

Big Data and Predictive Analytics in the Construction Industry: Applications, Status Quo, and Potential in Singapore's Construction Industry

Jasmine Ngo¹; Bon-Gang Hwang²; and Chenyue Zhang³

¹Dept. of Building, National Univ. of Singapore, Singapore (corresponding author). E-mail: jasminengo@u.nus.edu

²Dept. of Building, National Univ. of Singapore, Singapore

³CTBH Singapore, Singapore

ABSTRACT

Despite the hype for big data and predictive analytics (BDPA) in the recent decade, development of BDPA in the construction industry is significantly falling behind other industries. Hence, this paper discusses the key applications of BDPA, status quo, and potential of BDPA implementation in Singapore's construction industry. Based on the literature review, twenty-five determinants that impact an BDPA implementation were identified and formed the basis of the survey questionnaire. The study found that only 7.1% of the construction organizations adopt BDPA. The top applications were found to be cost estimation, delay prediction, and energy management, and consumption prediction. The top predictive analytics methods adopted are simple linear regression, decision tree, time series forecast, and multiple regression. The survey results indicate that construction organizations have suitable data, operations, technical support and training, and culture for BDPA implementation, but should work on the adoption of BDPA technologies. This research provides better understanding of the applications, status quo, and potential of BDPA implementation in the construction industry, and provides a guide for organizations to implement BDPA into their practices, and serves as a starting point for future research on increasing implementation of BDPA in the construction industry in both the local and global contexts.

INTRODUCTION

The big data (BD) market is forecasted to more than double its market size from 2018 and reach a market value of US\$103 billion in 2027 (Wikibon and SiliconANGLE 2018). With increasing amount of data generated at high velocity from various sources in the construction industry and the advancements in data analytics techniques, the use of predictive analytics has been increasing. Big data and predictive analytics (BDPA) can be applied throughout the project lifecycle and present great potential for the construction industry. However, there are limited BDPA studies that focus on the construction industry. Hence, this study discusses the determinants that impact implementation of BDPA, key BDPA applications in the construction industry and assesses the status quo and potential of BDPA implementation in Singapore's construction industry. The findings of this study can serve as a starting point for future research on increasing BDPA implementation and to develop a guide for organizations to implement BDPA in both the local and global context as BDPA technologies advance and have increasing potential to improve the performance of the construction industry.

BACKGROUND

Big Data and Big Data Analytics

BD has become a ubiquitous term globally. While Big Data Engineering (BDE) provides supporting functions like data storage and processing for analytics, Big Data Analytics (BDA) is concerned with the extraction of knowledge from large datasets to support decision-making (Bilal et al. 2016). Analytics is the process of developing actionable insights through problem definition and the application of statistical techniques on existing or simulated data (Sivarajah et al. 2017).

Analytics is typically categorized into descriptive, diagnostic, predictive and prescriptive analytics. Descriptive analytics categorizes and classifies data to generate information on existing organization performance and is most frequently used (Evans and Lindner 2012). Diagnostic analytics evaluates the potential causes of a problem using exploratory data analysis. Predictive analytics uses statistical techniques to predict the future based on historical data. Prescriptive analytics explores potential decisions that involve complex objectives to optimize organizational performance (Wang et al. 2016). The use of predictive analytics is increasing with growing volume of data and more affordable processing power. On the other hand, there are only a few notable examples of good prescriptive analytics currently due to the insufficient number of dimensions captured in existing databases (Banerjee et al. 2013).

Common predictive analytics techniques fall under the broader category of supervised machine learning techniques, including regression and classification. Regression is a technique to predict the target variable's numerical value based on input variables while classification is the process of emulating decisions automatically through trained data sets (Bilal et al. 2016). Common variations of regression include simple linear regression, multiple linear regression and logistic regression. Prominent classification algorithms include decision tree, Naïve Bayes, neural networks, K-nearest neighbor, support vector machine, random forest, time series forecasting and partial least squares (Bilal et al. 2016; Dietrich et al. 2015). These techniques may be extended to deal with BD (Bilal et al. 2016).

