



PRIMARY RESEARCH

Logistics business analytics for achieving environmental sustainability

Pei-Ju Wu *

Department of Transportation Technology and Management, Feng Chia University, Taichung, Taiwan

Index Terms

Logistics
Business analytics
Environmental sustainability
Operations management

Received: 5 July 2016**Accepted:** 10 October 2016**Published:** 19 December 2016

Abstract. Although many logistics firms seek to attain environmental sustainability, they have little knowledge of how to generate appropriate eco-strategies that can achieve environmental sustainability. Moreover, few studies have investigated how to go green in the field of logistics management by exploring logistics big data. This study aims to tackle environmental sustainability issues in logistics systems through big data analytics. Specifically, this study proposes a research framework of green logistics business analytics involving descriptive, predictive, and prescriptive analytics. The proposed logistics business analytics involve descriptive analytics, predictive analytics, and prescriptive analytics for firms to seek environmental sustainability. Logistics companies can utilize the proposed business analytics framework to identify key eco-drivers and thereby implement effective environmental logistics management. The proposed model may stimulate further research in the field of green logistics and may help address issues regarding economic and environmental sustainability.

© 2016 The Author(s). Published by TAF Publishing.

INTRODUCTION

Implementing environmental sustainability is very complicated. But, why so many firms try to go green? One reason is that they would like to achieve competitive advantage. More and more environmental regulations have been imposed on firms. If firms do not follow environmental rules, they will lose chance to fulfill economic sustainability. Hence, competitive advantage issues such as losing market and paying penalties are less likely to achieve economic sustainability. Another reason for firms to go green is that the society puts much pressure on firms to better protect the environment.

In order to satisfy the social concerns, firms have tried hard to engage in environmental efforts and eventually to maintain their reputation. Nonetheless, few studies have determined what eco-approaches are worthwhile for firms to implement to achieve environmental sustainability in logistics systems. Moreover, most logistics firms have little knowledge about how to make use of their massive data to

create valuable insight. Hence, the purpose of this study is to tackle environmental sustainability issues in logistics systems through big data analytics. In particular, this study proposes a logistics business analytics framework comprehending descriptive, predictive and prescriptive analytics. The rest of this paper is organized as follows: Section 2 presents literature review related to big data and text mining. Section 3 illustrates the proposed logistics business analytics. Finally, Section 4 conducts the discussions and conclusions of the study.

LITERATURE REVIEW

A literature review of studies related to big data and text mining has been conducted.

Big Data

Wamba, Akter, Edwards, Chopin & Gnanzou (2015) proposed a big data taxonomy to retrieve business value based

* Corresponding author: Pei-Ju Wu

† Email: wupj@fcu.edu.tw

on a longitudinal case study. Dutta & Bose (2015) investigated the case of Ramco Cements Limited and retrieved critical factors for implementing a big data project. Regarding big data and supply chain management, Zhong *et al.* (2015) developed a big data methodology to gain frequent trajectory based on RFID logistics data. Ilie-Zudor *et al.* (2015) provided a guidance to employ big data for tackling logistics issues. Opresnik & Taisch (2015) utilized big data of manufacturers to create the value in servitization. Tachizawa, Alvarez-Gil & Montes-Sancho (2015) argued that the combination of smart cities and big data can enhance supply chain management. Zhao, Yeung, Huang & Song (2015) found that the business-failure prediction of supply chain finance clients can be enhanced by employing big dataset.

Text Analytics

Approximate 80% information is represented in textual formats and valuable knowledge can be retrieved for business management through text mining techniques (Ur-Rahman & Harding, 2012). Choudhary, Harding & Tiwari (2009) applied text mining to discover knowledge linkages and research gaps from the keywords and abstracts of 150 papers. Hata, Mizuno & Kikuno (2010) mentioned that employing large-scale text characteristics can improve fault-prone module detection according to spam filtering. Cao, Duan & Gan (2011) utilized a text mining approach to analyze the helpfulness voting behavior of online users. Ma *et al.* (2012) developed an ontology text mining to cluster research proposals. Leong, Lee & Mak (2012) used sentiment analysis to explore Short Message Service (SMS) texts for teaching evaluation. Al-Zaidy, Fung, Youssef & Fortin (2012) identified criminal networks through mining suspect's text documents.

