ORIGINAL PAPER

Business intelligence and analytics to support management in construction: a systematic literature review

Anderson Brunheira Lopes[iD],1 and Clodis Boscarioli[iD],1

1Universidade Estadual do Oeste do Paraná (UNIOESTE) - Campus de Foz do Iguaçu

*andersonlopes.foz@gmail.com; boscarioli@gmail.com

Abstract

Management is essential to meet the requirements defined in a project, such as business objectives and stakeholder expectations. The computational tools in the Business Intelligence and Analytics (BIA) category have great potential to contribute to management, providing important management information about the business. Tools of this type are widely used in the most varied sectors of the industry, but in the construction industry, the scenario is different, with much to progress. Therefore, the present work presents a survey of Business Intelligence and Analytics tools applicable to the construction sector and its possible applications, in order to present options for improving management in these organizations, based on evidence obtained in studies carried out. To this end, a systematic review of the literature was carried out, which analyzed 1407 articles from six databases, where several applications were identified, the most relevant in the area of cost management, budgeting and work safety. With that, it can be concluded that there are several BIA tools for construction, with different applications. Most software was developed for each case studied due to the unique characteristics of the construction sector. The large-scale adoption of the tools involves cooperation between companies, professional associations and universities. It verifies limitations in the research regarding the characterization of the companies, due to the absence of this data in the analyzed articles. It suggests that the challenges of implementing technologies and the verified limitations should be addressed in future studies.

Keywords: Business Intelligence and Analytics; Construction; Systematic Literature Review

Resumo

A gestão é essencial para atender aos requisitos definidos em um projeto, como objetivos de negócios e expectativas das partes interessadas. As ferramentas computacionais na categoria Business Intelligence and Analytics (BIA) têm grande potencial para contribuir com a gestão, fornecendo informações gerenciais importantes sobre o negócio. Ferramentas desse tipo são amplamente utilizadas nos mais variados setores da indústria, mas na indústria da construção civil, o cenário é diferente, com muito a avançar. Portanto, o presente trabalho apresenta um levantamento das ferramentas de Business Intelligence e Analytics aplicáveis ao setor da construção civil e suas possíveis aplicações, de forma a apresentar opções para a melhoria da gestão nessas organizações, com base em evidências obtidas em estudos realizados. Para tanto, foi realizada uma revisão sistemática da literatura, que analisou 1407 artigos de seis bases de dados, onde foram identificadas diversas aplicações, sendo as mais relevantes na área de gestão de custos, elaboração de orçamento e segurança do trabalho. Com isso, pode-se concluir que existem diversas ferramentas de BIA para a construção civil, com diferentes aplicações. A maioria dos software foram desenvolvidos para cada caso estudado devido às características únicas do setor da construção. A adoção em larga escala das ferramentas passe pela cooperação entre empresas, entidades de classe e universidades. Verifica limitações na pesquisa quanto à caracterização das empresas, devido à ausência desses dados nos artigos analisados. Sugere que os desafios de implementação das tecnologias e as limitações verificadas sejam abordados em estudos futuros.

Palavras-Chave: Business Intelligence e Analytics; Construção Civil; Revisão sistemática da literatura

1 Introduction

An efficient management is essential to comply with the requirements initially stipulated in a project, such as the business goals and the expectations of stakeholders. On the other hand, inefficient management can result in missed deadlines, budget overruns, poor quality and rework (PMBOK, 2017).

An important tool to aid in the management is data analysis, which provides information to support decision-making through analysis of records of the existing processes in organizations, providing important business management information. In this context, there are different data analysis tools, and the most used are based on concepts of data warehouse (DW), OLAP (Online Analytical Processing), dashboards, data mining and visual data analysis. The set of those technologies for data analysis is part of the umbrella concept called Business Intelligence and Analytics (BIA), of interest in this research.

BIA tools are known as techniques, technologies, systems, practices, methodologies and applications capable of providing, through the analysis of business data, a better understanding to organizations, enabling the decision at the right time (Chen et al., 2012)

Such tools are widely used in various industry sectors. In the construction industry, focus of this work, the scenario is different, since there is a considerable gap between construction companies and those of other sectors of industry in the search for new technologies and improvement of their management principles. Therefore, an analysis of the potential benefits obtained with the deployment of analytic technologies is pertinent, and can improve the performance of management of those organizations, which usually have archaic procedures and information systems that fail to provide resources to support the strategic decision making.

Faced with those facts and seeking to contribute to the development of construction companies, this paper seeks to investigate the implementation of the tools of Business Intelligence and Analytics to support decision-making in construction, in order to display the usage trends and enable the development of future work, in order to present options for companies in the sector, based on evidence obtained in experiences of other companies and experiments already carried out.

The paper is structured as follows: Section 2 presents the definitions of BIA and the technologies that comprise this concept. Section 3 presents a systematic literature review, which sought to answer the research questions, whereas Section 4 presents the our conclusions and perspectives.

2 Business Intelligence and Analytics applications to support construction management

BIA tools are those that seek to make the management data available in useful organizations in the decision-making process and are based on concepts of DW, OLAP, dashboards, data mining and visual data analysis. It is

a recent term, whose definitions have slight differences between researchers in the area (Sharda et al., 2014).

The concept of Business Intelligence (BI), which integrates the acronym BIA is an umbrella term that encompasses architectures, tools, databases, analytical tools, applications and methodologies to provide interactive access to data (sometimes in real time, enabling data manipulation, providing valuable insights for managers to make the best decisions (Turban et al., 2009). Basically, one may say that the BI process is based on the transformation of data into information that becomes decisions, and finally actions (Sharda et al., 2014). Also according to these authors, a BI system is composed of four main elements: data warehouse, tools of data analysis and mining, BPM (business performance management) and an visual and interactive interface such as a dashboard.

Many researchers use the term analytics instead of the term BI (Sharda et al., 2014). Even though many authors have defined the terms a little differently, some see the term analytics as the process of development of actions based on insights generated by historical data, focusing on the predictive aspect, differing, in part, from the original concept of BI, which focused on understanding what had happened and what was happening in organizations.

One can say that, as technology advanced, data analyses began to generate, for data mining, statistical tools and models of visualization, an overview of the behavior of all the available data of the company. With the perspective of the general behavior of the data and the support of analytical tools, it became possible to find hidden patterns, with the power to generate predictions of future scenarios, generating new insights. In the 1980's, the tools focused on reporting what had happened through statistical reports of business data, to, in the next decade, evolve to an analysis of the causes of events, through OLAP. In the 2000's, technology has evolved, allowing real-time data monitoring that could be used by executives, managers, and even operational employees, through dashboards, to improve the decision-making process. Ten years later, with the development of the concept of Big Data, the focus was on the prediction and the question to be answered was "what will happen?".

With this, the concept of Business Intelligence increasingly tends to use analytical tools, focusing on not only understanding what happened or better viewing what is going on in companies, but also enabling the anticipation of actions, in order to maximize future results, based on predictions and prescriptions. With this, the present study adopted the term BIA to describe the concept of study.

The technology known as data warehouse is one of the pillars of the concept of BIA and also the oldest of them, dated 1990 (Sharda et al., 2014). In a simple definition, we describe this technology as a repository that stores pre-processed and non-volatile data, arising from multiple - and probably heterogeneous - databases (Chau, Anson, Ying and Jianping, 2003). It can be defined as:

> "A set of data produced to support decision-making; it is also a repository of current and historical data of possible interest to managers across the organization. The data are usually structured so that they are available in a format ready for the analytical processing activities (e.g. OLAP, data

mining, look ups, generation of reports, and other decision-support applications). Therefore, a data warehouse is a collection of data driven by subject, integrated, variable in time and non-volatile, which provides support for the decision-making process of management"(Turban et al., 2009, p. 57).

Sharda et al. (2014) describe data warehouse as a repository that stores data of potential interest, current and historical, usually structured so that they are available to be used in analytical activities, data mining and reports. Data warehouses contain sub groups of data, called data marts, which can be defined as "a subset of a data warehouse, which usually consists of a single thematic area" (Turban et al., 2009, p. 58).

Before being loaded into the data warehouse, the data are extracted and processed, to then be loaded in DW in a process known as ETL (extraction, transform and load) (Panwar and Bhatnagar, 2019). This process is one of the most important from the data warehouse and its development generally consumes 70% of the development time of the solution (Sharda et al., 2014). The ETL process begins with the extraction of data from one or more databases, such as spreadsheets, e-mail, database, among others. After the extraction, there is the processing, which consists of converting the data extracted from its prior form to the new format in the data warehouse. Finally, the processed data are loaded (Sharda et al., 2014). The big difference from the data warehouse in relation to conventional databases in providing data is its structure, which is based on two-dimensional modeling. This type of modeling has two main models, called star schema and snow flake schema, being the first one simpler, containing only one central fact table, surrounded by several tables of dimension (Sharda et al., 2014, Rehman and Soomro, 2020).

