# Proposal: Researching supporting returns of big data and analytics investment in supply chain management

David A. Wagle

School of Business, Northcentral University TIM7020: Databases & Business Intelligence Dr. Marty Crossland

April 4, 2021

|  |  |  |
| --- | --- | --- |
|  | **Contents** |  |
| [**Background**](#_bookmark0) |  | **3** |
| [**Problem Statement**](#_bookmark3) |  | **4** |
| [**Goal**](#_bookmark4) |  | **6** |
| [**Relevance and Significance**](#_bookmark5) |  | **7** |
| [**Literature Review**](#_bookmark6) |  | **8** |
| [**Approach**](#_bookmark7) |  | **10** |

# Background

Big Data & Analytics (BDA), Blockchain, and Artifical Intelligence (AI) are currently major topics in relationship to supply chain management (SCM). Google Trends shows that while interest in SCM from business and industry world-wide has remained relatively flat (see Figure [1)](#_bookmark1) there has been substantial increases in interest in BDA and Blockchain, which moderate growth in interest for AI (see Figure [2).](#_bookmark2) Similarly, a brief search of the Roadrunner Search Discovery Service on the topic of “Supply Chain Management” filtering on the terms for BDA, AI, and Blockchain yielded 3,213

peer-reviewed articles published in business journals over the year 2020 alone. A similar search for 2016 yielded only a few hundred such articles.

Off those 2020 studies, only address cost effectiveness or financial risk. Yet, there is significant disagreement as to the value these technologies can provide to organizations.

For the purposes of this discussion, “Big Data” will be considered data that suffices the conditions of volume, variety, and velocity (Hofmann, [20](#_bookmark21)17).

# Figure 1

*Google Trend interest in Supply Chain Management*

Volume is the measure of the size of the data. It should be noted that the amount of data being discussed with respect to BDA, Blockchain, and AI involves tens or hundreds of GB of data. These volumes are not present in many BDA environments, but they are present in SCM (Appuswamy et al., [2013).](#_bookmark14) In 2010, supply chains were dealing with

# Figure 2

*Google Trend interest in BDA, AI, & Blockchain*

100GB of data daily (The Economist, [2010).](#_bookmark30) That number is increasing annually with the proliferation of RFID and Internet of Things (IOT) (Valuates Reports, [2020).](#_bookmark33) Variety is the complex nature of disparate data types including text data, image data, audio data, and so forth. Finally, velocity is the speed at which data is generated and the rate at which it must be processed to provide near-real-time information to the organization (Hofmann, [2017).](#_bookmark21)

Roßmann et al. [(2018)](#_bookmark26) demonstrate that business leaders and SCM experts see little value in BDA, while computer science experts feel strongly that there is value to be had.

This disagreement derives from the fact that supply chains are inherently complex, uncertain, and require significant investment in data management systems (Govindan et al., [2017).](#_bookmark18) There is little doubt that BDA, Blockchain, and AI have the potential to enhance operational decision-making (Gunasekaran et al., [2017).](#_bookmark19) There is, however, uncertainty that he required investments necessary to drive such improvements can be

recouped sufficiently (Arunachalam et al., [2018).](#_bookmark15) Similar disagreements exist for blockchain (Kouhizadeh et al., [2021)](#_bookmark23) and AI (NEDSI, [2019)](#_bookmark25)

# Problem Statement

Given the uncertainty expressed by business leadership as to the value of these technologies in practice, and the gap between the perceived value as expressed by CS

experts compared to business leaders, a clear problem for researchers emerges. What are the characteristic factors that must be present for a business to successfully adopt BDA, Blockchain, and AI into their SCM systems in order to see a sufficient return on investment?

This question is important and timely as SCM expenses are increasing at

double-digit rates yet the impact of characteristics of the organization on SCM adoption and responsiveness are unclear (Story et al., [2021).](#_bookmark28) This lack of a sound research-based answer is especially difficult on small-to-mid-sized enterprises (SME) where these technologies are known to be underutilized. Willetts et al. [(2020)](#_bookmark35) identify that this lack of utilization is likely driven by a host of factors, and they are able to identify several:

* **Data Issues:** Data lacking sufficient quality, data silos, data complexity.
* **Knowledge Issues:** Skills and awareness insufficient to engender confidence of success.
* **Regulatory Issues:** Various legal issues generally around privacy and security.
* **Technical Issues:** Lacking sufficient infrastructure to succeed.
* **Organizational Issues:** Cultural and leadership impacting adoption.
* **Resource Issues:** Financial constraints limiting adoption.

