

# Challenges and drivers for data mining in the AEC sector

Vian Ahmed and Zeeshan Aziz

*School of the Built Environment, University of Salford, Manchester, UK*

Algan Tezel

*University of Huddersfield, Huddersfield, UK, and*

Zainab Riaz

*Suleman Dawood School of Business, Lahore University of Management Sciences, Lahore, Pakistan*

## Abstract

**Purpose** – The purpose of this paper is to explore the current challenges and drivers for data mining in the AEC sector.

**Design/methodology/approach** – Following a comprehensive literature review, the data mining concept was investigated through a workshop with industry experts and academics.

**Findings** – The results showed that the key drivers for using data mining within the AEC sector is associated with the sustainability, process improvement, market intelligence, cost certainty and cost reduction, performance certainty and decision support systems agendas in the sector. As for the processes with the greatest potential for data mining application, design, construction, procurement, forensic analysis, sustainability and energy consumption and reuse of digital components were perceived as the main process areas. While the key challenges were perceived as being, data issues due to the fragmented nature of the construction process, the need for a cultural change, IT systems used in silos, skills requirements and having clearly defined business goals.

**Originality/value** – With the increasing abundance of data, business intelligence and analytics and its related concepts, data mining and Big Data have captured the attention of practitioners and academics for the last 20 years. On the other hand, and despite the growing amount of data in its business context, the AEC sector still lags behind in utilising those concepts in its end products and daily operations with limited research conducted to explore those issues at the sector level. This paper investigates the main opportunities and barriers for data mining in the AEC sector with a practical focus.

**Keywords** Technology, Decision support systems, Information and communication technology (ICT) applications

**Paper type** Research paper

## Introduction and background

Business intelligence and analytics (BI&A) and the related field of Big Data analytics have gained great significance over the past two decades (Chen *et al.*, 2012). Top performing organisations are three times more likely to use sophisticated methods in their use of data analytics than lower performers, making analytics a major differentiating factor (Davenport and Harris, 2007; Hopkins *et al.*, 2010). BI&A is often referred to as the techniques, technologies, systems, practices, methodologies and applications that analyse critical business data to help an enterprise better understand its business and make timely business decision (Chen *et al.*, 2012). According to a report by Gartner (2015), with the mobile and cloud computing technologies, the dynamic BI&A market is undergoing a fundamental shift from highly centralised, IT-led consolidation and standardisation projects to a more mobile, user-centric and interactive styles of analysis and reporting without requiring the users to have static location and IT/data science skills. These business-centric practices and methodologies can be applied to various high-impact applications such as in: e-commerce,

The authors would like to acknowledge the contribution of support of XpertRule (<http://xpertrule.com/>) and TSB for their support for this study.



market intelligence, e-government, healthcare, security, determining human/asset location and manufacturing etc. (Turban *et al.*, 2008; Brown *et al.*, 2011).

A closely related term to BI&A is data-driven decision making (DDD), which refers to the practice of making decisions on the analysis of data rather than purely on intuition, and can translate into increased productivity and market value (Provost and Fawcett, 2013). However, the act of gathering and storing large amounts of data for eventual analysis and decision making is not new. Over the past 20 years, the amount, type and flow of data in many organisations have exponentially increased and is now classified as Big Data (Gantz and Reinsel, 2011; Chen *et al.*, 2014). Moreover, Big Data is accumulating from multiple sources such as systems, sensors and mobile devices, etc., at an alarming velocity, volume and variety to an extent that 90 per cent of the data in the world today has been created in the last two years alone (Zikopoulos *et al.*, 2013; IBM, 2017). This influx of data offers insights for making right decisions at the right time based on efficient aggregation and analysis of internal and external data in organisations.

Given the competitive and complex nature of the construction industry, and the constant transformation in its methodologies and processes, more and more businesses need to rely on fact-based decision making and analysis to guarantee their survival (Witten *et al.*, 2016; Shmueli *et al.*, 2017; Tixier *et al.*, 2017). However, the topic is still novel and the volume of research in the AEC domain, particularly in terms of works investigating the subject at the sector/industry level is limited (Bilal *et al.*, 2016; Ahmed *et al.*, 2017). This paper therefore attempts to develop an understanding of the potential and challenges of data analytics and data mining in the AEC sector and to present a scoping a study, providing a ground for future research in the field.

### **Business intelligence (BI), analytics and data mining: what is the difference?**

BI is an umbrella term that includes the applications, infrastructure and tools, and best practices that help gain access to and analysis of information to improve and optimise decisions and performance (Gartner, 2017). Business analytics is considered a subset of BI; however, there are some differentiating factors which can be understood from Table I.

Table I shows that business analytics is BI that relies on statistical modelling, predictive and forecasting techniques. Davenport (2013) suggests that there are three types of analytics; descriptive, which reports on the past to report “what has happened?”; predictive, which uses models based on past data to predict the future and to determine “what could happen?”; and prescriptive, which uses models to specify optimal behaviours, actions and answer “what should we do?”

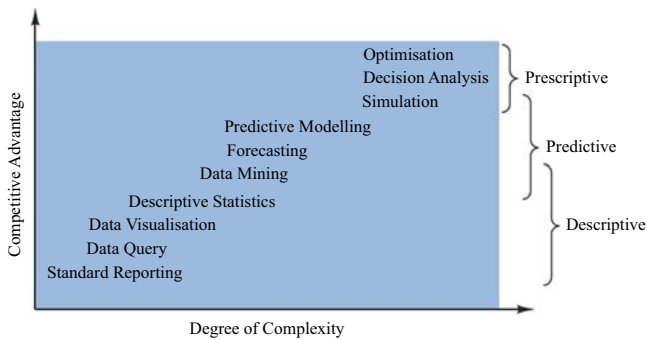
Data mining analyses large quantities of data in order to discover meaningful patterns by using a combination of pattern-recognition rules, statistical rules, as well as rules drawn from machine learning (an area of computer science) (Ledolter, 2013; Larose, 2014). Figure 1 highlights the positioning of data mining on a business analytics spectrum. It can be seen that it mainly falls under predictive analytics where the complexity is higher than descriptive analytics. According to Camm *et al.* (2015), data mining can be classified as:

- supervised learning: defined as a set of tools used for prediction (linear model, logistic regression, linear discriminant analysis, classification trees etc.); and
- unsupervised learning: an exploratory data analysis technique used for identifying groups (e.g. patterns, clusters, association rules, etc.) in the data set of interest.