Applications of BDPA in the Construction Industry

BDPA can be applied throughout the project lifecycle. In the pre-tender stage, BDPA can be applied to support site selection, as BD can synthesize demographic, economic and labour information to generate intelligent location selection decisions (Ballard-Bloomfield and Brook 2018). In the design and planning stage, BDPA can be applied in tender price evaluation (Zhang et al. 2015), failure prediction model for construction firms (Alaka et al. 2018), prediction of cost overrun (Williams and Gong 2014) and project delays (Bilal et al. 2016). In the construction stage, BDPA can be used to minimize unsafe worker behavior (Han et al. 2012), construction waste management (Lu et al. 2015) and construction site management (Tibaut and Zazula 2018). In the operations and maintenance stage, BDPA can be used in energy management to assess the energy saving performance and to predict energy consumption based on historical trends (Linder et al. 2017; Song et al. 2017).

Government Initiatives in Singapore

Singapore's government has introduced several initiatives to enhance BDPA adoption in the construction industry. Various government authorities specifically – Building and Construction

Authority (BCA), Smart Nation and Digital Government Office (SNDGO) and Infocomm Media Development Authority (IMDA), have introduced initiatives to provide technical, financial and regulatory support for organizations looking to implement advanced technologies. These agencies also serve as platforms to bridge the public and private sectors and promote the training of technologically skilled individuals. These initiatives aim to reduce the risks associated with BDPA adoption and increase the supply of skilled professionals.

Determinants that Impact BDPA Implementation

Twenty-five determinants that impact BDPA implementation have been identified and categorized into five determinant groups (DG). These determinants would be used as the basis to assess the potential for BDPA implementation.

In DG1, the determinants are related to the data characteristics for BDPA implementation and include: *data volume* (D1); *data velocity* (D2); *data variety* (D3); and *data veracity* (D4) (Beyer and Laney 2012; IBM 2013). Sufficient data is required for predictive analytics to identify patterns of value. In addition, the reliability of the data inputs is crucial to ensure the reliability of the analysis results.

The next DG is concerned with organization operations and consists of: *clear standard operating procedure* (SOP) (O1); *financial capability* (O2); *sufficient time* (O3); *dedicated team* (O4); and *regular review* (O5) (Gupta and George 2016; Hazen et al. 2012; Phillips 2013). Organizations have to allocate sufficient resources, commitment and develop the right processes to maximize the potential of BDPA.

The third DG (DG3) is related to technology and software which includes: *data visualization tool* (T1); *cloud-based platforms for data processing and analytics* (T2); *open source software for BDPA* (T3); and *new forms of databases for data storage* (T4) (Davenport 2014; Gupta and George 2016). To implement BDPA, the technology and software adopted should have sufficient capability, scalability, and an user-friendly interface.

Availability of technical support and training (DG4) also impacts BDPA implementation, and can be measured by the *provision of technical support* (S1); *training* (S2); and *skill sets and competency* (S3) (Gupta and George 2016; Hazen et al. 2012; Phillips 2013). Workers should be competent in handling and managing data, and supported by the organization in terms of provision of technical support and data analytics training to implement BDPA.

Lastly, organizational culture (DG5) plays an important role in the implementation of BDPA and consists of: *data-driven culture* (C1); *top management support for innovation* (C2); *alignment with organization's vision* (C3); *continuous improvement* (C4); and *adaptability to change* (C5) (Bi and Cochran 2014; Davenport and Dyché 2013; Gupta and George 2016; Liang et al. 2007). It is crucial that the organization's top management support and lead the organization into adopting BDPA to maximize the potential of BDPA. Furthermore, for organizations to stay relevant, *industry trends on the innovative technologies adopted* (C6) and *the research and adoption of data analytics tools* (C7) should be followed. At the same time, the organization should *stay updated on the government policies, regulations and initiatives related to the innovation technologies for the industry* (C8), and *BDPA* (C9), so that organizations are kept aware of the market direction and opportunities in BDPA implementation.

RESEARCH METHODS AND DATA PRESENTATION

The research process consists of four steps. In Step 1, a literature review was conducted to establish a foundation for the study and the development of the survey questionnaire. Pilot

interviews were carried out with industry experts to validate the survey questionnaire in Step 2. Step 3 was to administer the survey questionnaire to industry practitioners to assess the status quo of BDPA implementation in Singapore and its key applications. In Step 4, analysis of the responses were conducted through Statistical Package for the Social Sciences (SPSS) to determine the status quo and potential of BDPA implementation.