It is this proposal’s intention to limit the scope of inquiry to Organizational Issues. This is for several reasons. Importantly, the first three items all amount to a single issue of awareness. There are methods, tools, and techniques to clean and prepare even the most complex data environments for data analytics use (Anne Laurent et al., [2020;](#_bookmark13) Souibgui

et al., [2019).](#_bookmark27) Similarly, meeting complex legal, governance, compliance, and security requirements is a problem with available solutions (Hager et al., [2020;](#_bookmark20) Kennedy A. Torkura et al., [2017).](#_bookmark22) Skills issues can be addressed through training, consulting, communities of practice and other methods. Finally, while financial resources are a legitimate problem when talking about expensive SCM solutions, one presumption of this proposal is that some SME’s who otherwise have the financial resources to embark on an SCM BD&A

strategy are non-the-less not doing so.

Thus the problem statement is to provide research that outlines organizational characteristics that mediate success when SME’s adopt BD&A, Blockchain, and AI capabilities into their SCM IT strategy. Specifically what identifiable characteristics should practitioners be able to recognize within their own organizations which will be positive predictors of likely success, and to what degree are these predictors accurate.

# Goal

Ideally, the problem statement could be answered with a definitive set of questions which if answered in the affirmative would provide a practitioner a guaranteed probability of success. However, this is highly unlikely as a plausible outcome for this type of research. The predictive power of social science research, such as identifying organizational characteristics, has been long known to be limited (Young, [1932).](#_bookmark36) This has not changed greatly over recent times (Taagepera, [2008).](#_bookmark29) Moreover, the typical method of acquiring information about organizational characteristics is through methods such as structured interviews and surveys. Such interactions have a notoriously low response rate from senior leadership, meaning the resulting data is often fairly small with marginal effect sizes. Such self-selecting surveys also suffer from other potential statistical defects (Zikmund, [199](#_bookmark37)7).

In order to overcome these issues, this proposal has the specific goal of being able to “survey” thousands of SMEs leadership characteristics without their being overtly aware that they are being studied, and without them having the ability to self-select out of the study. To accomplish this ethically, it is of course necessary to rely entirely on publicly available data. Thus the central goal of this study is to utilize Big Data & Analytics methods of textual analysis to identify SMEs that are engaging in a SCM BD&A, Blockchain, or AI transformation by automatically analyzing SEC filing and press releases for key phrases. Then, using similar textual analytical tools, the system will use publicly facing data from social media sites such as Twitter, Facebook, Linked-In, Glassdoor and

others to identify employees and analyze two key dimensions: their perception of key leadership characteristics, and their level of engagement. This study will track the success of the SME implementation effort again through press releases, SEC filings, and other public statements. Finally, financial performance and market share of the companies will be evaluated for the year period prior to the start of the SCM journey and for two years after the initial launch.

The companies collected will be broken into three sets randomly, a training set, a validation set, and a research set. The training set will be used to train the model. The validation set will be used to determine the expected performance of the model, and the research set will be used to make actual predictions and test them against observed outcomes.

Success for this study will be measured in three ways. First, the development of a predictive data mining model that can determine the likelihood of success with confidence interval superior to random guessing. Second, a set of categorized organizational characteristics which allow a description of SMEs who are ideally suited for SCM BD&A success. Third, the ability to built a descriptive matrix of organizational characteristics, employee perceptions of leadership characteristics, and employee engagement that mediate SMC BD&A strategy success.

# Relevance and Significance

This problem exists as for practitioners in SMEs. BD&A, Blockchain, and AI are being utilized by larger firms to gain competitive advantage at a rapid rate (Alazab et al., [2021;](#_bookmark11) Willetts et al., [2020).](#_bookmark35) Further, SME practitioners are unclear as to how to ensure that they have the best chance to extract value from their large IT investments (Walsh

et al., [2010).](#_bookmark34)

This impact is likely significant as SCM first-mover advantage in these technologies appears to offer some durable advantages over the near term (Chen et al., [2015).](#_bookmark17) Given the

financial advantages larger firms have over SMEs when it comes to IT investment, delayed, smaller investments in these technologies may lead to the erosion of market position for the SME firms, all other factors being equal.

This study will be significant in the scope of the study and in that it will use methods to provide empirical evidence of various leadership and firm characteristics related to success without disclosing the fact of the study to leadership. This will prevent any placebo or observational effects impacting the study. Further, if successful, the resulting textual analysis model would allow firms to perform their own analysis to identify areas for self-improvement prior to undertaking such large-scale efforts to improve their chance of success. This will give leaders and practitioners more insight into organizational and leadership gaps that can be relatively easily addressed.