The fact table usually has a large number of rows, which contain the facts to be observed in the reports, such as the volume of sales, profit margins, costs of production, among other. Dimension tables contain the attributes of the data contained in the fact table's columns, such as places of sale, temporal range of sales, sold brands, main buyers, among others (Sharda et al., 2014). With data structured in this way, this repository is used as the data source for the OLAP technology with intuitive interface, where the user can observe the data by diverse perspectives, having greater chances of obtaining insights to make better decisions based on information Ahmad et al. (2004).

The OLAP data structure is based on a cube concept, which allows rapid data visualizations. The cube structure allows data to be analyzed almost instantaneously (Sharda et al., 2014), and are often utilized in the decision making process (Ribeiro et al., 2020). Using OLAP, the user can navigate through specific data, as well as historical data, changing the prospects of visualization (Sharda et al., 2014). In a more current concept, it can be defined as the set of technologies that enable rapid data processing, presenting them generally as a multidimensional cube (Konikov, 2018). Using OLAP allows making instant look ups, analyzing slices from the cube, according to the desired perspective. It is still possible to select sub groups of data, allowing, for example, the analysis of sales of a certain product in a city, belonging to a region.

The operations that can be used in an OLAP model are slice, where data are selected from a single face of the cube, dice, which allows extracting from a sub cube the selection of two or more faces of the cube, as well as the operations of drill down, roll up, pivot and rotation, which allow examining the data with greater accuracy or generalization, as well as by different perspectives.

In the same line of OLAP, but with different and more intuitive visual, there is the tool known as dashboard, which can be defined as a visual presentation of critical data, focusing on aiding executives to make decisions in seconds (Sharda et al., 2014). The dashboards are known as the front end, term that describes the screen developed for the end user of the application, where the interaction with the system is simple. One may also say that:

"Dashboards provide visual displays of important information, consolidated and organized into a single screen to be absorbed quickly and easily. From the dashboard, it is easy to see, for example, if all Key Performance Indicators (KPI) are within the expected performance standards" (Ferreira et al., 2016, p. 42).

Most of the times, the dashboards "display quantitative measures about what is happening. To accelerate the assimilation of numbers, which must be placed in context. This can be done by comparing the numbers relevant to other reference numbers" (Turban et al., 2009, Abou-Ibrahim and Hamzeh, 2020, p. 228), so that one can observe if the number is good or bad. A practical example is the indicator of estimated cost for a particular work step, in comparison to the cost carried out up to the moment of the lookup. It is easy to observe if the process has a tendency to meet the planned cost or to burst the budget, demanding urgent actions. In general, a dashboard has the following characteristics:

"1. They use visual components (like bars of performance, indicators, gauges, lights) to highlight immediately the data and exceptions that require action; 2. They are transparent to the user, i.e., require minimal training and are extremely easy to use; 3. Combine data from multiple systems and form a single vision, summarized and unified; 4. Enable the achievement of drill down in data sources or reports, providing more details about the comparative and evaluative context that is behind; 5. Feature a dynamic and practical vision with occasional data updates, which allows the end user to be updated about any recent changes in business; and have few or no customized code to implement, deploy and maintain." (Turban et al., 2009, p. 229).

Finally, another BIA tool is data mining. This technology arose due to the exponential increased volume of existing data, from the technological development and the new habits of humans, increasingly connected. With this, the data became ubiquitous and - in most cases - without any cost, demanding only complementary analysis tools.

Data mining can be defined as "a process that uses statistical and mathematical techniques, artificial intelligence and automatic learning to extract and identify useful information and subsequent knowledge from databases" (Turban et al., 2009, p. 153). The data mining models consist basically in classification, grouping, association, sequence discovery, visualization, regression and prediction.

The classification is the most common data mining activity and aims to analyze historical data stored and

generate a model that can predict the future behavior. The model consists of training the algorithm in a data group, expecting the model to be used to predict the classes of other unclassified records (Turban et al., 2009). The usual classification tools include neural networks, decision trees, logistic regression and discriminant analysis (Sharda et al., 2014).

Another data mining activity is data grouping, also known as data clustering. In this case, the data in a database are divided into segments whose members have similar characteristics (Sharda et al., 2014). The association is also a data mining activity, which seeks to establish relationships between items that appear in a record, such as items that are usually sold together in a supermarket (Turban et al., 2009).

There are also the activities of sequence discovery, which seeks to identify associations along the time and also the activities of visualization, which, as the name says, seeks to present an overview of the data through mining. The activities of linear and non-linear regression use statistical tools to design the data to a predicted value. Finally, there are the activities of prediction, which assess future values based on analyses of the data mass, through the projection of temporal series and statistical tools Sharda et al. (2014). There are several free software of BIA with the aforementioned tools, which can be marketed without adding costs to companies. There are also free versions of commercial software.

In addition to being used in the industrial area, tools of this type are widely employed in sectors such as E-commerce, through intelligent market, where the data analysis systems automatically recommend products of possible interest to customers. There are still major use in the area of health, public safety and governmental systems (Chen et al., 2012).

The BIA systems generally used in industries are based on analysis of structured, pre-stored data from different systems of business control, allowing a better data visualization (Chen et al., 2012).

Since the construction industry has unique characteristics, once it demands from field work where a large amount of relevant data is generated, which are stored in various formats, extracting managerial information can be a difficult task (Martínez–Rojas et al., 2016). With this, the construction industry has been adopting technologies to mitigate this problem, but more slowly than the other industries (Lam et al., 2010, Ahuja et al., 2010, Majrouhi Sardroud, 2015). Some of the reasons cited for the lower adherence to those technologies include the lack of skills and knowledge about the technological tools (Majrouhi Sardroud, 2015, Lu et al., 2015, Viljamaa and Peltomaa, 2014) and the organizational culture of enterprises (Gajendran and Brewer, 2012).

Thus, there becomes evident the potential contribution of the topic of study for the construction sector, in order to identify which tools of BIA can add value to the management of these companies. In addition, the next section presents a systematic literature review, which aims to identify which data analysis tools are used and how they were applied in the construction industry.

3   Systematic Literature Review

According to Tranfield et al. (2003), Kitchenham (2007), a systematic literature review is a way to identify and interpret, based on pre-defined parameters, all studies concerning a particular topic. The individual studies evaluated in a SLR are called primary studies, making the SLR a secondary study about the subject.

In the management area, the literature review is an essential tool, used to manage the diversity of knowledge and guide a specific academic investigation (Tranfield et al., 2003). An effective review creates a solid basis for the advancement of knowledge, facilitates the development of theories, limits areas in which research is abundant and reveals areas in which research is needed (Webster and Watson, n.d.). This SLR was conducted according to the guidelines of Tranfield et al. (2003), Kitchenham (2007), and is composed of the following steps: (i) Planning of the SLR; (ii) Development of the SLR and (iii) Reporting the SLR. Next, there us the planning of this review.

3.1   Systematic review planning

According to Kitchenham (2007), the planning of a SLR involves: (i) Identifying the need for a SLR; (ii) Commissioning the SLR; (iii) Defining research questions; (iv) Developing the SLR Protocol and (v), Validating the protocol. Next, the items (i), (iii) and (iv) are presented, considered the most relevant to understand the work.

3.1.1   *Identifying the need for a SLR*

Since no similar studies were identified, there came the decision to prepare this systematic review aiming to identify and summarize the topics that address the subject. The literature review also seeks to gather this information to allow other authors to conduct their research from the present study, which may be used for the deepening of their research.

3.1.2   *Defining the research questions*

Assuming the need for executing a SLR, research questions were defined to be answered by the SLR. The questions relate to the use of technologies for data analysis to support the management in construction. With this, one seeks to identify how the studies report this relationship, whether by describing which software/technologies have been used or in which sector and the purpose of their use.

Q1. Which data analysis tools have been used in the construction area? The purpose of this question is to identify, in the first place, if there are data analysis software/technologies used in the construction industry, to list them, if they exist. Eventually, tools that initially did not have BIA characteristics may be identified, but they might contribute to supplying information to those technologies.

Q2. What is the purpose of using these technologies, and in what areas? Identified the software and technologies, this question seeks to elucidate in which stage of the management process they are applied and how they are used.