McCart, Berndt, Jarman, Finch & Luther (2013) argued that statistical text mining models can discover falls in ambulatory care clinical data. He (2013) found that case-based reasoning systems can be enhanced through text mining. Anwar & Abulaish (2014) utilized text mining to retrieve essential users, terms, sessions, and user-groups from chat logs data for cyber-crime investigation. Saha, Bose & Mahanti (2016) applied text mining to evaluate fraudulent loan processing of banks. Jin, Yan, Li & Li (2016) utilized text mining to discover healthcare information of an online Q & A community. The applications of text mining can be further summarized into business intelligence, supply chain management, and environmental sustainability.

Text mining and business intelligence: Yoon (2012) employed text mining of Web news to explore signals of business opportunities. He, Zha & Li (2013) adopted text mining to obtain crucial business knowledge of the pizza industry from Facebook and Twitter. Mostafa (2013) utilized Twitter text mining to investigate consumer brand sentiments (e.g., Nokia, T-Mobile, IBM, KLM and DHL). Chung (2014) identified crucial business intelligence elements from news articles by BizPro text mining system. Lee & Lee (2015) discovered service chances through text mining of service documents.

Text mining and supply chain management: Moriizumi, Chu, Cao & Matsukawa (2011) adopted an IBM text mining tool to identify supply chain risk drivers. Chae (2015) utilized Twitter analytics to better understand supply chain practice and research.

Kundu, Jain, Kumar & Chandra (2015) adopted the text analysis of literature review to retrieve the taxonomy of supply chain management. Abrahams, Fan, Wang, Zhang & Jiao (2015) devised a text analytic framework to detect product defect and found that distinctive terms, product characteristics, and semantic features are critical to identify product defect.

Text mining and environmental sustainability: Barke-meyer, Figge & Holt (2013) utilized text mining to better understand media agendas of sustainability covering various areas. Te Liew, Adhitya & Srinivasan (2014) utilized text mining to explore sustainability trends and practices and found that environment is a critical sustainability factor in the process industries. Shapira, Gok, Klochikhin & Sensier (2014) measured green industries, green manufacturing jobs, and green goods innovation by text mining. Rivera, Minsker, Work & Roth (2014) discovered sustainability issues and region-specific sustainability indicators through text mining techniques. Nishant, Goh & Kitchen (2016) found that water, energy, resource efficiency, and supplier integration are critical factors to achieve environmental sustainability based on text mining of Indian firms.

RESEARCH METHODOLOGY

Research Context

The proposed logistics business analytics involve descriptive analytics, predictive analytics, and prescriptive analytics for firms to seek for environmental sustainability. Figure 1 represents the framework of the proposed logistics business analytics.

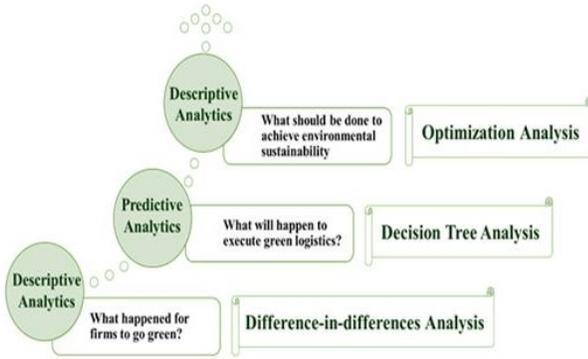


FIGURE 1 . The proposed business analytics for green logistics

Descriptive Analytics

In descriptive analytics, a difference-in-differences research design can be used to identify the impact of each eco-strategy on firm’s performance over time. For instance, the impact of key eco-drivers (e.g., emission reductions) on firm’s economic and social performance for a certain group of firms (e.g., top 20%) can be investigated through a difference-in-differences analysis. A difference-in-differences research design (Figure 2) is a causal inference methodology that approximates randomized experiments. Specifically, a difference-in-differences design can eliminate biases, such as omitted variable biases, when identifying causal relationships (Black & Kim, 2012; Imbens & Wooldridge, 2009). However, the greatest challenge in applying this methodology is its identification of treatment and control groups that have similar characteristics excluding the treatment of interest. That is, treatment firms implement green logistics and control firms do not.