# Literature Review

While BD&A, Blockchain, and AI are among the fastest areas of computer growth, facilitated and driven by cloud computing /parencitegartnerTrendsImpactingCloud2020, with large-scale integrators investing billions to support cloud-based services for clients (Accenture, [2020;](#_bookmark8) Talia, [2013),](#_bookmark31) adoption lags for SMEs. These SMEs face a serious strategic challenge in determining when and how to develop BD&A capabilities. This challenge is primarily around balancing the extensive up-front costs and risks with the perceived long-term benefits (Ajimoko, [2018).](#_bookmark9)

It is known that data-driven companies are more profitable then their competitors (Alsghaier et al., [2017).](#_bookmark12) But researchers have noted that there is little understanding as to how BD&A and related investments help some data-driven companies become successful but not others (Moreno et al., [2020).](#_bookmark24) The crux of the issue is that while research has also demonstrated that becoming data-driven is necessary for long-term competitive advantage (Alsghaier et al., [2017)](#_bookmark12) the mechanism for translating BD&A capabilities into corporate success has not been an adequate focus of research (Trieu, [2017).](#_bookmark32) And in some business

domains, the business leaders even disagree with CS experts that there is much more value to be gained by increased technology investment. Roßmann et al. [(2018),](#_bookmark26) used a delphi research methodology to gather a large, multi-domain, heterogeneous collection of 73 experts from both academia and business. They investigated the realm of SCM and how BD&A and other associated technologies would impact the supply chain in the future. The experts had some interesting points of agreement, but the more interesting results is that business SCM experts simply don’t see the new technologies as being a reliable producer of value for the businesses. The basis of this disagreement in the SCM space rests on the fact that supply chains are already extremely lean and some amount of laxity in the supply chain is always necessary to be able to provide resilience in the face of any number of unforseeable events.

Value creation has been recognized as coming from firm characteristics that above and beyond IT investment (Božič & Dimovski, [2019)](#_bookmark16). Thus, SMEs are faced with a

“no-win” situation wherein they are being asked to expend capital, embrace technological uncertainty, and take on associated risk without clearly proven demarcations of success. One are where non-IT investment is known to matter to the success of BD&A and AI investments is around the various dimensions of data quality. Souibgui et al. [(2019)](#_bookmark27) outlined 15 dimensions of data quality and show how each impacts BD&A and AI viability utilizing 4 different “extract, transform, load” tools. This research importantly demonstrates that investment per-se isn’t always the limiting factor to success as much as matching the investment to the characteristics of the problem that must be solved.

Further, research such as that done by Božič and Dimovski [(2019)](#_bookmark16) show that firm characteristics such as absorptive capacity, the ability to innovate both radically and incrementally at the same time, is a key characteristic linked to successful BD&A adoption. But studies such as this are both relatively new and relatively rare in the research literature. Further, this study is limited to firms in Slovenia, and there is ample reason to question if there might not be culture confounding variables that would make the results

less portable across cultural boundaries.

# Approach

In order to explore the characteristics of the firm that predicate BD&A adoption success, the research will utilize BD&A techniques from machine learning and natural language (ML/NL) processing. It is well understood that these techniques can be used to predict the success or failure of human activities based on how the activities are discussed. For example, Alaphat and Jiang [(2020)](#_bookmark10) showed that using ML/NL processing, it is possible to predict how successful research will be from the grant application.

Thei two specific hypotheses to be tested is the following: H1: The future financial and market success of a company engaging in a BD&A, Blockchain, or AI project is predictable based on existing characteristics of a firm; and, H2: The characteristics of a firm identifiable in H1 are knowable through ML/NL processing of social media posts about the firm.

The proposed research consists of several phases:

1. Identify key firm attributes to be examined
2. Configure an ML/NL engine to categorize for those terms
3. Build a web-bot to identify publicly traded SMEs who are initiating a BD&A, AI, or Blockchain initiative.
4. Track for mentions of the SME on social media sites and feed the ML/NL engine from those feeds
5. Track the financial performance of the SME in terms of key financial market indicators and market share for one year prior and 2 years after the press release
6. Track news feeds for mention of the BD&A, AI, or Blockchain project
7. Using the ML/NL engine, determine if performance is predictable and assign weight to each characteristics

For step (1) business leaders, change management experts, and academics will be surveyed to determine which characteristics should be used to initially seed the model. It should be noted that ML/NL tools can be allowed to discover and add their own categories and this model will be run in that mode. On step (2) the cloud-based IBM Watson™ engine will be configured to explore the input space. Step (3) will be a simple

web-scrapping utility that will be coded to examine typical business news feeds such as Yahoo! Finance, Bloomberg, and others. Once step (3) identifies a company that is announcing a new BD&A, Blockchain, or AI initiative, the name of the company, its stock symbol, and the names of all of its corporate officers and board members will be entered into a second web-scrapping tool that will collect all mentions from Twitter, Glassdoor, Linked-In, and Facebook from the past year. No identifying information of the posters will be retained, only the post content. That content will be fed into the ML/NL engine to compute characteristics of the company as seen by employees, customers, and other stakeholders. At the same time, the ticker symbol of the SME will be sent into a data collection tool to collect financial and market data for the prior year.