The ability to embed analytics and optimisation in DDD process promises huge benefits for customers and markets (Davenport, 2013). The benefits have also been demonstrated conclusively in a study conducted on how DDD affects organisations’ performance (Brynjolfsson *et al.*, 2011). The research shows statistically that 1 standard deviation higher

Business intelligence	Business analytics	Reference
A broader concept that comprises of technologies and processes used to understand and analyse business performance Encompasses accessing data, reporting and analytics to provide answer questions about an organisation's business activities Strong association with conventional, technical reporting	A subset of business intelligence Involves rigorous use of data, statistical and quantitative analysis, explanatory and predictive modelling, and fact-based management to drive decisions and actions	Davenport and Harris (2007), Power <i>et al.</i> (2015)
Employs a fixed set of metrics for the evaluation of historical performance and for carrying out future planning Only queries and report generation does not guide actions to be taken for the improvement of the organisation and so lacks decision support system capabilities Develops intelligence through descriptive analytics majorly	Generates greater value and has a more proactive approach towards data than conventional BI Involves use of more dynamic and developed problem solving tools or methods such as optimisation Involves gaining future insights and analysing data to extract meaningful information and actionable decisions which shows its decision support system capabilities Inclination towards automation or prescriptive analytics and improved forecasting models	Davenport and Harris (2007), Laursen and Thorlund (2010) Wang (2014)
Uses structured data often stored in relational database management systems	Can make use of unstructured data that is widely available through various sources such as the web and the internet	Gangotra and Shankar (2016), Laursen and Thorlund (2010) Chen <i>et al.</i> (2012)

**Table I.**  
Distinguishing factors between BI&A



**Figure 1.**  
The spectrum of business analytics

**Source:** Adapted from SAS (Camm *et al.*, 2015)

on the DDD scale is associated with a 4–6 per cent increase in productivity. The research also shows that DDD is correlated with higher return on assets, return on equity, asset utilisation and market value, and the relationship seems to be causal.

**Data and business performance in the AEC sector**

The AEC sector, in general, is characterised by its slow pace of change, low productivity, waste, fragmentation and long established processes and ways of doing business that have not changed over the decades (Behera *et al.*, 2015; Arashpour *et al.*, 2017). Whilst there have been incremental improvements in workflows and processes, there has not been a fundamental change in how the business is done. In contrast, other industry sectors

(e.g. aerospace, automotive, aviation) have undergone paradigm shifts in the way they deliver value to their customers (Crowley, 1998; Green *et al.*, 2005; Bonev *et al.*, 2015). The increasing scrutiny over construction costs by clients, coupled with the increasing pressure on data transparency, has made a strong case for better management of construction data.

Better utilisation of data and analytics is particularly important in early project decision making, which has a huge influence on the downstream life-cycle stages (Bilal *et al.*, 2016). It is expected that in the near future, decision making will become more complex because of increasing demands and expectations, coupled with paucity of funds and fast paced technology growth. In early project stages, the ability to influence the key project parameters is maximum and the cost of any proposed design changes is minimal. Thus, time and effort spent upfront in the planning phase could provide huge dividends throughout the project life cycle.

With many data formats flowing from incompatible software, such as Autodesk's DWG (abbreviation of "drawing"), Bentley's DGN (abbreviation of "design") for engineering activities, Microsoft's DOC/XLS/PPT (document format), Adobe's PDF (document format), RM/MPEG (video format) and JPEG (image format), and constantly increasing data volumes distributed over fragmented ICT systems (i.e. ERP systems) and relational databases of different organisations, data in the AEC sector are typically of the Big Data category (Jiao *et al.*, 2013). Adding to the existing data volumes and complexity, automatically data generating technologies such as laser scanning, advanced photogrammetry, mobile computing and sensors have been fast penetrating into the sector's daily operations (Chen and Kamara, 2011; Razavi and Haas, 2011; Kim *et al.*, 2013). Alongside the technical data interoperability issues, due to the project centric, dynamic and fragmented nature of the sector, the problem of data integration, and subsequently data analytics, throughout the life cycle of a project among multiple collaborative organisations remain unsolved (Bakis *et al.*, 2007; Cerovsek, 2011).

Recently, the building information modelling (BIM) concept has been widely discussed as a single project-life cycle data repository (Howard and Björk, 2008; Babič *et al.*, 2010). Nonetheless, it is still a challenge to completely integrate both design (e.g. geometric and parametric descriptions) and management (e.g. scheduling, monitoring, work assignments) related data onto a singular data representation through the project-life cycle with additional standardisation, interoperability and data sharing issues (Jiao *et al.*, 2013). Ironically, with the improved capabilities of BIM with systems like seismic, lighting and energy simulation in project design, the extension of the use of BIM models from design to construction and maintenance, and the increasing integration of BIM systems with mobile and cloud computing, even bigger data sets to manage have been generated for the sector (Popov *et al.*, 2010; Kim *et al.*, 2011; Redmond *et al.*, 2012). Robust analytics that holistically infer from diverse data resources and those singular project data representations are yet to come.

Given bigger data sets being generated within almost each project life-cycle and fast developing Big Data analytics opportunities, the sector lags behind in obtaining value from those data as an asset and DDD as a managerial competence (Bilal *et al.*, 2016; Ahmed *et al.*, 2017). Organisations in the AEC sector can succeed based on their ability to achieve constant yield from each project. There is tremendous value in turning operational and transactional data "into tactical information" in an industry where profit can be so severely influenced by exceptions (Kamara *et al.*, 2002). BI&A can be used to analyse past performance for planning, identify exceptions for action and trend current performance for forecasting. Many construction executives often make decisions based on their intuitions and subjective judgements, justified by years of experience and expertise (Mohemad *et al.*, 2010). However, today's fierce competition requires precision and process improvements to avoid any, if not all, risks associated with construction financial and project management, which requires BI&A to be an integral part of the sector's business strategy. The rest of this section looks into the potential of data mining in the AEC sector.

### Data mining in the AEC sector

Data mining is the application of specific algorithms or techniques for extracting knowledge patterns from data sets (Fayyad *et al.*, 1996). The patterns discovered must be meaningful in that they lead to some advantage in decision making. Although the traditional way of making decisions in construction organisations is based on chief executives' experience and gut feelings, the fierce market conditions in construction today requires precision and process improvement to avoid any risks associated with financial and project management issues (Mehta, 2010). Being a relatively recent phenomenon for construction, data mining techniques were mostly introduced to the AEC sector at the beginning of the 2000s (Hammad *et al.*, 2013). In the AEC sector, BI&A can be used in the following: analysing data from the generic, firm-wide level to the more specific, individual level; performing "what if" analyses for cash flow, revenue and profit; eliminating cumbersome Excel spread sheets; and integrating information from different financial software resources such as SAP or Tally for a complete view of the business (Mehta, 2010). Specific data mining applications exploiting Artificial neural networks, genetic algorithms, neuro-fuzzy systems, bayesian networks, case based reasoning, text mining, decision trees, clustering and regression on smaller, structured construction data warehouses have been applied and brought a number of benefits to the AEC sector:

- Historical project data can assist AEC professionals in answering specific questions about the business, the performance of interested operations, business trends and what can be done to improve the business and operations in general (Rujirayanyong and Shi, 2006). The practice of inference from that historic data is broadly called data mining, which is the act of extraction of implicit, previously unknown and potentially useful information and knowledge from recorded data sets (Frawley *et al.*, 1992).
- Data mining is also used for optimising structural engineering design (Leu *et al.*, 2001; Tinoco *et al.*, 2011; Gandomi *et al.*, 2016), energy-efficient and sustainable building design (Kim *et al.*, 2011; Naganathan *et al.*, 2016), supporting collaborative design processes (Chiu and Lan, 2005), construction operation simulation (Akhavian and Behzadan, 2013), investment feasibility analysis (Yun and Caldas, 2009), future cost forecasting (Wilmot and Cheng, 2003), analysing building performance (Ahmed *et al.*, 2009; Fan *et al.*, 2015), construction accident analysis (Abdelhamid and Everett, 2000; Huang and Hinze, 2003; Ding *et al.*, 2018), safety clash detection (Tixier *et al.*, 2017), optimising bid selection policies (Art Chaovalitwongse *et al.*, 2011), hierarchical construction document classification (Caldas and Soibelman, 2003), asphalt paving pattern prediction and analysis for transportation projects (Nassar, 2007) and predictive cost overrun and mitigation analysis (Ahiaga-Dagbui and Smith, 2014; Williams and Gong, 2014).
- Data mining also provides useful tools that help to explain how building systems and facilities that were once thought to be completely chaotic have predictable patterns (Peitgen *et al.*, 1992; Fan *et al.*, 2015). A report by Liew and Rosenblatt (2003) revealed that applying data mining techniques could greatly help facility managers to identify crucial maintenance cost patterns and to form preventive maintenance plans. Reffat *et al.* (2006) gave a classification of possible facilities management benefits with their corresponding data mining techniques: unveiling the correlations between facilities service parameters and malfunctions, and some other independent parameters (i.e. seasons or weather conditions), dynamic adjustments of lighting, temperature and air conditioning, prioritising maintenance works, calculating benefits from successful maintenance practices, maintenance resource allocations and better planning in maintenance works.

Having outlined those data mining implementations and possibilities, it should be noted most of the data mining discussions in construction are coming from specific academic

research efforts, which do not illustrate the general extent of the use of data mining in the sector with its drivers and challenges. Previous studies of data mining in the construction industry have mainly focussed on introducing new data mining techniques, comparing between techniques and comparing between data mining techniques and traditional methods (Yun and Caldas, 2009). Also, the amount of data within the sector has become increasingly overwhelming, as well as the growth of many business, government and scientific databases coupled with the data generated by individuals and even inanimate objects (i.e. sensors, mobile and wearable computing) has begun to far outpace our ability to interpret and digest those data holistically with conventional approaches.

Nassar (2007) illustrated some key issues relating to the implementation of data mining in construction:

- Lack of standardisation in the construction industry as to collecting and storing project and company-specific data: this industry fragmentation significantly hinders the uptake of data mining techniques in practice.
- The way in which the information is stored in the construction industry is generally not very organised and patchy. In the research presented by Nassar (2007), for example, various pieces of information had to be collected from bulletins, reports, as well as electronic databases and then re-structured into one database in order to facilitate data mining.
- There is a need for a unified data model for construction data. This unified data model would be similar to the current building product model utilised in the Industry Foundation Classes but would focus primarily on construction specific data. One layer would capture project-specific data such as cost, estimate, schedule and productivity. The second layer would capture company-specific data, such as profitability, bids and bonding capacity. These two layers would be obviously interrelated so that information can be indexed from one layer to another. The third layer would capture industry-specific data such as employment rates, industry-wide productivity rates and financial ratios.

A study by Silwattananusarn and Tuamsuk (2012) looked into the available literature on data mining applications for knowledge management (KM). The study divided knowledge resources into eight groups as to which knowledge object to be stored and manipulated in KM and how data mining aids for healthcare, retailing, financial banking, small and middle business, entrepreneurial science, collaboration and teamwork, general business and construction organisations. The study revealed that only 19 articles containing both “Data Mining” and “Knowledge Management” in their title, keyword or abstract fields had been published between 2007 and 2012. For construction, the authors stated that a large part of enterprise information was available in the form of textual data. This leads to the influence of text mining techniques to handle textual information source for industrial knowledge discovery and management solutions, which have been applied in a very small number of conceptual research efforts in construction so far.

In light of this context, it can be inferred that there is a lack of empirical research exploring the data mining concept for construction (Soibelman and Kim, 2002; Gajzler, 2010; Bilal *et al.*, 2016; Ahmed *et al.*, 2017). Therefore, this paper aims at understanding the current challenges and drivers for data mining in the AEC sector.

### Research methodology

This paper shares the outcome of an industry workshop that was carried to develop a scoping study for an SME IT group (XpertRule) in order to explore the potential and challenges for utilising data mining in the AEC sector and whether these align with what has already been identified from the literature. The focus group consisted of a mixture of

65 participants from the UK construction sector (AEC) and academia with a few participants from IT service providers to the AEC sector. When researchers seek depth or execute scoping studies at industry/sector level on a subject, focus groups are frequently preferred to quickly capture rich insights from relevant individuals, enabling the exploration of novel or complex phenomena (Love *et al.*, 2010; Anvuur *et al.*, 2011; Liamputtong, 2011; Krueger and Casey, 2014; Loosemore, 2014; Loudoun and Townsend, 2017).

Details of the workshop attendees can be seen in Table II. The attendees were identified and invited selectively. Senior- and middle-level managers from prominent organisations in the industry were contacted initially for the workshop. Also, the academic participants were invited by relevance of their research interests in the built environment to the subject area. Contacts details of the participants were obtained from professional organisations, the industry partner's network, as well as from the authors' and their institutes' contact records.

It was anticipated that the level of understanding of the subject among the AEC sector professionals and some higher education sector participants might have been limited. Therefore, four presentations on issues associated with construction data types, data management and DDD in the AEC sector, the data mining and BI&A concepts, their current use in the AEC sector compared to other sectors, their possible future use across the project life-cycle and opportunities facilitated by emerging IT solutions were given to the attendees by subject experts.

After the presentations, it was assumed that the participants had gained a satisfactory understanding of BI, data mining and their applications across different sectors for the scoping research presented in the paper. Following the presentations and initial discussions, the participants' feedback was collated in two steps as explained below. At the end of the workshop, initial findings were presented to the attendees and also, the attendees were contacted within two weeks after the workshop to inform them of the collated findings and to obtain their further comments and suggestions for validation. Some minor comments were received from especially senior managers and academics that were incorporated into the final discussion.