3.1.3   Developing the SLR protocol

The SLR protocol is the step that documents the methodology to be used to carry out the systematic review. According to Kitchenham (2007), the protocol is composed of (i) Definition of databases; (ii) Definition of the research strategy; (iii) Definition of the criteria for the selection of primary studies; (iv) Definition of the method to extract information and (v) Definition of the form of data presentation. Next, there are the details of the items (i), (ii) and (iii). The others will be presented in the course of work.

A tool was used to conduct this SLR. It is a software called LaPES – Software Engineering Research Laboratory (2012), which provides support from the definition of the protocol until the extraction and analysis of results. The software allows the bibliographic data of applicants for primary studies to be imported in .bib or .ris format, facilitating the initial application of the inclusion and exclusion criteria, which take into account the reading of the titles and abstracts of the works without the need to download all full articles. In addition, the system generates records of the application of the criteria, as well as of the extracted data, allowing, at the end, the generation of a report in .xls format with data that allow the creation of complementary graphs.

Some criteria were defined to choose the databases, namely: Containing studies related to IT, management, technology, construction and allowing searching with Boolean operators. With these criteria, the following repositories were defined as databases: ACM Digital Library; Engineering Village; IEEE Xplore Digital Library; Material Science Engineering: A (Elsevier); Science Direct; Scopus and Web of Science, in order to maximize the chances of finding primary studies. Furthermore, they are well known data bases and globally disseminated.

As a research strategy, a search string was built. As the same string cannot be used, because the databases have mechanisms for advanced search with different characteristics, small adjustments were made to be used according to the databases. The search string is presented below.

```
("data warehouse"OR "OLAP"OR "dashboard*"OR "business
intelligence"OR "Data mining"or "business analytics") AND
("construction management"OR "Construction Planning"OR
"construction enterprise*"OR "construction work*"OR
"building construction"OR "construction company*")
```

The SLR was conducted taking into account scientific papers published from the year 2000, therefore, the period covered by the survey comprises approximately 19 years, since the application of the strings occurred in July 2019. With the time range, it was possible to observe the trends of technological development about the SLR theme over the years. Some databases allowed restricting the year of publication only after applying the search string. In these cases, the time and study type restrictions were applied manually on the page of search results of the database. More details about the inclusion and exclusion criteria of studies will be presented below.

As selection strategy of primary studies, a set of inclusion and exclusion criteria was defined based on the research scope, in order to ensure that only studies related to the SLR subject were chosen. To be included, the study needed to meet one or more inclusion criteria. To be disregarded, needed to fit any of the exclusion criteria. Below, we list the inclusion and exclusion criteria:

Inclusion Criteria
I1 – The paper addresses Business Intelligence and Construction directly and not only mentions in the text;
I2 – The paper addresses OLAP and Construction directly and not only mentions in the text;
I3 – The paper addresses data warehouse and Construction directly and not only mentions in the text;
I4 – The paper addresses Business Analytics and Construction directly and not only mentions in the text.;
I5 – The paper addresses Dashboards and Construction directly and not only mentions in the text;
I6 – The paper addresses Data Mining and Construction directly and not only mentions in the text.

Exclusion Criteria
E1 – The paper is not fully available online;
E2 – paper is not written completely in English;
E3 – paper not related to the area of construction;
E4 – The paper addresses BIA software/technologies, but not related to management in construction;
E5 – The paper does not address BIA in construction;
E6 – The document is not an paper, being in fact a technical report, book or book chapter, proceeding, preface or editorial.

3.2   Execution of the systematic review

This section presents the main steps to develop the SLR: (i) identification and selection of primary studies; (ii) verification of the quality of primary studies and (iii) data extraction and synthesis of primary studies.

3.2.1   Identification of primary studies

The identification of candidates for primary studies occurred with the implementation of search strings on the engines of advanced search of the respective databases. At this stage, there were 1350 candidates for primary studies. After reading the titles, abstracts and keywords, the pre-defined aforementioned inclusion and exclusion criteria were applied. After this stage, 772 were rejected, 465 studies were identified as duplicates and 113 studies were accepted for the next step.

3.2.2   Primary studies quality assessment

To verify the quality of primary studies, new inclusion/exclusion criteria were defined, which were applied after a complete reading of the papers. Paper published reviewed by peers that responded to at least one of the questions defined in the SLR were kept. Moreover, there was exclusion of duplicated papers that could not be identified with just the reading of the title and abstract summary, due to the small changes made by the authors, imperceptible in the first readings of titles, abstracts and keywords.

After applying the criteria for verification of quality of studies, 59 scientific papers remained. Table 1 shows the amount of studies for each phase of the selection process. Of the primary studies identified, 37 were published in journals and 22 were published in conferences.

Table 1: Primary studies selected by database repositories

DataBase	Candidates	Pre-selected	Final result
Scopus	462	54	28
Science Direct	247	13	10
ACM Digital	245	11	6
Material Science	211	13	8
Engineering Village	170	18	6
Web of Science	61	4	1
IEEE Xplore Digital	11	3	3
Total	1407	116	62

Aiming to standardize the nomenclature of studies and at their general presentation, Table 2 and Table 3 present the lists with the final studies identified in conferences and journals, respectively. For studies published in conferences/events, the coding used was E01, E02 and so on. For studies published in journals, the coding used was J01, J02, and so on.

Table 2: Conference studies

Year	ID	Authors
2002	E01	Cao et al. (2002)
2004	E02	Xiaopeng et al. (2004)
2005	E03	Rujirayanyong and Shi (2005)
2006	E04	Zhou and Ding (2006)
2007	E05	Kim et al. (2007)
2007	E06	Vuori (2007)
2008	E07	Chang and Guan (2008)
2008	E08	Barth and Formoso (2008)
2009	E09	Wang et al. (2009)
2009	E10	Zhigang and Yajing (2009)
2012	E11	Riaz and Husain (2012)
2012	E12	Guerriero et al. (2012)
2013	E13	Ahiaga-Dagbui and Smith (2013)
2014	E14	Gui et al. (2014)
2015	E15	Riaz et al. (2015)
2017	E16	Kushiro et al. (2017)
2017	E17	Seong et al. (2017)
2017	E18	Montaser, A. ; Montaser (2017)
2018	E19	Girsang et al. (2018)
2018	E20	Tianyi et al. (2018)
2018	E21	Konikov (2018)
2018	E22	Konikov et al. (2018)
2020	E23	Çelik (2020)

Table 3: Journal studies

Year	ID	Authors
2001	P01	Rezgui (2001)
2003	P02	Chau, Anson, Ying and Jianping (2003)
2003	P03	Chau, Cao, Anson and Zhang (2003)
2004	P04	Wang et al. (2004)
2004	P05	Ahmad et al. (2004)
2006	P06	Tam et al. (2006)
2006	P07	Rujirayanyong and Shi (2006)
2007	P08	Moon et al. (2007)
2010	P09	Szelka and Wrona (2010)
2010	P10	Cheng et al. (2010)
2011	P11	Rezaei et al. (2011)
2011	P12	Ma et al. (2011)
2012	P13	Cheng, Leu, Cheng, Wu and Lin (2012)
2012	P14	Cheng, Su and Cheng (2012)
2013	P15	Horta and Camanho (2013)
2013	P16	Cheng et al. (2013)
2013	P17	Fan and Li (2013)
2013	P18	Ma et al. (2013)
2013	P19	Chong and Phuah (2013)
2013	P20	Li and Zhang (2013)
2014	P21	Chi et al. (2014)
2014	P22	Hammad et al. (2014)
2014	P23	Ahiaga-dagbui et al. (2014)
2014	P24	Williams and Gong (2014)
2015	P25	Shin (2015)
2015	P26	Lu et al. (2015)
2015	P27	Cheng et al. (2015)
2016	P28	Tixier et al. (2016)
2016	P29	Huang (2016)
2016	P30	Martínez-Rojas et al. (2016)
2016	P31	Bilal et al. (2016)
2017	P32	Lee and Yi (2017)
2018	P33	Zhang et al. (2018)
2018	P34	Shin et al. (2018)
2018	P35	Ratajczak et al. (2018)
2019	P36	You and Wu (2019)
2019	P37	Ajayi et al. (2019)
2020	P38	Huang and Hsieh (2020)
2020	P39	Ghazal and Hammad (2020)

Q1. Which data analysis tools have been used in the construction area? Two attributes to be checked were defined, namely: the data analysis technology used (data warehouse, OLAP, dashboards or data mining), as well as the software used. At the end, all software/technologies identified in the studies were listed, whether of BIA or to support the use of technologies.