For each treatment firm, this study attempts to identify a control firm that is similar in all regards excluding the point that it has not engaged in green logistics over the specified time horizon. After identifying treatment and control firms, covariate similarity can be evaluated by testing for differences between treatment firms and control firms in mean (t-statistic), median (Wilcoxon signed rank test S-statistic), and distributions (Kolmogorov-Smirnov statistic) of variables of interest. Kernel density plots can also be drawn to visualize the differences between distributions of variables of interest for the treatment firms and control firms. Furthermore, treatment distinction will be conducted to determine whether control firms differ from treatment firms in terms of the expected treatment (eco-performance change).

Once the difference between the treatment and control firms in terms of treatment and covariate similarity has been confirmed, the first difference in performance be-

tween two specified periods is evaluated for both firms. Then, the second difference (difference-in-differences) is the difference between the difference in outcome for the treatment and control firms. The initial identification of the treatment and control firms is briefly described as follows:

Treatment firms: Treatment firms are those that have implemented green logistics. After identifying firms that have engaged in green logistics, these firms will be selected from a certain top percent (e.g., top 20%) of firms in terms of a key eco-driver. These firms must also have improved in performance related to a key eco-driver during the research period.

Control firms: To control for potential industry or economy-wide effects, the performance of each treatment firm is compared with that of a suitable control firm. As a surrogate for not having engaged in green logistics, all control firms are required to have low eco-performance in relation to a key eco-driver and have no increase in performance in relation to this key eco-driver during the research period. Specifically, control firms will be selected from a certain bottom percentage (e.g., the bottom 20%) of firms in terms of a specific key eco-driver, and they must not have experienced an improvement in terms of performance as it relates to a key eco-driver during the research period.

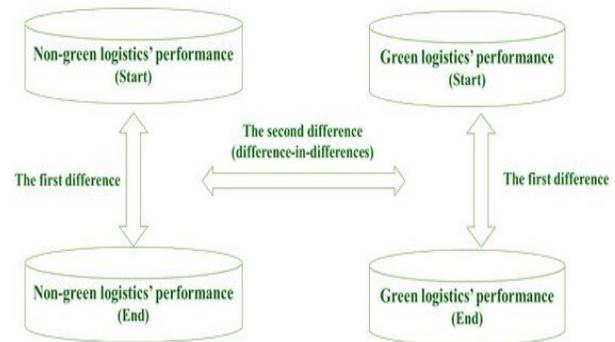


FIGURE 2 . Difference-in-differences analysis for green logistics

Predictive Analytics

In predictive analytics, decision tree analysis in data mining can be used to predict sustainable green logistics approaches. Herein, sustainability in a logistics system represents a green logistics approach with positive financial growth. The primary process to establish a decision tree model is as follows:

Select target and input attributes: According to the research purpose, the target attributes of sustainable green

logistics (positive environmental performance with economic prosperity) can be selected from all attributes. Then, the attributes related to the target attributes can be chosen as input attributes to construct the classification model.

Separate the dataset: Before developing a data mining model, the chosen dataset should be divided into training data and test data. Herein, a training data set is used to conduct a classification model while a test data set is utilized to avoid over-fitting or under-fitting.

Set up the split rule of decision tree development: Decision tree algorithms and split parameters, such as split rule, maximum branch, and maximum tree depth, should be determined. It should be noted that the parameters need to be further examined and adjusted according to the accuracy of classification and meaningful managerial implications.

Implement decision tree modeling: The conceptual decision tree model of sustainable green logistics is represented in Figure 3. An importance ranking of data attributes for the splitting decision tree can be acquired via decision tree modeling. Moreover, classification rules of sustainable green logistics can eventually be created.

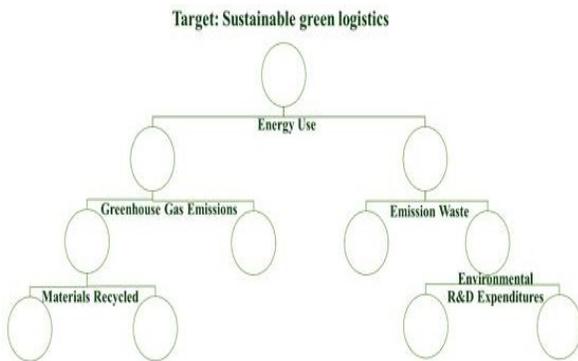


FIGURE 3 . The conceptual decision tree model of sustainable green logistics

Classification results from decision tree analysis can be interpreted and evaluated, including the priority rank of input variables for the splitting decision tree, the accuracy of the decision tree classification rate, and the coverage rate of leaf nodes. The importance of each input variable to achieve sustainable green logistics can be determined by evaluating the split priority of input variables on the decision tree. Furthermore, examining the accurate classification rate of validation data and test data as well as the coverage rate of leaf nodes can ensure the predictive ability of classification results by the decision tree. Additionally, the split and span