The following steps were taken to analyse the focus group findings (Liamputtong, 2011; Krueger and Casey, 2014): transcribing the written feedback and the discussion voice records of the participants; noting the main ideas emerging for each question;

Sector role	Years of experience in sector			Grand total
	< 10	10–20	> 20	
AEC		22	14	36
Client		3	2	5
Contractor		10	9	19
Designer		4	3	7
Facilities manager		5		5
Higher education	10	4	7	21
Lecturer	2	2	1	5
PhD researcher	6			6
Professor		1	3	4
Research fellow	2			2
Senior lecturer		1	3	4
IT	2	4	2	8
Database		2		2
Software	2		2	4
Consultant		2		2
Grand total	12	30	23	65

**Table II.**  
Details of the  
workshop attendees

---

reviewing, classifying and grouping the recurring ideas; identifying the main themes associated with those ideas; reporting the main findings; and seeking validation and reviewing the findings.

#### *Step 1 – panel discussion*

A panel formed of a senior construction manager, a senior academic in the built environment and a director of an IT company led the discussion about the potential for the use of data mining for BI in the AEC sector.

#### *Step 2 – participants' feedback*

The participants were asked to give written feedback to include three main points in response to the following four questions:

- RQ1. what are the key drivers to use data mining within construction?
- RQ2. which processes within construction hold the greatest potential for data mining applications?
- RQ3. what are the key challenges to application of data mining within construction?
- RQ4. what steps should be taken by the construction industry to drive the uptake of data mining?

The rest of the paper shares the results generated and captured from the above steps.

### **Results and analysis**

#### *Step 1 – panel discussion*

The panel discussion evolved around the attendees' feedback for the four main questions. Based on the discussions, a few generic conclusions were made:

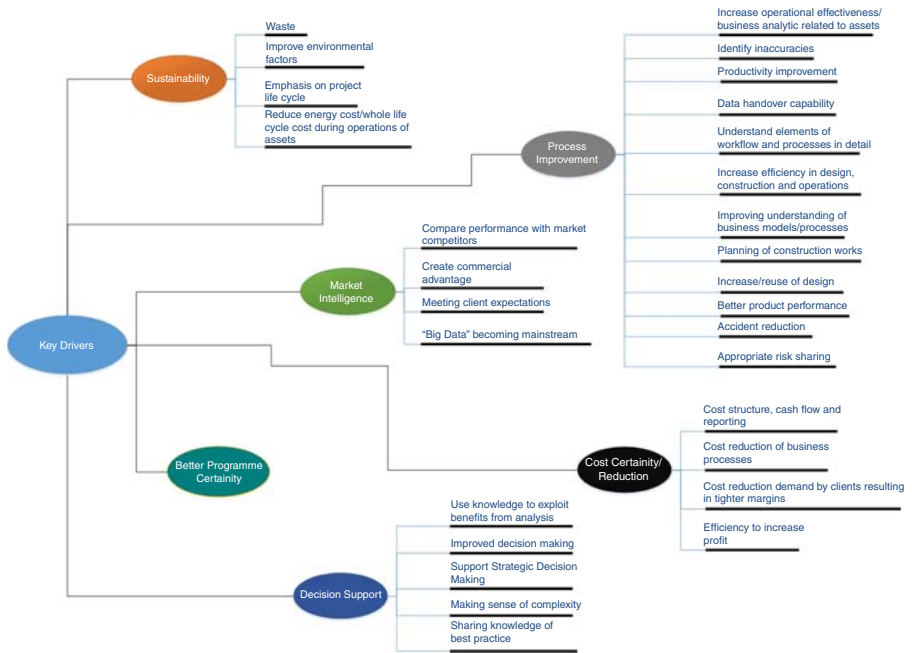
- Data mining and BI&A is an area that is not well explored in construction, and it requires academic research and investigation.
- There is a great potential in using data mining in order to improve the sustainability aspects of projects, reducing project waste, improving processes, improving market intelligence, supporting the decision-making process and improving business certainty.
- Data mining principles and concepts can be applied throughout the construction life cycle from conception to operation.
- The main challenges that face the application of data mining in construction have to do with obtaining the right data, understanding its relevance and importance, the intellectual rights, cultural and trust issues, skills and cultural changes that need leadership to drive this concept.
- There is a potential for construction companies and academics to work together on knowledge transfer schemes.

#### *Step 2 – participants' feedback*

The participants were asked about the key drivers for using data mining in construction. Figure 2 shows the summary of the participants' feedback, whereby six key areas the participants identified are as such:

- (1) Sustainability – where data mining can play an important role in waste management and recycling on construction sites, which helps in responding to the sustainability agenda and managing the percentages of recyclable material and the source of waste.





**Figure 2.**  
Key drivers to use data mining within construction

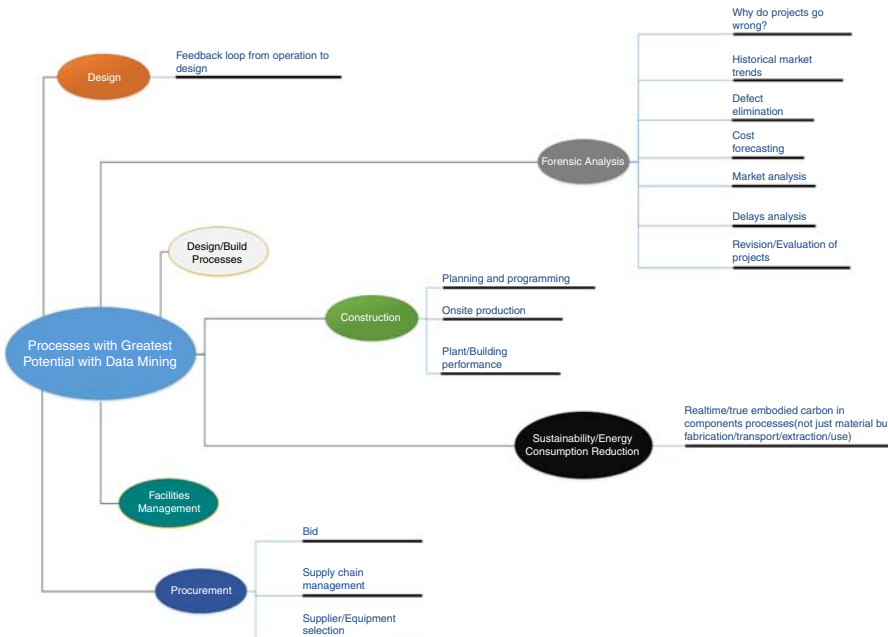
This in return plays an important role in improving the environmental factors around a project such as the percentages of pollutant materials and the level of pollution generated by the construction process. Project generated data can also help predict the time and costs wasted in the life cycle of a project and the pattern of behaviours around these processes to reduce waste.