Q2. What is the purpose of using these technologies, and in what areas? Again, since this was an exploratory question, the strategy adopted for extraction was the listing of the various forms of use of the BI software/technologies in construction identified in the studies.

3.2.3   *Data extraction and synthesis of primary studies*

After completing the previous steps, the 59 selected papers were fully read, searching for answers to the listed questions. Since each question has a goal, distinct strategies were defined for each question, as can be seen below.

3.3   The findings of the SLR

From this section, the results obtained with the implementation of the systematic review will be presented.

3.3.1 Identified tools

The primary studies evaluated presented different answers to this question. Initially, the search result was presented by the software mentioned in the studies. Notably, 27 studies mentioned that the software used was developed specifically for the problem situation, which represents more than 64.28% of the answers. Secondly, there comes the commercial software, led by Primavera, with. Table 4 shows the responses provided by each primary study identified.

Table 4: *Software found by study*

Software	Study ID
Accept System	P35
IBM SPSS	P38
Matlab	E11
Microsoft Clustering	E11
Power BI	E23
Pentaho	E19
Primavera	E03, P07, P22
Qlik	E19
Rapid Miner	P24
SAP	P12
Own Software	E01, E04, E08, E10, E12, E13, E15, E17, P01, P02, P04, P05, P07, P08, P09, P11, P15, P16, P18, P19, P20, P21, P23, P27, P30, P32, P37
Statsoft Statistica	P10
Weka	P39
Wordstat	P28

When analyzing the BIA technologies used, Fig. 1 shows the results. There were 33 records of use of data warehouse, which correspond to 35.48% of the total number of records. Additionally, on 35 occasions, the application of data mining as a tool became evident, which represents 37.63% of the records. For the use of OLAP tool, it occurred on 21 occasions, which represents 22.58% of the records. Finally, the use of the dashboards tool was evidenced on 4 occasions, which corresponds to approximately 4.30%.

Therefore, the data analysis tool most used in the construction industry is data mining, followed by data warehouse and OLAP, on several occasions used in combination.
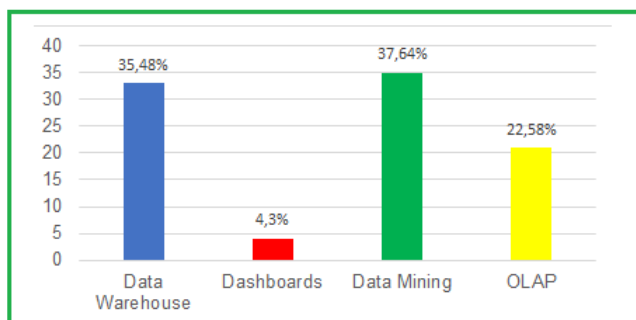

Figure 1: BIA Technologies found in the literature review

A more detailed analysis allows affirming – based on the studies analyzed in the present paper – that the OLAP tool is used only in conjunction with the data warehouse tool, since there were no records of both tools alone, confirming, therefore, what the literature states about the topic, since the database for OLAP is always a data warehouse. Table 5 lists the technologies mentioned in studies and their identifications.

Table 5: Technologies found by study

Technology	Study ID
Data Warehouse	E01, E02, E03, E04, E05, E09, E13, E15, E18, E19, E21, E22, P01, P02, P03, P04, P05, P07, P08, P10, P11, P12, P13, P18, P19, P20, P22, P23, P30, P31, P36, P39
Dashboards	E08, E12, P35
Data Mining	E02, E04, E07, E10, E11, E13, E14, E15, E16, E17, E20, P06, P13, P14, P15, P16, P17, P21, P22, P23, P24, P25, P26, P27, P28, P29, P31, P32, P33, P34, P36, P37, P38, P39
OLAP	E01, E04, E05, E09, E19, E21, P01, P02, P03, P05, P08, P10, P11, P13, P19, P20, P22, P30, P31, P36, P39

3.3.2 Identified application area

There are several uses of data analysis for construction, in various sectors and phases in the process of project implementation. Fig. 2 presents the obtained results.
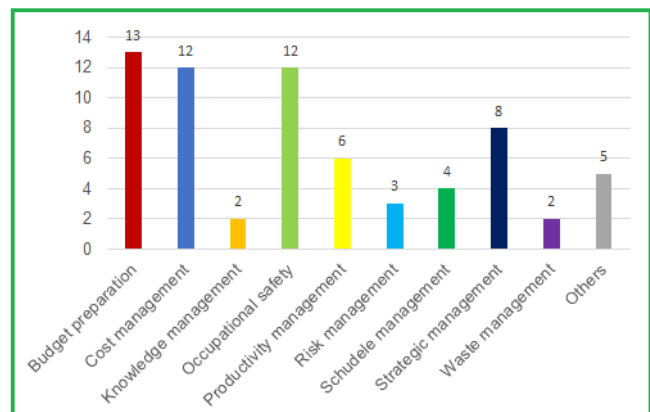

Figure 2: BIA application areas found in the literature review

It is possible to observe 14 different uses of tools to support the management in construction. The areas of preparation of the budgets of works, occupational safety and cost management hold the greatest records of use, showing, which correspond, in conjunction, to approximately 55.22% of the total. There were also in strategic management, with 8 records, productivity management, with 6, schedule management, with 4, and

risk management with 3 records followed by knowledge management and waste management, with 2 reports each. In addition to those, there were reports of use in the areas of inventory management, vendor management, quality management, personnel management and decision support for the preparation of structural design, each one with 1 record evidenced. Table 6 shows the uses and their respective studies.

Table 6: BIA application areas in construction by study

Application area	Study ID
Preparation of budgets	E03, E10, E13, P07, P08, P12, P16, P19, P22, P24, P29, P30, P38
Productivity management	E08, E14, E20, P11, P33, P35
Quality management	P27
Costs management	E01, E04, E19, E23, P02, P03, P19, P23, P24, P25, P29, P39
Inventory management	P07
Vendor management	P19
Personnel management	E07
Waste management	P20, P26
Risk management	E02, P19, P32
Knowledge management	P01, P19
Schedule management	E19, P03, P04, P35
Strategic management	E05, E06, E11, E12, E15, P05, P15, P36
Structural project DSS	P09
Occupational safety	E16, E17, P06, P10, P13, P14, P17, P18, P21, P28, P34, P37

In a further analysis, aiming to know more deeply the *modus operandi* of construction companies that make use of data analysis tools, a relationship was established between application area and the types of software used in each one of them, as shown in Fig. 3.

Most analyzed occurred through the use of their own software, i.e., those designed and programmed specifically for the company in question. In 13 of the 14 recorded, an own software was used, totaling 32 relationships. Then, there is the use of the software Primavera, related three times to the preparation of budgets and once to inventory management, as the IBM SPSS for budget preparation. Pentaho, Power BI, Qlik and Weka were related to applications of schedule and cost managements. The SAP was related to the preparation of budgets and cost management, but only providing data to the data warehouse, whereas the Statsoft Statistica had an application in the area of occupational safety.

It is believed that the large number of applications developed specifically for each case, to the detriment of commercial programs, is due to the fact that most of these applications used analytical concepts through data mining applications, which require greater customization due to the characteristics of the available data and the objectives pursued with the use of the application.

In addition to the relationship between software and application, there was an analysis of the relationship between technology and application, in order to identify
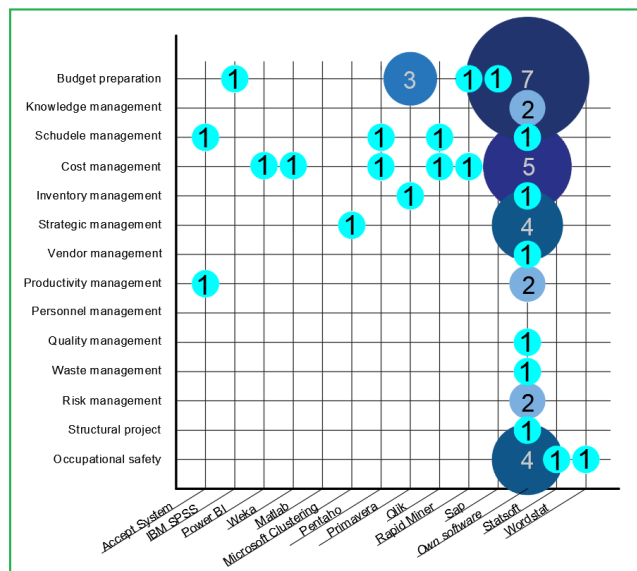

Figure 3: Relation between application area and software

the technologies most used for a particular purpose. The technologies considered were dashboards, data mining, data warehouse and OLAP. Fig. 4 shows the relationships found. The interpretation principle is the same used in the previous image, that is, the larger the diameter of the circumference, the greater the number of relations between the application and the technology.