The survey was sent to 170 target respondents and a total of 42 responses were received. The survey response rate is 24.7% and is within the acceptable range (Alreck and Settle 2004). The organizations represented the major business domains in the construction industry including contractors (28.6%), consultants (19.0%), property management firms (19.0%), quantity surveyors (11.9%), developers (9.5%), government agencies (9.5%) and others (2.4%). Furthermore, over half (57.1%) of the organizations have more than 20 years of experience in the construction industry and 31.0% of organizations have more than 10 years of experience of using predictive data analytics.

In addition, survey respondents' positions are heterogeneous in nature, with a good coverage of project managers (33.3%), facility managers (19.0%), engineers (16.7%), quantity surveyors (11.9%), government officers (9.5%), architects (7.1%) and others (2.4%). About two-thirds (64.3%) of the respondents have more than 10 years of experience in the construction industry. The respondents also had varying number of years of using predictive data analytics, where about half of them (54.8%) had 0 to 5 years of experience, 23.8% with 5 to 10 years of experience and 21.4% of them with more than 10 years of experience of using predictive data analytics.

The heterogeneity of the respondent panel provides reliable and meaningful analysis. Shapiro-Wilk test was used to test the normality of the sample. For non-normal distributions, Mann-Whitney nonparametric test was performed to test if there are any statistical differences between two groups of respondents.

DATA ANALYSIS AND DISCUSSION

The applications of data analytics, status quo of BDPA implementation and the potential for BDPA implementation have been analyzed. The results will be further detailed in this section.

Applications of Data Analytics in the Construction Industry

The most common applications of data analytics in the construction industry were found to be cost estimation (29.1%), delay prediction (18.2%) and energy management and consumption prediction (12.7%). However, the applications differ according to the organizations' business domains. This is expected as the value of BDPA is generated when data is analyzed to support decision-making in business operations (Phillips 2013). For contractor and consultant firms, the top applications are related to project management, specifically delay prediction, cost estimation and construction waste and site management. Developer firms are concerned with site selection and business failure prediction, which are in alignment with the interests of the business. Property management firms are concerned with energy management and consumption prediction. Quantity surveying firms are concerned with project costs, using data analytics for tender price evaluation and cost estimation. BDPA can be applied in several areas, presenting great potential for the construction industry to benefit from BDPA.

Table 1. Summary of Analysis

Determinant Group	Code	Mean	Shapiro-Wilk Test	Mann Whitney U Test (Industry Experience)	Mann Whitney U Test (Data Analytics Experience)
DG1 – Data Characteristics	D2	3.738	0.000	0.154	0.714
	D3	3.857	0.000	0.989	0.609
	D4	3.357	0.000	0.459	0.149
DG2 – Organization Operations	O1	3.167	0.001	0.075	0.006*
	O2	3.048	0.001	0.832	0.119
	O3	3.143	0.002	0.430	0.122
	O4	2.524	0.001	0.834	0.888
	O4*	3.111	0.032	0.233	0.003*
	O5	3.238	0.002	0.571	0.259
DG4 – Technical Support and Training	S1	3.310	0.002	0.285	0.586
	S2	3.476	0.001	0.048*	0.021*
DG5 – Organizational Culture	C1	4.024	0.000	0.064	0.003*
	C2	4.190	0.000	0.337	0.014*
	C3	3.786	0.000	0.605	0.287
	C4	4.381	0.000	0.888	0.763
	C5	4.286	0.000	0.699	0.189
	C6	3.833	0.000	0.361	0.589
	C7	3.714	0.000	0.336	0.094
	C8	4.095	0.000	0.756	0.965
	C9	3.810	0.000	0.979	0.853

*Respondents who have indicated that their organizations have no existing data analytics team were directed to indicate if their organization has plans to form such a team in the near future

Status Quo of BDPA Implementation

On the organizational level, only 7.1% of the organizations practice BDPA, and 4.8% of the organizations practice advanced analytics on BD. On the other hand, 14.3% of the organizations do not practice any form of advanced analytics. Majority of the organizations (61.9%) practice advanced and predictive analytics, but do not apply to BD. The results indicate that the existing level of BDPA implementation in Singapore's construction industry is low.

