innovation and IS Services	2 9	Education (curriculum design)
1 2	Software development	3 0	ERP and decision-making – business intelligence
1 3	Classification and clustering	3 1	Data warehouses and data ontology
1 4	Customers (e-commerce and e- marketing)	3 2	Competitive advantage
1 5	Machine learning	3 3	Big data
1 6	Literature reviews – decision support and information dissemination	3 4	Visualization
1 7	Graph and network theory	3 5	Trust, security, and privacy
1 8	Big data technologies and architecture	3 6	Medicine and drug development