We cite as examples of proprietary applications that made use of data mining the studies of Chi et al. (2014), which used text mining for analysis of work safety reports for the prediction of safety risks in works and Riaz et al. (2015) who used the LVQ algorithm for grouping data in "R" language to obtain predictive information about strategic decision making. Both studies focused on predictions and, like other studies of the type found in RSL, have great technical complexity for development and implantation, which can make the use in small and medium-sized companies unfeasible.

The technologies most cited in the papers were data mining and data warehouse, with 33 and 32 records, respectively. The OLAP and dashboards technologies were also listed, with 23 and 4 records, respectively. The data mining technology had a greater number of use in the area of occupational safety in construction, with 10 records, followed by preparation of budgets and cost management. There were also 4 records of strategic management, 3 of management of productivity and 2 of risk management. Personnel management, waste management and risk management had 1 record each.

The data warehouse technology was related to the preparation of budgets and cost management, with 7 and 9 records, respectively, for each use and with 4 records for strategic management. Use in the schedule management were also checked, as well as in occupational safety, with 3 records and inventory management, vendor management, productivity management and waste management, with one record each. OLAP was reported with use in cost management and preparation of budgets, with
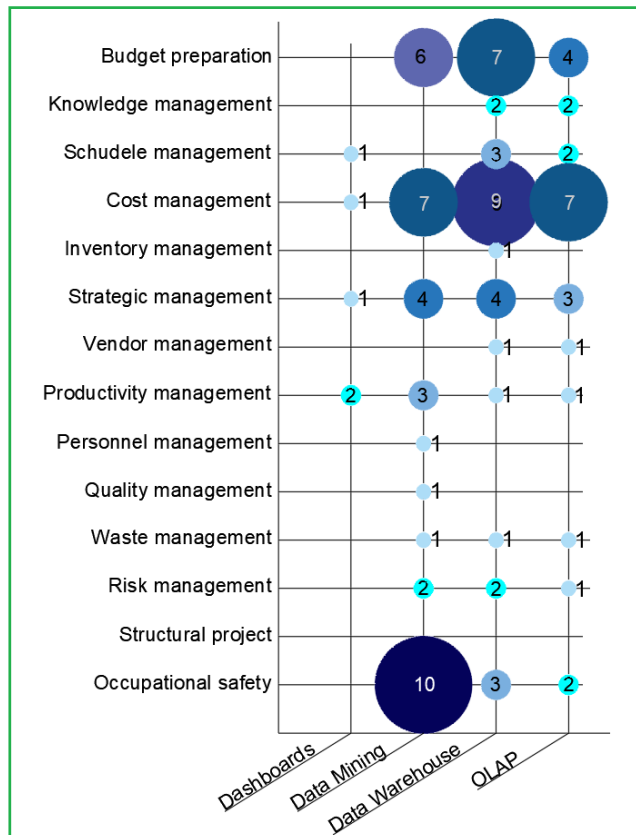
Figure 4: Relation between application area and software

6 and 4 records, respectively, as well as in strategic management, with 3 records and knowledge management and schedule management, with 2 records each and vendor management, productivity management, waste management and risk management, with 1 record each. The dashboards technology had use related to productivity management, with 2 records and schedule management and strategic management, with 1 record each.

Finally, seeking to take advantage of the temporal range used in the systematic review, Fig. 5 shows the technologies used according to the analyzed year. The period considered was from 2001 to 2018, since the year 2019 was partially covered by the survey, which could distort the result presented.

From 2001 to 2003, there was no use of the data mining technology, being registered only the use of data warehouse and OLAP. The first record of use of data mining occurred in the year 2004, with its use increasing significantly. However, are they are exploratory studies, from data collected and prepared for mining, such as case studies, being a small number of those with solution coupled to the company's software.

The data warehouse technology was verified in all years, with the largest number of records in 2013, with 4 occurrences. This fact was already expected, since the data warehouse technology is the basis of the process of data analysis, on which other tools are applied.

The use of OLAP remained constant over the years, not

being checked publications that reported use only in 2005, 2008, 2015 and 2017. Finally, the dashboards technology, the least reported, had records in 2008, 2012, 2018 and 2020.

3.3.3   Synthesis of studies

This section seeks to present an overview about how the main studies have proposed the development of solutions used, in order to obtain familiarity with the theme, to, then, propose the technical solution to be used in the case study.

With this, we start introducing the study of (Cao et al., 2002), which describes the combined application of data warehouse and OLAP technologies to support the cost management in construction by developing a customized solution. For this purpose, they used the star schema of multidimensional modeling, which is a database that allows using OLAP and consists of fact tables and dimension tables. While the fact tables contain quantitative data about data management of works, dimension tables contain descriptive data, which compose the cube dimensions. With the data warehouse defined in this way, the SQL (Structured Query Language) were created, which relate data between the fact tables and dimension tables, returning the necessary data.

(Cao et al., 2002) integrate the OLAP to data warehouse designed to allow the extracted data to be analyzed as a multidimensional cube, which can be exploited by front-end tools such as Visual Basic, MS Access and MS Excel.

(Chau, Anson, Ying and Jianping, 2003) and (Chau, Cao, Anson and Zhang, 2003) also proposed the use of data warehouse and OLAP tools for cost management in works. For this purpose, they proposed a system capable of acquiring data in different locations, such as databases and external files via an ETL. The extracted data were stored in a data warehouse, and, after that, with the use of OLAP, data are used in a front-end tool.

Wang et al. (2004) proposed the use of the data warehouse tool in conjunction with 4D planning software in construction. One can say that this 4D software incorporates information regarding the progress of the work and the schedule, so that one can observe the three-dimensional construction of the model within planned space-time. As the 4D software has a lot of information in its database, a data warehouse was created, with data in a star schema, containing fact and dimension tables, in order to respond, at any time, questions like "where was the material x applied?"

In their study, Xiaopeng et al. (2004) show the use of data warehouse technology in the preparation phase of processed data from data mining for risk management. In the studied case, the use of association rules was proposed, indicating that the application of data mining remained jeopardized due to data scarcity, given the nature of the industry itself.

Rujirayanyong and Shi (2005) describe the implementation of data warehouse for preparation of budgets of works, using Primavera commercial software as a data source. From the database of this system, they proposed the steps to develop for the development of technology and cited as a critical point in the development of the ETL, acronym for extraction,
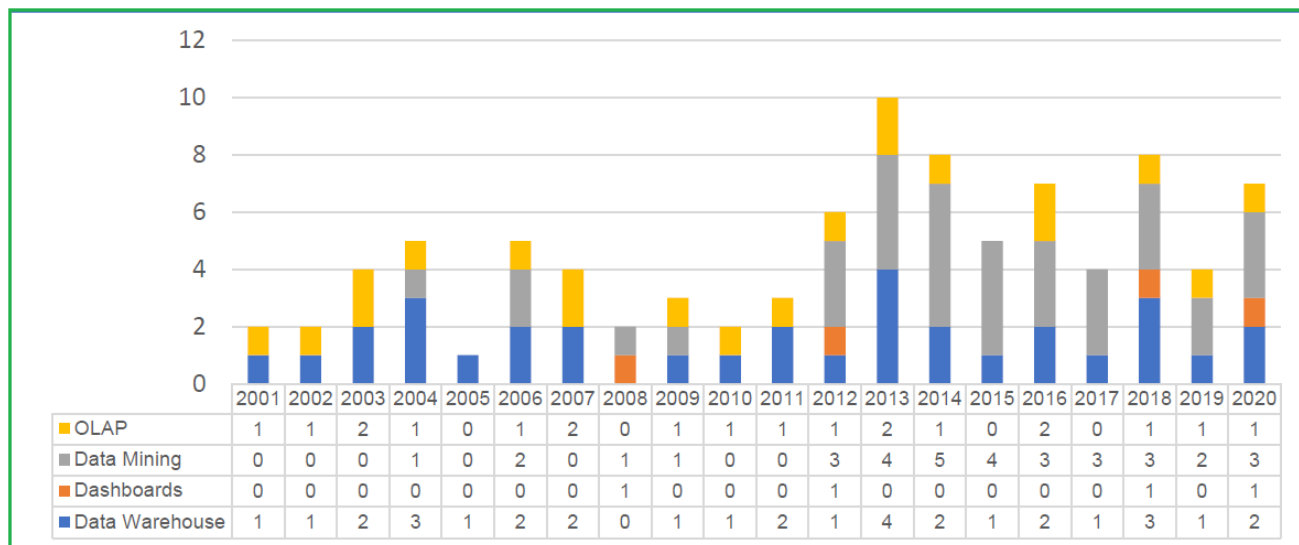
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
OLAP	1	1	2	1	0	1	2	0	1	1	1	1	2	1	0	2	0	1	1	1
Data Mining	0	0	0	1	0	2	0	1	1	0	0	3	4	5	4	3	3	3	2	3
Dashboards	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	1
Data Warehouse	1	1	2	3	1	2	2	0	1	1	2	1	4	2	1	2	1	3	1	2

Figure 5: Technologies used between 2001 a 2018

transform and load, a process in which the rules are defined to acquire data in databases, preparation and loading in data warehouses, as presented in the theoretical framework of this work.