of the decision tree should be carefully analyzed from managerial perspectives to avoid an over-fitting model. Thereafter, managers can develop their green logistics strategies that foster economic prosperity at the same time.

Prescriptive Analytics

In prescriptive analytics, optimization models can be utilized to seek for optimal green logistics value. An optimization model of green logistics (Fig. 4) involves four components as follows:

1. Objective function: An objective function with green logistics aims to minimize environmental pollution (e.g., emissions, noise, etc.) while maximize operations' profitability.
2. Constraints: Operations limitations (e.g., limited green resources, flow conservation, capacity of logistics members, etc.) that restrict pursuing optimization should be formulated in constraints.
3. Decision variables: Decision variables identify decisions to be made such as optimized green logistics flows, optimized green logistics network, etc.
4. Parameters: Parameters describe information which is given in the model. For example, total amount of hazardous waste produced per mile for green logistics.

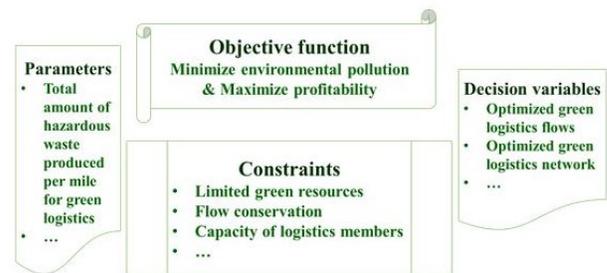


FIGURE 4 . The conceptual optimization model of sustainable green logistics

DISCUSSION AND CONCLUSION

Most firms lack knowledge and plans about how to go green. This study proposed a logistics business analytics framework for firms to achieve environmental sustainability. This study gives academics and managers a holistic view of strategies for making logistics decisions about going green with economic prosperity. The results from this study revealed some crucial properties as follows: The proposed logistics business analytics comprise of three components including descriptive analytics, predictive analytics, and prescriptive analytics. Descriptive analytics investi-

gate what happened for firms to go green. Difference-in-differences analysis can be applied in descriptive analytics. Moreover, predictive analytics address what will happen to execute green logistics.

Decision tree analysis with a data mining technique can be utilized in predictive analytics. Furthermore, descriptive analytics explore what should be done to achieve environmental sustainability. Optimization models can be employed in descriptive analytics. The proposed model may stimulate further research in the field of green logis-

tics and may help address issues regarding economic and environmental sustainability.

Acknowledgment

This research was supported in part by the Ministry of Science and Technology, Taiwan, under Contract No. MOST 104-2410-H-035-001. This support is gratefully appreciated. Any errors or omissions remain the sole responsibility of the author.