- (2) Process improvement – the participants’ response under this category covered a wide range of activities to do with the different stages of a project from inception to handover and beyond, whereby project data can help with increasing the operational effectiveness within a building, whether related to the management of the occupancy of the building or energy consumption, identifying the causes of waste during the design and construction process, predicting risk and identifying health and safety measures to improve the project performance.
- (3) Market intelligence – the participants considered data mining and business analytics to play an important role in gaining a competitive advantage through a better understanding of the construction market and specific demand and supply patterns, as well as in serving the client needs to provide a product with a competitive advantage in terms of cost, time and quality.
- (4) Cost certainty and cost reduction – the participants’ responses also aligned with the literature, which puts much emphasis on the use of data mining to predict cost certainty in the business world, and to work out patterns of behaviour for cost reduction, which, in return, helps to improve the project efficiency and the project profit margins.
- (5) Performance certainty – the participants felt that data mining techniques would be useful to improve the cost, time and quality of construction projects through gaining a better understanding of the history of project performance, and gaining a prediction of improved performance certainty for future projects.

- (6) Decision support systems – the participants perceived data mining as a tool to help with decision support systems through the sharing of good practice, in order to understand the roots of problems and to enable the project team to create “what-if-scenarios” to improve the decision making process, making sense of existing project strategies with an attempt for continuous improvement that is driven by data.

The participants were also asked to give their views of the “Processes with the Greatest Potential for Data Mining Applications in Construction”. Figure 3 shows a summary of the participants’ feedback, where design, construction, procurement, forensic analysis, sustainability and energy consumption and the reuse of digital components were perceived as the main process areas:

- Design – the Participants indicated that data generated from the design feedback loop gives architects and clients a good insight into the lessons learnt from previous projects at the design stage, and to define patterns of behaviours at this stage.
- Construction – the majority of the participants agreed that there is a lot of data generated from the construction stage particularly at the planning and programming stage, where data mining can analyse the time and cost efficiencies and efficiencies of different stages of projects to help enhance the performance of on-going, as well as future projects. In addition, the participants also saw great potential for data mining to improve the off-site manufacturing processes, where the same principles apply, as well as data on plant and machinery performance on construction sites.
- Forensic analysis – the participants gave a broad range of factors that could benefit from data analytics and data mining techniques in order to develop an understanding of the general trends as to why project are delayed, project defects’ causes and effects, the market trends in the AEC sector over the years in relation to the supply chain and the client requirements.

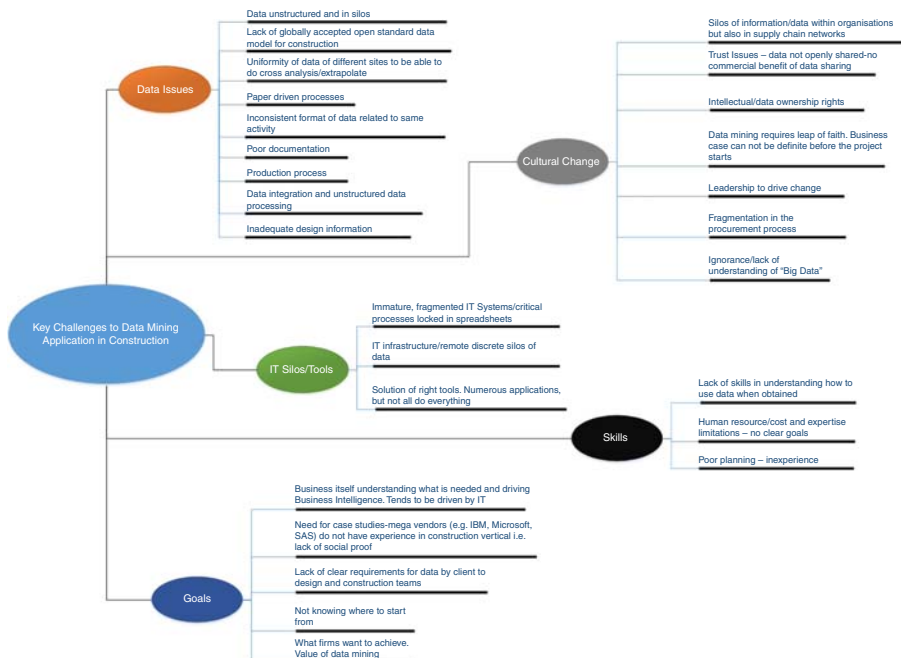


**Figure 3.** Processes with the greatest potential for data mining applications in construction

- Sustainability and energy consumption – the participants saw real potential for extracting data to detect patterns of behaviour through real time and truly embedded carbon processes and components such as the fabrication, transportation and extraction of construction materials and building components.
- Reuse of digital components – the participants recognise that there is a great potential in using data mining techniques, if the existing project digital materials, such as documents, images, designs, programmes, etc., are made use of to extract the relevant data.

The participants were asked about the key challenges that face the application of data mining within the AEC sector. The result in Figure 4 groups these challenges as being:

- Data generation issues – much of the data-related challenges evolved around the availability of massive data sets generated throughout the life cycle of a construction project. The data are often times in a paper based format and/or available in silos and in an unstructured manner.
- Data formatting issues – the participants perceived the attainment of data as one of the main challenges that face the AEC sector due to the fragmented nature of its processes, which forces some of the stakeholders to the projects to work in silos, while the actual data are either available in paper format (not on a shared drive where everyone can access), in inconsistent formats, or is unstructured, poorly and inconsistently documented and not easy to access. The participants also drew attention to the lack of an existing standard or a uniform model for reporting and storing information, particularly when it comes to work generated on different construction sites.