In another study, Rujirayanyong and Shi (2006) proposed the use of a data warehouse in conjunction with the database of the Primavera software for the preparation of budgets of works by inventory management. Through analysis of historical data, they developed a data warehouse comprising 16 dimension tables and 10 fact tables with data extracted through an ETL. One of the dimension tables included data such as the material's name, the unit of measurement and identification codes. Finally, they concluded that historical data can help managers make better decisions in construction. On the other hand, the difficulty cited by them regards the existence of a model ready for each company, which shall schedule the solution according to one's needs.

Zhou and Ding (2006) demonstrate the application of data mining tool in conjunction with the data warehouse and OLAP tools. In their work, they proposed an architecture that uses data mining to bring insights for the cost management of works. Overall, the proposed scheme consists of databases, data warehouses and data marts, OLAP servers and front-end tools.

The work of Moon et al. (2007) used the data warehouse and OLAP tools for the project's budget preparation. Through statistical analysis of historical data, it can generate data that have improved the cost estimates of future works, with levels of uncertainty calculated.

Kim et al. (2007) propose the use of data warehouse and OLAP tool for strategic management in construction. To do this, they describe the management process in 5 steps, which include measures, storage, analysis, report and action. In Step 1, through the use of the Balanced Scorecard (BSC), measurable indicators are defined, associated with the data. At this stage, the weights for each criterion were also defined. In Step 2, storage, a data warehouse is used. In Step 3, there was a predicted use of OLAP, which allowed, by the data stored, the composition of the data cube, which allowed the slicing and visualization of data according to users' needs. In step 4, reports, visual information of the data was generated in .XML format, without the need for additional programming. Finally, in step 5, the data collected and analyzed were used to support the strategic decision, through search tools, comparisons, history, among others.

Cheng et al. (2010) proposed the use of data warehouse and OLAP tools, in conjunction with the software StatSoft to find cause and effect relationships in records of occurrences of accidents at work. In this context, 1,347 reports of accidents were analyzed, recorded between the years 2000 and 2007. After the implementation of statistical tests in the sample data found, OLAP was used to find the data distribution and find correlation coefficients of attributes. At the end, data revealed the main factors that affect the causes of accidents, including the type of construction, the size of the company, the experience of workers, failure in the use of safety equipment, lack of safety infrastructure (scaffolding, platforms etc.), among others.

The following year, Rezaei et al. (2011) proposed the use of the data warehouse and OLAP tools to develop a system for the management of employees' productivity in construction companies certified with ISO 9001. The system of web platform used the data warehouse tool to collect data from the ongoing processes to measure the performance of employees and OLAP to view the data. In addition to those tools, the computerized system scanned the process of quality management and systematized the communication within the company. After the deployment of the system in two companies, the authors managed to identify cost reduction of 50.00% in the administrative area of construction companies, as well as the reduction between 40 and 60% at the time of execution of the activities.

Ahiaga-dagbui et al. (2014) proposed the joint use of data warehouse and data mining technologies for preparation of budgets of works. For this purpose, they initially analyzed the characteristics of the available data, through clustering. After identifying patterns, they chose those without significant amounts of missing data and outliers to compose the initial database. On these data, the bootstrap and artificial neural networks techniques were applied, considered by authors to be the most appropriate for the data found, which have non-linear variables. After the analysis of 1,600 works carried out between 2004 and 2012, 15 cases were selected for analysis by stratified sampling for the final test of the model. The rest of the data was divided in the ratio 80:20 for training and testing of the network, respectively. Among the models proposed, the one with the best performance managed to estimate the cost of the work with an error of only 3.67%.

Hammad et al. (2014) made use of data warehouse tools to discover knowledge in databases of the construction industry, with a DW of snow flake modeling and data mining techniques based on data grouping. In this way, they managed to obtain information about the characteristics of the productive process, as well as were able to generate information for decision making. Meanwhile, Williams and Gong (2014) used an algorithm to classify data to create a prediction model of overflows of budgets of works. In addition, they used text mining tools and achieved an accuracy rate of 43.72%.

Shin (2015) presented the results of the application of the mining technique of the regression trees category for the preparation of estimates of the costs of construction of buildings. After determining the factors that affect the cost of a work, selecting the data and applying the model on data from 234 works carried out, the error rate obtained ranged between 5.8 and 6.05%, slightly better than the error rate obtained with the use of model based on neural networks.

Riaz et al. (2015) demonstrate the use of data warehouse and data mining tool to provide a tool to support the strategic decision. For this purpose, they propose the use of the LVQ algorithm (learning vector quantization) modified for data grouping of various works, in order to find insights that can help in the decisions of works in progress.

Girsang et al. (2018) also proposed the use of data warehouse and OLAP tools, in conjunction with the software Pentaho. With the support of this software, an ETL was configured to provide data that were arranged in a star schema in the data warehouse. With these data organized, they used the OLAP tool for analysis and creation of dashboards that helped in the decision making process.

Finally, Ghazal and Hammad (2020) propose the use of data mining tools, to build a model capable of predicting construction costs and eventual budget extrapolations. For that, they used the Weka software, where techniques implemented in it were used in the tasks of grouping and classifying data. Finally, he obtained a model capable of estimating the costs of a work with a reasonable accuracy of 60.87

The methodologies used by the authors presented different ways of applying BIA technologies and tools.

In the studies presented, important information for the decision-making process was generated. It was demonstrated, therefore, that the analytical applications, whether commercial or developed specifically for each organization, have great potential for contribution to the civil construction sector.

4 Conclusions

The main objective of the study was to obtain an overview of the use of the BIA tools to support the management in the construction sector from the answer to both research guiding questions of the SLR.

The development of this systematic review allowed verifying the existence of different types of software and tools used in various sectors and phases of projects. For question 1, "Which data analysis tools have been used in the construction area?", it was possible to observe the use of the software Accept System, IBM SPSS, Matlab, MS Clustering, Pentaho, Power BI, Primavera, Qlik, Rapid Miner, SAP, Statsoft Statistica Weka and Wordstat, but more than 64.00% of the software types checked were not commercial and had been specifically developed for each company. It is believed that the large number of solutions developed specifically for each case is due to the characteristics inherent to the construction sector, which is known to involve field work, in temporary locations and with large data production in different formats Martínez-Rojas et al. (2012).

On the other hand, analytical applications, especially BIA applications, were originally developed to provide executive information Negash and Gray (2008), which would justify the need to develop alternative and adapted solutions, since commercial software would not be able to fully serve the construction sector. Cao et al. (2002) stated that the fact that the works were temporary and specific, that is, different from each other, required different forms of use for BIA tools, proposing a use with short-term vision to the detriment of the use for long-term strategic levels, as usual. Another possible justification, based on the fact that 70% of the records of use of own software were verified in articles published in journals, is that this type of study has a higher chance of publication, to the detriment of presenting results of the use of commercial tools. As for the technologies, there were frequent uses of Data Warehouse, Dashboards, Data Mining and OLAP.

For question 2, "What is the purpose of using these technologies, and in what areas?", it was possible to observe 14 different applications, and the main forms of utilization of BIA technologies for management support were preparation of budgets of the works, cost management of works in progress and in the area of occupational safety. In addition to these, there were also uses in the areas of knowledge management, risk management, strategic management, schedule management, productivity management, inventory management, vendor management, waste management and structural design.