REFERENCES

- Abrahams, A.S., Fan, W., Wang, G.A., Zhang, Z.J., & Jiao, J. 2015. An integrated text analytic framework for product defect discovery. *Production & Operations Management*, 24(6): 975-990. DOI: [10.1111/poms.12303](https://doi.org/10.1111/poms.12303)
- Al-Zaidy, R., Fung, B.C., Yousef, A.M., & Fortin, F. 2012. Mining criminal networks from unstructured text documents. *Digital Investigation*, 8(3): 147-160. DOI: [10.1016/j.diin.2011.12.001](https://doi.org/10.1016/j.diin.2011.12.001)
- Anwar, T., & Abulaish, M. 2014. A social graph based text mining framework for chat log investigation. *Digital Investigation*, 11(4): 349-362. DOI: [10.1016/j.diin.2014.10.001](https://doi.org/10.1016/j.diin.2014.10.001)
- Barkemeyer, R., Figge, F., & Holt, D. 2013. Sustainability-related media coverage and socioeconomic development: A regional and North-South perspective. *Environment & Planning: Government & Policy*, 31(4): 716-740. DOI: [10.1068/c11176j](https://doi.org/10.1068/c11176j)
- Black, B., & Kim, W. 2012. The effect of board structure on firm value: A multiple identification strategies approach using Korean data. *Journal of Financial Economics*, 104(1): 203-226. DOI: [10.1016/j.jfineco.2011.08.001](https://doi.org/10.1016/j.jfineco.2011.08.001)
- Cao, Q., Duan, W., & Gan, Q. 2011. Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. *Decision Support Systems*, 50(2): 511-521. DOI: [10.1016/j.dss.2010.11.009](https://doi.org/10.1016/j.dss.2010.11.009)
- Chae, B.K. 2015. Insights from hashtag supply chain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165: 247-259. DOI: [10.1016/j.ijpe.2014.12.037](https://doi.org/10.1016/j.ijpe.2014.12.037)
- Choudhary, A.K., Harding, J.A., & Tiwari, M.K. 2009. Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5): 501-521. DOI: [10.1007/s10845-008-0145-x](https://doi.org/10.1007/s10845-008-0145-x)
- Chung, W. 2014. Biz pro: Extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management*, 34(2): 272-284. DOI: [10.1016/j.ijinfomgt.2014.01.001](https://doi.org/10.1016/j.ijinfomgt.2014.01.001)
- Dutta, D., & Bose, I. 2015. Managing a big data project: The case of Ramco cements limited. *International Journal of Production Economics*, 165, 293-306. DOI: [10.1016/j.ijpe.2014.12.032](https://doi.org/10.1016/j.ijpe.2014.12.032)
- Hata, H., Mizuno, O., & Kikuno, T. 2010. Fault-prone module detection using large-scale text features based on spam filtering. *Empirical Software Engineering*, 15(2): 147-165. DOI: [10.1007/s10664-009-9117-9](https://doi.org/10.1007/s10664-009-9117-9)
- He, W. 2013. Improving user experience with case-based reasoning systems using text mining and Web 2.0. *Expert Systems with Applications*, 40(2): 500-507. DOI: [10.1016/j.eswa.2012.07.070](https://doi.org/10.1016/j.eswa.2012.07.070)
- He, W., Zha, S., & Li, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3): 464-472. DOI: [10.1016/j.ijinfomgt.2013.01.001](https://doi.org/10.1016/j.ijinfomgt.2013.01.001)
- Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P., & Monostori, L. 2015. Advanced predictive-analysis-based decision support for collaborative logistics networks. *Supply Chain Management: An International Journal*, 20(4): 369-388. DOI: [10.1108/SCM-10-2014-0323](https://doi.org/10.1108/SCM-10-2014-0323)
- Imbens, G.W., & Wooldridge, J.M. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1): 5-86. DOI: [10.1257/jel.47.1.5](https://doi.org/10.1257/jel.47.1.5)
- Jin, J., Yan, X., Li, Y., & Li, Y. 2016. How users adopt healthcare information: An empirical study of an online Q & A community. *International Journal of Medical Informatics*, 86: 91-103. DOI: [10.1016/j.ijmedinf.2015.11.002](https://doi.org/10.1016/j.ijmedinf.2015.11.002)