The most popular BDPA technique adopted was found to be simple linear regression. Decision tree is another common technique among the top three methods adopted by organizations. Organizations with more data analytics experience favour time series forecast

while those with less data analytics experience favour multiple linear regression. It was noted that organizations prefer common analytical methods such as regression instead of complex machine learning techniques. This may be due to the lack of skilled professionals, lack of tangible resources for research and implementation, reluctance by management to invest in innovation and the lack of perceived return on investment (Sepasgozar et al. 2016).

Potential of BDPA Implementation in Singapore's Construction Industry

With the low BDPA implementation rate in Singapore's construction industry, the potential of BDPA implementation has been analyzed according to the determinants identified above. Table 1 shows the summary of the analysis conducted.

Determinant Group 1 - Data Characteristics

Majority indicate that existing data volume is within the range of 1 – 100TB (69.1%) and is expected to increase in the next five years. Most of the data was found to be unstructured data from contract documents, schedules, drawings, specifications and site images. Advancements and increased adoption of technologies such as Internet-of-Things and Cyber-Physical Systems will increase the data volume, velocity, variety and veracity generated (Bilal et al. 2016). With regular progress reports and updates on project progress, the respondents largely agree that there is sufficient data velocity, variety and veracity in their organizations. No statistical differences were found in the responses between organizations of varying years of experience. The results indicate that construction organizations have suitable data for BDPA implementation.

Determinant Group 2 - Organization Operations

The results indicate that organization operations are moderately suitable for BDPA implementation. However, it was found that 42.9% of the organizations do not currently have dedicated data analytics team. Out of the 42.9%, organizations with more data analytics experience were found to be more willing to prepare for such teams in the near future. This is reasonable as organizations with more data analytics experience may have better understanding of the benefits brought about by adopting data analytics (Liu et al. 2018). At the same time, organizations with more data analytics experience were found to have more defined SOP for data analytics. It is intuitive for experienced firms to have mature data management systems and procedures to guide employees through data analytics to ensure a minimum standard, consistency and accuracy of the data analysis conducted (Nakagawa 2005).

Determinant Group 3 - Technology and Software

Survey results indicate that 64.3% of the organizations do not adopt new forms of databases or data warehousing technologies. Similarly, 61.9% of the organizations do not adopt cloud-based platforms for data processing and analytics. In terms of data processing, Microsoft Excel (88.1%) is the top ranked technology adopted. The results indicate that adoption of technologies used for BDPA is relatively low, reflecting a general resistance to new technologies in the construction industry.

Determinant Group 4 - Technical Support and Training

The results indicate that organizations have moderate availability of technical support and

provision of structured data analytics training. However, more experienced organizations tend to provide more data analytics trainings. This may be because more project data can be accumulated from past projects, and mature data management systems have been developed, both of which require skilled employees to manage and maintain the databases.

Determinant Group 5 - Organizational Culture

The results indicate that the existing organizational culture appear to be suitable for BDPA implementation. However, it was found that organizations with more data analytics experience tend to have more data-driven culture and get more support from top management. These may be critical impetus for BDPA implementation. At the same time, it was found that organizations tend to keep track of government-related policies, regulations and initiatives more often than market trends and practices. This is reasonable as government-related policies, regulations and initiatives may include the various incentives provided by the government and new regulations that may affect the operations of the organizations (Alzadjali and Elbanna 2019).

Discussion

Although the results indicate that existing data, organizational operations, technical support and training and organizational culture are moderately suitable for BDPA implementation, the level of BDPA implementation and use of BDPA technologies in the industry remains low. This may be contributed by the nature of the industry, where construction projects are short term, complex with multiple stakeholders involved and made up of a majority of small and medium enterprise with limited financial capability to invest in new technologies (Oesterreich and Teuteberg 2016; Sepasgozar et al. 2016). As such, challenges such as interoperability issues, poor training and skills development of workers, unwillingness to invest in new technologies arise. It is critical for the government and industry to work together to further drive BDPA implementation. For example, the government can provide financial incentives for organizations that have implemented BDPA and monetary subsidies for employee trainings (Chan et al. 2017; Suprun and Stewart 2015).