In a further analysis, which crossed use data versus software used, it was possible to observe that, for preparation of budgets, productivity management, vendor

management, waste management, risk management, cost management, knowledge management, strategic management, occupational safety and structural design, the use of a software developed specifically for each company predominated. As for the use for schedule management, the use of commercial software predominated.

In addition, when relating use versus technologies, it was possible to verify that the tools focused on occupational safety, preparation of budgets and cost management of works usually made use of Data Mining in conjunction with Data Warehouse and OLAP. These results reveal that BIA technologies are used in different sectors of Construction, transcending the area of financial management and supporting even the production processes.

With the technological development increasingly accelerated and an economic scenario with increasing competition, it can be argued that the use of these technologies will be a determinant factor for construction companies to remain in the market in the long term. Furthermore, a change of culture for managers of the construction sector is necessary for this to occur so widely.

Further analyses may be necessary to understand the technical requirements for implementation of the technologies mentioned in companies, since this work presents a broad overview of the scenario, not deepening in the technical area, for reasons of limited scope.

In the primary studies identified, the authors cite the benefits of using these technologies. Future works can assess the challenges of deployment, as well as the limitations found in the use of BIA tools in construction companies.

References

Abou-Ibrahim, H. and Hamzeh, F. (2020). A visual dashboard to monitor BIM model dynamics, *Canadian Journal of Civil Engineering* **47**(2): 178–185. http://dx.doi.org/10.1139/cjce-2018-0411.

Ahiaga-dagbui, D. D., Smith, S. D., Ahiaga-dagbui, D. D. and Smith, S. D. (2014). Construction Management and Economics Dealing with construction cost overruns using data mining, (December): 37–41. https://doi.org/10.1080/01446193.2014.933854.

Ahiaga-Dagbui, D. and Smith, S. (2013). My cost runneth over: Data mining to reduce construction cost overruns, *Proceedings 29th Annual Association of Researchers in Construction Management Conference, ARCOM 2013*, number September, pp. 559–568. http://doi.org/10.13140/2.1.2087.776010.13140/2.1.2087.7760.

Ahmad, I., Azhar, S. and Lukauskis, P. (2004). Development of a decision support system using data warehousing to assist builders/developers in site selection, *Automation in Construction* **13**(4): 525–542. http://doi.org/10.1016/j.autcon.2004.03.001.

Ahuja, V., Yang, J. and Shankar, R. (2010). Benchmarking framework to measure extent of ICT adoption for building project management, *Journal of Construction Engineering and Management* **136**(5): 538–545. http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000155.

Ajayi, A., Oyedele, L., Davila Delgado, J. M., Akanbi, L., Bilal, M., Akinade, O. and Olawale, O. (2019). Big data platform for health and safety accident prediction, *World Journal of Science, Technology and Sustainable Development* **16**(1): 2–21. http://doi.org/10.1108/wjstsd-05-2018-0042.

Barth, K. and Formoso, C. (2008). Improvement of performance measurement systems using production management dashboards, *Proceedings of IGLC16: 16th Annual Conference of the International Group for Lean Construction*, pp. 769–780.

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**(3): 500–521. http://dx.doi.org/10.1016/j.aei.2016.07.001.

Cao, Y., Chau, K. W., Anson, M. and Zhang, J. (2002). An Intelligent Decision Support System in Construction Management by Data Warehousing Technique, pp. 360–369.

Çelik, U. (2020). 4D and 5D BIM: A system for automation of planning and integrated cost management, *Communications in Computer and Information Science* **1188 CCIS**: 57–69. http://dx.doi.org/10.1007/978-3-030-42852-5_5.

Chang, Y. and Guan, M. (2008). Data mining to improve human resource in construction company, *2008 International Seminar on Business and Information Management, ISBIM 2008*, Vol. 1, pp. 275–278. https://doi.org/10.1109/ISBIM.2008.187.

Chau, K. W., Anson, M., Ying, C. and Jianping, Z. (2003). Integration of data warehouse into knowledge-based system on construction management decision making, *HKIE Transactions Hong Kong Institution of Engineers* **10**(1): 8–13. https://doi.org/10.1080/1023697X.2003.10667895.

Chau, K. W., Cao, Y., Anson, M. and Zhang, J. (2003). Application of data warehouse and decision support system in construction management, *Automation in Construction* **12**(2): 213–224. http://doi.org/10.1016/S0926-5805(02)00087-0.

Chen, H.-c., Chiang, R. and Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly* **36**: 1165–1188. https://doi.org/10.2307/41703503.

Cheng, C. W., Leu, S. S., Cheng, Y. M., Wu, T. C. and Lin, C. C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry, *Accident Analysis and Prevention* **48**: 214–222. http://dx.doi.org/10.1016/j.aap.2011.04.014.

Cheng, C. W., Lin, C. C. and Leu, S. S. (2010). Use of association rules to explore cause-effect relationships in occupational accidents in the Taiwan construction industry, *Safety Science* **48**(4): 436–444. http://dx.doi.org/10.1016/j.ssci.2009.12.005.

Cheng, M.-y., Hoang, N.-d. and Wu, Y.-w. (2013). Automation in Construction Hybrid intelligence approach based on LS-SVM and Differential Evolution for construction cost index estimation : A Taiwan case study, *Automation in Construction* . http://dx.doi.org/10.1016/j.autcon.2013.05.018.

Cheng, Y. M., Su, J. Y. and Cheng, C. W. (2012). Applying Clustering Algorithms to Explore the Circumstances of Fatal Falls in Construction Industry, *Applied Mechanics and Materials* **174-177**: 2126–2132. https://doi.org/10.4028/www.scientific.net/amm.174-177.2126.

Cheng, Y., Yu, W. D. and Li, Q. (2015). GA-based multi-level association rule mining approach for defect analysis in the construction industry, *Automation in Construction* **51**(C): 78–91. http://dx.doi.org/10.1016/j.autcon.2014.12.016.

Chi, N. W., Lin, K. Y. and Hsieh, S. H. (2014). Using ontology-based text classification to assist Job Hazard Analysis, *Advanced Engineering Informatics*, Vol. 28, Elsevier Ltd, pp. 381–394. http://dx.doi.org/10.1016/j.aei.2014.05.001.

Chong, H. Y. and Phuah, T. H. (2013). Incorporation of database approach into contractual issues: Methodology and practical guide for organizations, *Automation in Construction* **31**: 149–157. https://doi.org/10.1016/j.autcon.2012.10.007.

Fan, H. and Li, H. (2013). Automation in Construction Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques, *Automation in Construction* **34**: 85–91. http://dx.doi.org/10.1016/j.autcon.2012.10.014.

Ferreira, R. P., Martiniano, A. and Sassi, R. J. (2016). Dashboard inteligente para apoio à tomada de decisão em empresa de courier, *Revista Gestão & Tecnologia* **16**(2): 39–72. Disponível em http://http://www.spell.org.br/documentos/download/42759.

Gajendran, T. and Brewer, G. (2012). Cultural consciousness and the effective implementation of information and communication, *Construction Innovation: Information, Process, Management* **12**: 179–197. http://dx.doi.org/10.1108/14714171211215930.

Ghazal, M. M. and Hammad, A. (2020). Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects, *International Journal of Construction Management* **0**(0): 1–15. http://dx.doi.org/10.1080/15623599.2020.1738205.

Girsang, A. S., Isa, S. M., Saputra, H., Nuriawan, M. A., Ghozali, R. P. and Kaburuan, E. R. (2018). Business Intelligence for Construction Company Acknowledgement Reporting System, *1st 2018 Indonesian Association for Pattern Recognition International Conference, INAPR 2018 - Proceedings*, IEEE, pp. 113–122. https://doi.org/10.1109/INAPR.2018.8627012.

Guerriero, A., Zignale, D. and Halin, G. (2012). A zoomable location-based dashboard for construction management, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7467 LNCS, pp. 207–210. http://dx.doi.org/10.1007/978-3-642-32609-7_29.

Gui, X., Hu, Z., Zhang, J. and Bao, Y. (2014). Assessing personal performance with M-SVMs, *Proceedings - 2014 7th International Joint Conference on Computational Sciences and Optimization, CSO 2014*, pp. 598–601. http://doi.org/10.1109/CSO.2014.115.

Hammad, A., AbouRizk, S. and Mohamed, Y. (2014). Application of KDD Techniques to Extract Useful Knowledge from Labor Resources Data in Industrial Construction Projects, *Journal of Management in Engineering* **30**(6): 05014011. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000280.

Horta, I. M. and Camanho, A. S. (2013). Company failure prediction in the construction industry, *Expert Systems with Applications* **40**(16): 6253–6257. http://dx.doi.org/10.1016/j.eswa.2013.05.045.