All companies identified in a 3 month period will be tracked for a period of two years. At the end of the 2 year period, their financial and market performance will again be collected. Collected data will be divided into three sets, a training set, a validation set, and the experimental set. This division will be done randomly.

The ML/NL model will be trained using the training set. Once trained, the model will be validated against the validation set, which should produce results similar to those predicted by the training set. This will be an iterative process until the model validates. Once validated, the experimental set will be fed to the model and the characteristics and weights will be collected.

Standard statistical methods will be applied to the result to determine how much of

the future stock and market performance can be attributable to the perceived characteristics. Step (6) is there as it will be assumed the project is being employed by the company unless there is specific communication to the public or shareholders that the project has been scrapped.

This experiment will require approximately 2,000 hours of experienced ML/NL and Big Data programmers in equal measure. Once the system is coded, it will be able to run automatically until the data collection step is complete, at which point normal statistical analysis efforts will be required.

# References

Accenture. (2020, September 17). *Accenture cloud first launches with $3 billion investment to accelerate clients’ move to cloud and digital transformation*. Retrieved February 27, 2021, from /news/accenture-cloud-first-launches-with-3-billion-investment-to- accelerate-clients-move-to-cloud-and-digital-transformation.htm

Ajimoko, O. J. (2018). Considerations for the adoption of cloud-based big data analytics in small business enterprises. *Electronic Journal of Information Systems Evaluation*, *21* (2), pp63-79-pp63–79. Retrieved February 27, 2021, from

<https://academic-publishing.org/index.php/ejise/article/view/130>

Alaphat, A., & Jiang, M. (2020). Smartfund: Predicting research outcomes with machine learning and natural language processing. *2020 IEEE International Conference on Big Data (Big Data), Big Data (Big Data), 2020 IEEE International Conference* *on*, 2857–2865. <https://doi.org/10.1109/BigData50022.2020.9378206>

Alazab, M., Alhyari, S., Awajan, A., & Abdallah, A. B. (2021). Blockchain technology in supply chain management: An empirical study of the factors affecting user adoption/acceptance. *Cluster Computing: The Journal of Networks, Software Tools* *and Applications*, *24* (1), 83. <https://doi.org/10.1007/s10586-020-03200-4>

Alsghaier, H., Akour, M., Shehabat, I., & Aldiabat, S. (2017). The importance of big data analytics in business: A case study. *American Journal of Software Engineering and* *Applications*, *Vol. 6*, pp. 111–115. <https://doi.org/10.11648/j.ajsea.20170604.12>

Anne Laurent, Dominique Laurent, & Cédrine Madera. (2020). *Data lakes*. London,

Wiley-ISTE.

Appuswamy, R., Gkantsidis, C., Narayanan, D., Hodson, O., & Rowstron, A. (2013).

Nobody ever got fired for buying a cluster, 1–12.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and

implications for practice. *Transportation Research Part E: Logistics and* *Transportation Review*, *114*, 416–436. <https://doi.org/10.1016/j.tre.2017.04.001>

Božič, K., & Dimovski, V. (2019). Business intelligence and analytics use, innovation ambidexterity, and firm performance: A dynamic capabilities perspective. *Journal of* *Strategic Information Systems*, *28* (4). <https://doi.org/10.1016/j.jsis.2019.101578>

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32* (4), 4–39. <https://doi.org/10.1080/07421222.2015.1138364>

\_eprint: https://doi.org/10.1080/07421222.2015.1138364

Govindan, K., Fattahi, M., & Keyvanshokooh, E. (2017). Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research*, *263* (1), 108–141.<https://doi.org/10.1016/j.ejor.2017.04.009>

Gunasekaran, A., Subramanian, N., & Papadopoulos, T. (2017). Information technology for competitive advantage within logistics and supply chains: A review.

*Transportation Research Part E: Logistics and Transportation Review*, *99*, 14–33.