CONCLUSION

BDPA implementation in Singapore's construction industry is still in its infancy with only 7.1% of the construction organizations adopting BDPA. This could be due to the general resistance to new technologies, lack of skilled professionals and excess resources for the development and training for BDPA implementation. The top applications of BDPA were found to be cost estimation, delay prediction and energy management and consumption prediction while the top predictive analytics methods adopted are simple linear regression, decision tree, time series forecast and multiple regression. Despite the low BDPA implementation rate, the survey results indicate that construction organizations have suitable data, operations, technical support and training and culture for BDPA implementation, but should work on the adoption of BDPA technologies. Hence, there is great potential for the construction industry to benefit from BDPA.

This study is limited in several aspects. The sample size is relatively small, although the various roles and business domains within the industry are represented. More reliable results may be produced with a larger sample size. Next, respondents are unable to clarify doubts immediately as the survey is conducted through an online platform. This is mitigated through the

provision of explanatory notes and diagrams in the questionnaire.

The findings from this study provide a better understanding of the key applications, status quo and potential of BDPA implementation in Singapore's construction industry. The findings of the study serve as a starting point for organizations looking to implement BDPA into their practices and for future research on increasing the implementation rate of BDPA in the construction industry, which can be extended to the global context as BDPA technologies continue to advance, presenting greater potential to improve the performance of the construction industry. To further increase the BDPA implementation in both the local and global construction industry, future studies may include developing a BDPA capability assessment tool, to allow construction organizations to evaluate their BDPA capabilities and to serve as a guide to raise their BDPA capability levels. Solutions to overcome the challenges of BDPA implemented can also be further investigated. In addition, the determinants that impact BDPA implementation should also be reviewed and updated in future studies.

REFERENCES

- Alaka, H., Oyedele, L., Owolabi, H., Akinade, O., Bilal, M., and Ajayi, S. (2018). "A Big Data Analytics Approach for Construction Firms Failure Prediction Models." *IEEE Transactions on Engineering Management*, 1–10.
- Alreck, P. L., and Settle, R. B. (2004). *The survey research handbook*. McGraw-Hill, Boston.
- Alzadjali, K., and Elbanna, A. (2019). "Smart Institutional Intervention in the Adoption of Digital Infrastructure: The Case of Government Cloud Computing in Oman." *Information Systems Frontiers*.
- Ballard-Bloomfield, K., and Brook, D. (2018). "How Can Big Data Enable More Efficient and Effective Location Decisions and Planning? - Area Development." <<https://www.areadevelopment.com/corporate-site-selection-factors/Q3-2018/big-data-efficient-effective-location-decisions-and-planning.shtml>> (May 30, 2019).
- Banerjee, A., Bandyopadhyay, T., and Acharya, P. (2013). "Data Analytics: Hyped Up Aspirations or True Potential?" *Vikalpa: The Journal for Decision Makers*, SAGE PublicationsSage India: New Delhi, India, 38(4), 1–12.
- Beyer, M. A., and Laney, D. (2012). "The importance of 'big data': A definition." <<https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition>> (Apr. 24, 2019).
- Bi, Z., and Cochran, D. (2014). "Big data analytics with applications." *Journal of Management Analytics*, Taylor & Francis, 1(4), 249–265.
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., and Pasha, M. (2016). "Big Data in the construction industry: A review of present status, opportunities, and future trends." *Advanced Engineering Informatics*, 30, 500–521.
- Chan, A. P. C., Darko, A., and Ameyaw, E. E. (2017). "Strategies for promoting green building technologies adoption in the construction industry-An international study." *Sustainability*, 9(6).
- Davenport, T. H. (2014). "How strategists use 'big data' to support internal business decisions, discovery and production." *Strategy and Leadership*, 42(4), 45–50.
- Davenport, T. H., and Dyché, J. (2013). *Big Data in Big Companies*.
- Dietrich, D., Heller, R., Yang, B., and EMC Education Services. (2015). *Data science and big data analytics : discovering, analyzing, visualizing and presenting data*.