**Figure 4.**  
Key challenges to the application of data mining within construction

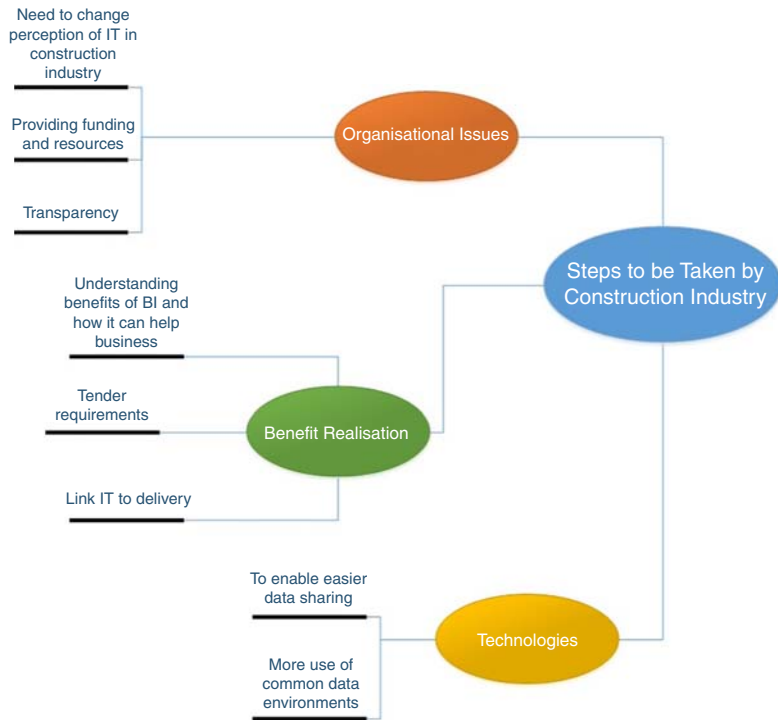
- Need for cultural changes – the participants clearly recognised the need for cultural changes in the way people work, and share information openly for the benefit of the business itself, where a certain level of trust needs to be achieved with good leadership to drive this change. Leadership would also mean training people to show the value in sharing project data, and investing in IT and infrastructure systems. There also needs to be some clear protocols around data sharing rights and intellectual property rights, while not many construction companies have tackled these issues, particularly where much of the data are generated through the design and procurement process.
- IT tools used in silos – the participants emphasised the importance of having an appropriate IT infrastructure that is capable of coping with different file formats and sharing, distributing and updating information that sit in silos of spread sheets. Therefore, having an IT structure that helps to deal and interpolate the data is key.
- Skills requirements – another interesting set of challenges that was highlighted by the participants evolves around the skills required to understand how the data could be used when generated, and how to plan ahead to store the data so that it makes sense when it is generated, and for there to be a clear differentiation between automated systems that store and generate data, and systems that produce meaningful data to serve the BI.
- Goals – the participants agreed that one of the main challenges is not knowing where to start from in terms of adopting data mining techniques, and that construction organisations need to have defined goals and business objectives to be achieved from the data mining with a realised value. Additionally, having clear requirement strategies from different stakeholders, such as the design and construction teams, is of critical importance. Therefore, to drive the value of data mining, the participants suggested that sharing good case studies from different software vendors would be useful.

Finally, the participants were asked about the “Steps to be taken by the Construction Industry to Drive the uptake of Data Mining”. The results can be seen in Figure 5. The participant’s responses were categorised as follows:

- Organisational issues – this category mainly focussed on the organisational mindsets, and the need to change the perceptions in terms of using IT and for it not only to be a communication tool but for it to be a mean for BI. In alignment with the previous points made, the participants also felt that openness and transparency (which are the resultants of trust) are key to providing and sharing valuable data, and for this to happen, sufficient resources and investments are required to promote good practices that will reap the long-term benefits.
- Benefit realisation – again, the participants emphasised the need for organisations to demonstrate the benefits that can be realised through different processes, where tendering was given as an example, as that much of the data generated from this process can help with understanding the probability of success or failure of project prospects.
- Technologies – the participants agreed that changing the mindset to an understanding that the technology is there to enable data sharing through transparent and open environments is a step forward towards integrating and gathering the relevant data for BI.

### Summary and conclusions

BI&A enables DDD, which refers to the practice of basing decisions on the analysis of data rather than purely on intuition, that can translate into an increase in productivity, higher



**Figure 5.**  
Steps to be taken by  
the construction  
industry to drive  
uptake of data mining

return on assets, return on equity, better asset and resource utilisation, and higher market value. This paper argued that although DDD has been widely utilised in different sectors and on the rise in the business world, beyond some specific cases or research investigations demonstrating the use of some specific data mining techniques, its utilisation and penetration have been somewhat limited in the AEC sector. The findings are in accord with the recent sector level investigations by Bilal *et al.* (2016) and Ahmed *et al.* (2017). The wider applications for the sector can lead to data-driven predictions in project KPIs, such as cash flow, and how projects will generate or consume cash, liquidity of project assets, labour productivity, factors that cause project and margin variance, rectifying time and cost defects and improving project deficiencies, and better predictions in asset whole-life performance. Therefore, to further identify challenges and drivers for using data mining (DDD) and the steps that should be taken by the AEC industry to realise this potential, this paper shared the results of an industry workshop which brought together 65 academics and practitioners to explore these issues.

The results showed that there needs to be a cultural change in the way the sector works and its mindset. Organisation leaders will play an important role in this by a top-down approach, investing in technologies that bring together valuable data generated from different sites and fragmented processes, investing in awareness raising and training, and employing people with necessary skill sets (Ofori, 2008). However, it should be noted that the use of those integrating technologies by different project stakeholders should be facilitated as well with relational project delivery structures and commercial configurations that enable data sharing and transparency between the stakeholders (Porwal and Hewage, 2013).

The results also showed that the key drivers in using data mining within the AEC sector are related to the sustainability, process improvement, market intelligence, cost certainty and cost reduction, performance certainty and decision support system agendas in the sector. As for the processes with the greatest potential for data mining applications, design, construction, procurement, forensic analysis, sustainability and energy consumption and reuse of digital components were perceived as the main process areas. While the key challenges were perceived as being data issues due to the fragmented nature of the construction process, the need for a cultural change, IT systems used in silos, skills requirements and having clearly defined business goals.

## References

- Abdelhamid, T.S. and Everett, J.G. (2000), "Identifying root causes of construction accidents", *Journal of Construction Engineering and Management*, Vol. 126 No. 1, pp. 52-60.
- Ahiaga-Dagbui, D.D. and Smith, S.D. (2014), "Dealing with construction cost overruns using data mining", *Construction Management and Economics*, Vol. 32 Nos 7-8, pp. 682-694.
- Ahmed, A., Ploennigs, J., Gao, Y. and Menzel, K. (2009), "Analyze building performance data for energy-efficient building operation", *Proceedings of CIB W 26th W78 Conference on Information Technology in Construction, International Council for Building Research Studies and Documentation*, Vol. 78, Istanbul, 1-3 October.
- Ahmed, V., Tezel, A., Aziz, Z. and Sibley, M. (2017), "The future of Big Data in facilities management: opportunities and challenges", *Facilities*, Vol. 35 Nos 13-14, pp. 725-745.
- Akhavian, R. and Behzadan, A.H. (2013), "Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining", *Journal of Construction Engineering and Management*, Vol. 139 No. 11, p. 04013021.
- Anvuur, A.M., Kumaraswamy, M.M. and Mahesh, G. (2011), "Building 'relationally integrated value networks' (RIVANS)", *Engineering, Construction and Architectural Management*, Vol. 18 No. 1, pp. 102-120.
- Arashpour, M., Abbasi, B., Arashpour, M., Hosseini, M.R. and Yang, R. (2017), "Integrated management of on-site, coordination and off-site uncertainty: theorizing risk analysis within a hybrid project setting", *International Journal of Project Management*, Vol. 35 No. 4, pp. 647-655.
- Art Chaovalitwongse, W., Wang, W., Williams, T.P. and Chaovalitwongse, P. (2011), "Data mining framework to optimize the bid selection policy for competitively bid highway construction projects", *Journal of Construction Engineering and Management*, Vol. 138 No. 2, pp. 277-286.
- Babič, N.Č., Podbreznik, P. and Rebolj, D. (2010), "Integrating resource production and construction using BIM", *Automation in Construction*, Vol. 19 No. 5, pp. 539-543.
- Bakis, N., Aouad, G. and Kagioglou, M. (2007), "Towards distributed product data sharing environments – progress so far and future challenges", *Automation in Construction*, Vol. 16 No. 5, pp. 586-595.
- Behera, P., Mohanty, R.P. and Prakash, A. (2015), "Understanding construction supply chain management", *Production Planning & Control*, Vol. 26 No. 16, pp. 1332-1350.
- 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*, Vol. 30 No. 3, pp. 500-521.
- Bonev, M., Wörösch, M. and Hvam, L. (2015), "Utilizing platforms in industrialized construction: a case study of a precast manufacturer", *Construction Innovation*, Vol. 15 No. 1, pp. 84-106.
- Brown, B., Chui, M. and Manyika, J. (2011), "Are you ready for the era of 'Big Data'?", *McKinsey Quarterly*, October, 12pp., available at: <https://pdfs.semanticscholar.org/bb36/9ffa765e3df25f5b9ab3bfba2bd809f48fac.pdf> (accessed 29 February 2016).

- Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011), "Strength in numbers: how does data-driven decision making affect firm performance", *Proceedings of ICIS 2011*, available at: <http://dx.doi.org/10.2139/ssrn.1819486> (accessed 29 February 2016).
- Caldas, C.H. and Soibelman, L. (2003), "Automating hierarchical document classification for construction management information systems", *Automation in Construction*, Vol. 12 No. 4, pp. 395-406.
- Camm, J., Cochran, J., Fry, M., Ohlmann, J., Anderson, D., Sweeney, D. and Williams, T. (2015), *Essentials of Business Analytics*, Cengage.
- Cerovsek, T. (2011), "A review and outlook for a 'building information model' (BIM): a multi-standpoint framework for technological development", *Advanced Engineering Informatics*, Vol. 25 No. 2, pp. 224-244.
- Chen, H., Chiang, R.H. and Storey, V.C. (2012), "Business intelligence and analytics: from Big Data to big impact", *MIS Quarterly*, Vol. 36 No. 4, pp. 1165-1188.
- Chen, M., Mao, S. and Liu, Y. (2014), "Big data: a survey", *Mobile Networks and Applications*, Vol. 19 No. 2, pp. 171-209.
- Chen, Y. and Kamara, J.M. (2011), "A framework for using mobile computing for information management on construction sites", *Automation in Construction*, Vol. 20 No. 7, pp. 776-788.
- Chiu, M.L. and Lan, J.H. (2005), "Information and IN-formation: information mining for supporting collaborative design", *Automation in Construction*, Vol. 14 No. 2, pp. 197-205.
- Crowley, A. (1998), "Construction as a manufacturing process: lessons from the automotive industry", *Computers & Structures*, Vol. 67 No. 5, pp. 389-400.
- Davenport, T.H. (2013), "Analytics 3.0", *Harvard Business Review*, Vol. 91, pp. 64-72.
- Davenport, T. and Harris, J. (2007), *Competing on Analytics*, Harvard Business School Press, Boston, MA, p. 46.
- Davenport, T., D'Ignazio, B. and Leifer, C. (2013), *The Complete Guide to Business Analytics (Collection)*, FT Press, Upper Saddle River, NJ.
- Ding, L., Fang, W., Luo, H., Love, P.E., Zhong, B. and Ouyang, X. (2018), "A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory", *Automation in Construction*, Vol. 86, pp. 118-124.
- Fan, C., Xiao, F. and Yan, C. (2015), "A framework for knowledge discovery in massive building automation data and its application in building diagnostics", *Automation in Construction*, Vol. 50, pp. 81-90.
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From data mining to knowledge discovery in databases", *AI Magazine*, Vol. 17 No. 3, p. 37.
- Frawley, W., Piatetsky-Shapiro, G. and Matheus, C. (1992), "Knowledge discovery in databases: an overview", *AI Magazine*, Vol. 13 No. 3, pp. 57-70.
- Gajzler, M. (2010), "Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry", *Technological and Economic Development of Economy*, Vol. 16 No. 2, pp. 219-232.
- Gandomi, A.H., Sajedi, S., Kiani, B. and Huang, Q. (2016), "Genetic programming for experimental Big Data mining: a case study on concrete creep formulation", *Automation in Construction*, Vol. 70, pp. 89-97.
- Gangotra, A. and Shankar, R. (2016), "Strategies in managing risks in the adoption of business analytics practices", *Journal of Enterprise Information Management*, Vol. 29 No. 3, pp. 374-399.
- Gantz, J. and Reinsel, D. (2011), "Extracting value from chaos", *IDC Review*, Vol. 1142, pp. 1-12.
- Gartner (2015), "Magic quadrant for business intelligence and analytics platforms", available at: [www.gartner.com/doc/reprints?id=1-2AJGAKH&ct=150225&st=sb](http://www.gartner.com/doc/reprints?id=1-2AJGAKH&ct=150225&st=sb) (accessed 29 February 2016).