<https://doi.org/10.1016/j.tre.2016.12.008>

Hager, A., Goland, T., Sapio, N., & Hurt, I. (2020). Securing private medical data, and influencing medical device design to prioritize privacy: A systems analysis approach. *2020 Systems and Information Engineering Design Symposium (SIEDS), Systems and Information Engineering Design Symposium (SIEDS), 2020*, 1–3.<https://doi.org/10.1109/SIEDS49339.2020.9106633>

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

Kennedy A. Torkura, Muhammad I.H. Sukmana, & Christoph Meinel. (2017). Integrating continuous security assessments in microservices and cloud native applications.

*Utility and Cloud Computing*, 171–180. <https://doi.org/10.1145/3147213.3147229>

Kouhizadeh, M., Saberi, S., & Sarkis, J. (2021). Blockchain technology and the sustainable supply chain: Theoretically exploring adoption barriers. *International Journal of* *Production Economics*, *231*. <https://doi.org/10.1016/j.ijpe.2020.107831>

Moreno, V., Cavazotte, F., & de Souza Carvalho, W. (2020). Business intelligence and analytics as a driver of dynamic and operational capabilities in times of intense

macroeconomic turbulence. *Journal of High Technology Management Research*,

*31* (2). <https://doi.org/10.1016/j.hitech.2020.100389>

NEDSI. (2019). Artificial intelligence and supply chain management-applications and challenges. *Proceedings for the Northeast Region Decision Sciences Institute (NEDSI)*, 806–838. Retrieved April 1, 2021, from http://proxy1.ncu.edu/login?url=https://search.ebscohost.com/login.aspx?direct= true&db=bth&AN=138192773&site=eds-live

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

Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Ben Yahia, S. (2019). Data quality in etl process: A preliminary study. *Procedia Computer Science*, *159*, 676–687.<https://doi.org/10.1016/j.procs.2019.09.223>

Story, W. K., Deitz, G. D., & Richey, R. G. (2021). Influence of supply chain technology responsiveness on supply chain and market performance. *Journal of Marketing* *Theory & Practice*, 1–18. <https://doi.org/10.1080/10696679.2021.1872388>

Taagepera, R. (2008, July 24). *Making social sciences more scientific: The need for*

*predictive models*. OUP Oxford.

Talia, D. (2013). Clouds for scalable big data analytics. *Computer*, *46* (5), 98–101.<https://doi.org/10.1109/MC.2013.162>

The Economist. (2010). The data deluge.<https://www.economist.com/leaders/2010/02/25/the-data-deluge>

Trieu, V.-H. (2017). Getting value from business intelligence systems: A review and research agenda. *Decision Support Systems*, *93*, 111–124.<https://doi.org/10.1016/j.dss.2016.09.019>

Valuates Reports. (2020, March 25). *Rfid market size is expected to reach usd 26,435.12 million by 2025 | valuates reports*. Retrieved April 1, 2021, from

[https://www.prnewswire.com/news-releases/rfid-market-size-is-expected-to-reach-](https://www.prnewswire.com/news-releases/rfid-market-size-is-expected-to-reach-usd-26-435-12-million-by-2025--valuates-reports-301029588.html) [usd-26-435-12-million-by-2025--valuates-reports-301029588.html](https://www.prnewswire.com/news-releases/rfid-market-size-is-expected-to-reach-usd-26-435-12-million-by-2025--valuates-reports-301029588.html)

Walsh, G., Schubert, P., & Jones, C. (2010). Enterprise system investments for competitive advantage: An empirical study of swiss smes. *European Management Review*, *7* (3), 180–189. <https://doi.org/10.1057/emr.2010.12>

Willetts, M., Atkins, A. S., & Stanier, C. (2020). Barriers to smes adoption of big data analytics for competitive advantage. *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Intelligent Computing in Data*

*Sciences (ICDS),2020 Fourth International Conference On*, 1–8.

<https://doi.org/10.1109/ICDS50568.2020.9268687>

Young, K. (1932). Method, generalization, and prediction in social psychology. *Publications of the American Sociological Society: Papers and Proceedings, tewenty-senventh Annual Meeting*. Retrieved April 1, 2021, from<https://brocku.ca/MeadProject/Young/Young_1933.html>

Zikmund, W. (1997). *Business research methods*. Dryden. <https://books.google.com/books?id=XcIii1_gWGYC>

List of Figures

# Appendix

**\***

1. [Google Trend interest in Supply Chain Management](#_bookmark1) . . . . . . . . . . . . . . 3
2. [Google Trend interest in BDA, AI, & Blockchain](#_bookmark2) . . . . . . . . . . . . . . . . 4