- Evans, J. R., and Lindner, C. H. (2012). "Business Analytics: The Next Frontier for Decision Sciences." *Decision Line*, <http://faculty.cbpp.uaa.alaska.edu/afef/business_analytics.htm> (May 23, 2019).
- Gupta, M., and George, J. F. (2016). "Toward the development of a big data analytics capability." *Information & Management*, 53(8), 1049–1064.
- Han, S. U., Lee, S. H., and Peña-Mora, F. (2012). "A Machine-Learning Classification Approach to Automatic Detection of Workers' Actions for Behavior-Based Safety Analysis." *Computing in Civil Engineering* (2012), American Society of Civil Engineers, Reston, VA, 65–72.
- Hazen, B. T., Overstreet, R. E., and Cegielski, C. G. (2012). "Supply chain innovation diffusion: going beyond adoption." *The International Journal of Logistics Management*, Emerald Group Publishing Limited, 23(1), 119–134.
- IBM. (2013). "What is big data? - Bringing big data to the enterprise." <<http://www-01.ibm.com/software/data/bigdata/>> (Apr. 30, 2019).
- Liang, H., Saraf, N., Hu, Q., and Xue, Y. (2007). "Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management." *MIS Quarterly*, Management Information Systems Research Center, University of Minnesota, 31(1), 59–87.
- Linder, L., Vionnet, D., Bacher, J. P., and Hennebert, J. (2017). "Big Building Data-a Big Data Platform for Smart Buildings." *Energy Procedia*, Elsevier, 589–594.
- Liu, D., Lu, W., and Niu, Y. (2018). "Extended Technology-Acceptance Model to Make Smart Construction Systems Successful." *Journal of Construction Engineering and Management*, 144(6), 04018035.
- Lu, W., Chen, X., Peng, Y., and Shen, L. (2015). "Benchmarking construction waste management performance using big data." "Resources, Conservation & Recycling," Elsevier, 105, 49–58.
- Nakagawa, Y. (2005). "Importance of Standard Operating Procedure Documents and Visualization to Implement Lean Construction." *13th Annual Conference of the International Group for Lean Construction*, Sydney, Australia, 207–215.
- Oesterreich, T. D., and Teuteberg, F. (2016). "Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry." *Computers in Industry*, 83, 121–139.
- Phillips, J. (2013). *Building a Digital Analytics Organization : Create Value by Integrating Analytical Processes, Technology, and People Into Business Operations*. FT Press Analytics, Pearson FT Press, Upper Saddle River, N.J.
- Sepasgozar, S. M. E., Loosemore, M., and Davis, S. R. (2016). "Conceptualising information and equipment technology adoption in construction: A critical review of existing research." *Architectural Management*, 23, 158–176.
- Sivarajah, U., Kamal, M. M., Irani, Z., and Weerakkody, V. (2017). "Critical analysis of Big Data challenges and analytical methods." *Journal of Business Research*, Elsevier, 70, 263–286.
- Song, K., Kwon, N., Anderson, K., Park, M., Lee, H. S., and Lee, S. H. (2017). "Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups." *Energy and Buildings*, Elsevier, 156, 121–133.
- Suprun, E. V., and Stewart, R. A. (2015). "Construction innovation diffusion in the Russian Federation." *Construction Innovation*, 15(3), 278–312.

- Tibaut, A., and Zazula, D. (2018). "Sustainable management of construction site big visual data." *Sustainability Science*, Springer Japan, 13(5), 1311–1322.
- Wang, G., Gunasekaran, A., Ngai, E. W. T., and Papadopoulos, T. (2016). "Big data analytics in logistics and supply chain management: Certain investigations for research and applications." *International Journal of Production Economics*, 176, 98–110.
- Wikibon, and SiliconANGLE. (2018). "Global Big Data market size 2011-2027." *Statista*, <<https://www.statista.com/statistics/254266/global-big-data-market-forecast/>> (May 2, 2019).
- Williams, T. P., and Gong, J. (2014). "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers." *Automation in Construction*, Elsevier, 43, 23–29.
- Zhang, Y., Luo, H., and He, Y. (2015). "A System for Tender Price Evaluation of Construction Project Based on Big Data." *Procedia Engineering*, Elsevier, 606–614.