- Gartner (2017), "Gartner IT glossary: business intelligence (BI)", available at: [www.gartner.com/it-glossary/business-intelligence-bi/](http://www.gartner.com/it-glossary/business-intelligence-bi/) (accessed 15 December 2017).
- Green, S.D., Fernie, S. and Weller, S. (2005), "Making sense of supply chain management: a comparative study of aerospace and construction", *Construction Management and Economics*, Vol. 23 No. 6, pp. 579-593.
- Hammad, A., AbouRizk, S. and Mohamed, Y. (2013), "Application of KDD techniques to extract useful knowledge from labor resources data in industrial construction projects", *Journal of Management in Engineering*, Vol. 30 No. 6, p. 05014011.
- Hopkins, M.S., LaValle, S., Balboni, F., Kruschwitz, N. and Shockley, R. (2010), "10 data points: information and analytics at work", *MIT Sloan Management Review*, Vol. 52 No. 1, pp. 27-31.
- Howard, R. and Björk, B.C. (2008), "Building information modelling – experts' views on standardisation and industry deployment", *Advanced Engineering Informatics*, Vol. 22 No. 2, pp. 271-280.
- Huang, X. and Hinze, J. (2003), "Analysis of construction worker fall accidents", *Journal of Construction Engineering and Management*, Vol. 129 No. 3, pp. 262-271.
- IBM (2017), "What is Big Data?", available at: [www.ibm.com/big-data/us/en/](http://www.ibm.com/big-data/us/en/) (accessed 15 December 2017).
- Jiao, Y., Wang, Y., Zhang, S., Li, Y., Yang, B. and Yuan, L. (2013), "A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: integrating BIMs and SN", *Advanced Engineering Informatics*, Vol. 27 No. 2, pp. 173-188.
- Kamara, J.M., Augenbroe, G., Anumba, C.J. and Carrillo, P.M. (2002), "Knowledge management in the architecture, engineering and construction industry", *Construction Innovation*, Vol. 2 No. 1, pp. 53-67.
- Kim, C., Son, H. and Kim, C. (2013), "Automated construction progress measurement using a 4D building information model and 3D data", *Automation in Construction*, Vol. 31, pp. 75-82.
- Kim, H., Stumpf, A. and Kim, W. (2011), "Analysis of an energy efficient building design through data mining approach", *Automation in Construction*, Vol. 20 No. 1, pp. 37-43.
- Krueger, R.A. and Casey, M.A. (2014), *Focus Groups: A Practical Guide for Applied Research*, Sage Publication.
- Larose, D.T. (2014), *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Hoboken, NJ.
- Laursen, G. and Thorlund, J. (2010), *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*, John Wiley & Sons, Hoboken, NJ.
- Ledolter, J. (2013), *Data Mining and Business Analytics with R*, Wiley.
- Leu, S.S., Chen, C.N. and Chang, S.L. (2001), "Data mining for tunnel support stability: neural network approach", *Automation in Construction*, Vol. 10 No. 4, pp. 429-441.
- Liamputtong, P. (2011), *Focus Group Methodology: Principle and Practice*, Sage Publications.
- Liew, M.P. and Rosenblatt, J. (2003), *Using Data Mining Techniques for Improving Building Life Cycle*, CRC for Construction Innovation, Brisbane.
- Loosemore, M. (2014), "Improving construction productivity: a subcontractor's perspective", *Engineering, Construction and Architectural Management*, Vol. 21 No. 3, pp. 245-260.
- Loudoun, R. and Townsend, K. (2017), "Implementing health promotion programs in the Australian construction industry: levers and agents for change", *Engineering, Construction and Architectural Management*, Vol. 24 No. 2, pp. 260-274.
- Love, P., Davis, P., Ellis, J. and On Cheung, S. (2010), "Dispute causation: identification of pathogenic influences in construction", *Engineering, Construction and Architectural Management*, Vol. 17 No. 4, pp. 404-423.
- Mehta, S. (2010), "Business intelligence for construction industry", *The Master Builder*, September, pp. 196-198.



- Mohemad, R., Hamdan, A.R., Othman, Z.A. and Noor, N.M.M. (2010), "Decision support systems (DSS) in construction tendering processes", *IJCSI International Journal of Computer Science Issues*, Vol. 7 No. 2, pp. 35-45.
- Naganathan, H., Chong, W.O. and Chen, X. (2016), "Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches", *Automation in Construction*, Vol. 72, pp. 187-194.
- Nassar, K. (2007), "Application of data-mining to state transportation agencies' projects databases", *ITcon*, Vol. 12, pp. 139-149.
- Ofori, G. (2008), "Leadership for future construction industry: agenda for authentic leadership", *International Journal of Project Management*, Vol. 26 No. 6, pp. 620-630.
- Peitgen, H.O., Jurgens, H. and Saupe, D. (1992), *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, New York, NY.
- Popov, V., Juocevicius, V., Migilinskas, D., Ustinovichius, L. and Mikalauskas, S. (2010), "The use of a virtual building design and construction model for developing an effective project concept in 5D environment", *Automation in Construction*, Vol. 19 No. 3, pp. 357-367.
- Porwal, A. and Hewage, K.N. (2013), "Building information modeling (BIM) partnering framework for public construction projects", *Automation in Construction*, Vol. 31, pp. 204-214.
- Power, D.J., Sharda, R. and Burstein, F. (2015), *Decision Support Systems*, John Wiley & Sons.
- Provost, F. and Fawcett, T. (2013), "Data science and its relationship to Big Data and data-driven decision making", *Big Data*, Vol. 1 No. 1, pp. 51-59.
- Razavi, S.N. and Haas, C.T. (2011), "Using reference RFID tags for calibrating the estimated locations of construction materials", *Automation in Construction*, Vol. 20 No. 6, pp. 677-685.
- Redmond, A., Hore, A., Alshawi, M. and West, R. (2012), "Exploring how information exchanges can be enhanced through Cloud BIM", *Automation in Construction*, Vol. 24, pp. 175-183.
- Reffat, R.M., Gero, J.S. and Peng, W. (2006), "Improving the management of building life cycle: a data mining approach", *CRC Research Conference, Brisbane*.
- Rujirayanyong, T. and Shi, J.J. (2006), "A project-oriented data warehouse for construction", *Automation in Construction*, Vol. 15 No. 6, pp. 800-807.
- Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R. and Lichtendahl, K.C. Jr (2017), *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, John Wiley & Sons.
- Silwattananusarn, T. and Tuamsuk, K. (2012), "Data mining and its applications for knowledge management: a literature review from 2007 to 2012", *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, Vol. 2 No. 5, pp. 13-24.
- Soibelman, L. and Kim, H. (2002), "Data preparation process for construction knowledge generation through knowledge discovery in databases", *Journal of Computing in Civil Engineering*, Vol. 16 No. 1, pp. 39-48.
- Tinoco, J., Correia, A.G. and Cortez, P. (2011), "Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time", *Construction and Building Materials*, Vol. 25 No. 3, pp. 1257-1262.
- Tixier, A.J.P., Hollowell, M.R., Rajagopalan, B. and Bowman, D. (2017), "Construction safety clash detection: identifying safety incompatibilities among fundamental attributes using data mining", *Automation in Construction*, Vol. 74, pp. 39-54.
- Turban, E., Sharda, R., Aronson, J.E. and King, D. (2008), *Business Intelligence: A Managerial Approach*, Prentice Hall.
- Wang, J. (2014), *Encyclopaedia of Business Analytics and Optimization*, Business Science Reference, Hershey, PA.
- Williams, T.P. and Gong, J. (2014), "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers", *Automation in Construction*, Vol. 43, pp. 23-29.

- Wilmot, C.G. and Cheng, G. (2003), "Estimating future highway construction costs", *Journal of Construction Engineering and Management*, Vol. 129 No. 3, pp. 272-279.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.
- Yun, S. and Caldas, C.H. (2009), "Analysing decision variables that influence preliminary feasibility studies using data mining techniques", *Construction Management and Economics*, Vol. 27 No. 1, pp. 73-87.
- Zikopoulos, P., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D. and Giles, J. (2013), *Harness the Power of Big Data*, McGraw-Hill.

**Corresponding author**

Algan Tezel can be contacted at: [algantezel@gmail.com](mailto:algantezel@gmail.com)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)