- Kundu, A., Jain, V., Kumar, S., & Chandra, C. 2015. A journey from normative to behavioral operations in supply chain management: A review using Latent Semantic Analysis. *Expert Systems with Applications*, 42(2): 796-809.
DOI: [10.1016/j.eswa.2014.08.035](https://doi.org/10.1016/j.eswa.2014.08.035)
- Lee, C., & Lee, H. 2015. Novelty-focussed document mapping to identify new service opportunities. *The Service Industries Journal*, 35(6): 345-361. DOI: [10.1080/02642069.2015.1003368](https://doi.org/10.1080/02642069.2015.1003368)
- Leong, C.K., Lee, Y.H., & Mak, W.K. 2012. Mining sentiments in SMS texts for teaching evaluation. *Expert Systems with Applications*, 39(3): 2584-2589. DOI: [10.1016/j.eswa.2011.08.113](https://doi.org/10.1016/j.eswa.2011.08.113)
- Te Liew, W., Adhitya, A., & Srinivasan, R. 2014. Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry*, 65(3): 393-400. DOI: [10.1016/j.compind.2014.01.004](https://doi.org/10.1016/j.compind.2014.01.004)
- Ma, J., Xu, W., Sun, Y.H., Turban, E., Wang, S., & Liu, O. 2012. An ontology-based text-mining method to cluster proposals for research project selection. *IEEE Transactions on Systems, Man & Cybernetics-Part A: Systems & Humans*, 42(3): 784-790. DOI: [10.1109/TSMCA.2011.2172205](https://doi.org/10.1109/TSMCA.2011.2172205)
- McCart, J.A., Berndt, D.J., Jarman, J., Finch, D.K., & Luther, S.L. 2013. Finding falls in ambulatory care clinical documents using statistical text mining. *Journal of the American Medical Informatics Association*, 20(5): 906-914.
DOI: [10.1136/amiajnl-2012-001334](https://doi.org/10.1136/amiajnl-2012-001334)
- Moriizumi, S., Chu, B., Cao, H., & Matsukawa, H. 2011. Supply chain risk driver extraction using text mining technique. *Information*, 14(6): 1935-1945.
- Mostafa, M.M. 2013. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10): 4241-4251. DOI: [10.1016/j.eswa.2013.01.019](https://doi.org/10.1016/j.eswa.2013.01.019)
- Nishant, R., Goh, M., & Kitchen, P.J. 2016. Sustainability and differentiation: Understanding materiality from the context of Indian firms. *Journal of Business Research*, 69(5): 1892-1897. DOI: [10.1016/j.jbusres.2015.10.075](https://doi.org/10.1016/j.jbusres.2015.10.075)
- Opresnik, D., & Taisch, M. 2015. The value of big data in servitization. *International Journal of Production Economics*, 165: 174-184. DOI: [10.1016/j.ijpe.2014.12.036](https://doi.org/10.1016/j.ijpe.2014.12.036)
- Rivera, S.J., Minsker, B.S., Work, D.B., & Roth, D. 2014. A text mining framework for advancing sustainability indicators. *Environmental Modelling & Software*, 62: 128-138. DOI: [10.1016/j.envsoft.2014.08.016](https://doi.org/10.1016/j.envsoft.2014.08.016)
- Saha, P., Bose, I., & Mahanti, A. 2016. A knowledge based scheme for risk assessment in loan processing by banks. *Decision Support Systems*, 84: 78-88. DOI: [10.1016/j.dss.2016.02.002](https://doi.org/10.1016/j.dss.2016.02.002)
- Shapira, P., Gök, A., Klochikhin, E., & Sensier, M. 2014. Probing green industry enterprises in the UK: A new identification approach. *Technological Forecasting & Social Change*, 85: 93-104. DOI: [10.1016/j.techfore.2013.10.023](https://doi.org/10.1016/j.techfore.2013.10.023)
- Tachizawa, E.M., Alvarez-Gil, M.J., & Montes-Sancho, M.J. 2015. How smart cities will change supply chain management. *Supply Chain Management: An International Journal*, 20(3): 237-248. DOI: [10.1108/SCM-03-2014-0108](https://doi.org/10.1108/SCM-03-2014-0108)
- Ur-Rahman, N., & Harding, J.A. 2012. Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications*, 39(5): 4729-4739. DOI: [10.1016/j.eswa.2011.09.124](https://doi.org/10.1016/j.eswa.2011.09.124)
- Wamba, S.F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. 2015. How big data can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165: 234-246.
DOI: [10.1016/j.ijpe.2014.12.031](https://doi.org/10.1016/j.ijpe.2014.12.031)
- Yoon, J. 2012. Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Systems with Applications*, 39(16): 12543-12550. DOI: [10.1016/j.eswa.2012.04.059](https://doi.org/10.1016/j.eswa.2012.04.059)
- Zhong, R.Y., Huang, G.Q., Lan, S., Dai, Q.Y., Chen, X., & Zhang, T. 2015. A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics*, 165: 260-272.
DOI: [10.1016/j.ijpe.2015.02.014](https://doi.org/10.1016/j.ijpe.2015.02.014)
- Zhao, X., Yeung, K., Huang, Q., & Song, X. 2015. Improving the predictability of business failure of supply chain finance clients by using external big dataset. *Industrial Management & Data Systems*, 115:9, 1683-1703.
DOI: [10.1108/IMDS-04-2015-0161](https://doi.org/10.1108/IMDS-04-2015-0161)

— This article does not have any appendix. —