Huang, C. H. and Hsieh, S. H. (2020). Predicting BIM labor cost with random forest and simple linear regression, *Automation in Construction* **118**(May): 103280. https://doi.org/10.1016/j.autcon.2020.103280.

Huang, Q. (2016). Research on the Engineering project cost management and cost control based on data mining method, *Revista Tecnica de la Facultad de Ingenieria Universidad del Zulia* **39**(12): 370–376. http://doi.org/10.21311/001.39.12.47.

Kim, N., Park, M. and Lee, H.-s. (2007). Performance management method for construction companies, *24th International Symposium on Automation & Robotics in Construction (ISARC 2007)*, pp. 523–528. Available at http://www.iaarc.org/publications/fulltext/isarc2007-5.1_7_124.pdf.

Kitchenham, B. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering, *Biomedical and Environmental Sciences* **13**(1): 37–43. https://doi.org/10.1145/1134285.1134500.

Konikov, A. (2018). A selective study of Information technologies to improve operations efficiency in construction, *MATEC Web of Conferences*, Vol. 170, p. 01110. https://doi.org/10.1051/matecconf/201817001110.

Konikov, A., Kulikova, E. and Stifeeva, O. (2018). Research of the possibilities of application of the Data Warehouse in the construction area, *MATEC Web of Conferences*, Vol. 251, p. 03062. http://doi.org/10.1051/matecconf/201825103062.

Kushiro, N., Fujita, Y. and Aoyama, Y. (2017). Extracting Field Oversees' Features in Risk Recognition from Data of Eyes and Utterances, *IEEE International Conference on Data Mining Workshops, ICDMW*, Vol. 2017-Novem, pp. 590–596. http://doi.org/10.1109/ICDMW.2017.83.

Lam, P. T. I., Wong, F. W. H. and Tse, K. T. C. (2010). Effectiveness of ict for construction information exchange among multidisciplinary project teams, *Journal of Computing in Civil Engineering* **24**(4): 365–376. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000038.

LaPES - Software Engineering Research Laboratory (2012). StArt - State of the Art through Systematic Review Tool. Available at http://lapes.dc.ufscar.br/ferramentas/start-tool.

Lee, J. and Yi, J.-s. (2017). Predicting Project's Uncertainty Risk in the Bidding Process by Integrating Unstructured Text Data and Structured Numerical Data Using Text Mining, *Applied Sciences* **7**(11): 1141. https://doi.org/10.3390/app7111141.

Li, Y. and Zhang, X. (2013). Web-based construction waste estimation system for building construction projects, *Automation in Construction* **35**: 142–156. http://dx.doi.org/10.1016/j.autcon.2013.05.002.

Lu, W., Chen, X., Peng, Y. and Shen, L. (2015). Benchmarking construction waste management performance using big data, *Resources, Conservation and Recycling* **105**: 49–58. http://dx.doi.org/10.1016/j.resconrec.2015.10.013.

Ma, L., Luo, H. B. and Chen, H. R. (2013). Safety risk analysis based on a geotechnical instrumentation data warehouse in metro tunnel project, *Automation in Construction* **34**: 75–84. http://dx.doi.org/10.1016/j.autcon.2012.10.009.

Ma, Z., Lu, N. and Wu, S. (2011). Identification and representation of information resources for construction firms, *Advanced Engineering Informatics* **25**(4): 612–624. http://dx.doi.org/10.1016/j.aei.2011.08.008.

Majrouhi Sardroud, J. (2015). Perceptions of automated data collection technology use in the construction industry, *Journal of Civil Engineering and Management* **21**. http://dx.doi.org/10.3846/13923730.2013.802734.

Martínez-Rojas, M., Marin, N. and Amparo Vila, M. (2012). The Role of Information Technologies to Address Data Handling in Construction Project Management, *Journal of Computing in Civil Engineering* **30**(4): 1–11. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.

Martínez-Rojas, M., Marín, N. and Miranda, M. A. V. (2016). An intelligent system for the acquisition and management of information from bill of quantities in building projects, *Expert Systems with Applications* **63**: 284–294. http://doi.org/10.1016/j.eswa.2016.07.011.

Montaser, A. ; Montaser, A. (2017). *Proceedings of the 34rd ISARC*, Taipei, Taiwan. https://doi.org/10.1080/01446190422000186040.

Moon, S. W., Kim, J. S. and Kwon, K. N. (2007). Effectiveness of OLAP-based cost data management in construction cost estimate, *Automation in Construction* **16**(3): 336–344. https://doi.org/10.1016/j.autcon.2006.07.008.

Negash, S. and Gray, P. (2008). Business Intelligence, *in* F. Burstein and C. W. Holsapple (eds), *Handbook on Decision Support Systems 2*, Springer, Berlin, chapter 2, pp. 175–193.

Panwar, A. and Bhatnagar, V. (2019). Data Lake Architecture, *International Journal of Organizational and Collective Intelligence* **10**(1): 63–75. http://dx.doi.org/10.4018/ijoci.2020010104.

PMBOK (2017). *Um Guia do Conhecimento em Gestão de Projetos*, 6 edn, Project Management Institute, Inc, Newton Square.

Ratajczak, J., Schimanski, C. P., Marcher, C., Riedl, M. and Matt, D. T. (2018). *Cooperative Design, Visualization, and Engineering*, Vol. 3190, Springer International Publishing. https://doi.org/10.1007/b100089.

Rehman, I. and Soomro, T. R. (2020). Proposed Framework for HEC Pakistan Data Warehouse, pp. 1–7. http://dx.doi.org/10.1109/icisct49550.2020.9080040.

Rezaei, A. R., Çelik, T. and Baalousha, Y. (2011). Performance measurement in a quality management system, *Scientia Iranica* **18**(3 E): 742–752. http://dx.doi.org/10.1016/j.scient.2011.05.021.

Rezgui, Y. (2001). Review of information and the state of the art of knowledge management practices in the construction industry, *Knowledge Engineering Review* **16**(3): 241–254. https://10.1017/S026988890100008X.

Riaz, M. N. and Husain, S. A. (2012). Intelligent decision support system for construction project monitoring, *2012 15th International Multitopic Conference, INMIC 2012*, pp. 2–7. https://doi.org/10.1109/INMIC.2012.6511501.

Riaz, M. N., Husain, S. A., Ali, A. and Shamshad, T. (2015). Modified LVQ based clustering analysis for decision making in construction management, *ICOSST 2015 - 2015 International Conference on Open Source Systems and Technologies, Proceedings*, pp. 108–114. https://doi.org/10.1109/ICOSST.2015.7396411.

Ribeiro, W. M., Lima, A. B. and Oliveira, D. (2020). OLAP parallel query processing in clouds with C-ParGRES, *Concurrency Computation* **32**(7): 1–44. http://dx.doi.org/10.1002/cpe.5590.

Rujirayanyong, T. and Shi, J. J. (2005). Company-Wide Project Data Integration for a Construction Organization, pp. 1–10. https://doi.org/10.1061/40754(183)85.

Rujirayanyong, T. and Shi, J. J. (2006). A project-oriented data warehouse for construction, *Automation in Construction* **15**(6): 800−807. https://doi.org/10.1016/j.autcon.2005.11.001.

Seong, H., Choi, H., Cho, H., Lee, S., Son, H. and Kim, C. (2017). Vision-Based Safety Vest Detection in a Construction Scene, *Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC)*. https://doi.org.br/10.22260/isarc2017/0039.

Sharda, R., Sursun, D. and Turban, E. (2014). *Business intelligence and analytics: Systems for decision support*, 10ª edn, Pearson. http://doi.org/10.4324/9781315206455-12.

Shin, D.-p., Park, Y.-j., Seo, J. and Lee, D.-e. (2018). Association Rules Mined from Construction Accident Data, **00**(0000): 1−13. http://doi.org/10.1007/s12205-017-0537-6.

Shin, Y. (2015). Application of boosting regression trees to preliminary cost estimation in building construction projects, *Computational Intelligence and Neuroscience* **2015**. https://doi.org/10.1155/2015/149702.

Szelka, J. and Wrona, Z. (2010). Application of Analytic Databases to Support Decision Making in Structural Engineering / Zastosowanie Analitycznych Baz Danych Przy Podejmowaniu Decyzji W Obszarze Budownictwa Ladowego, *Archives of Civil Engineering* **56**(2): 173−192. https://doi.org/10.2478/v.10169-010-0009-6.

Tam, C. M., Tong, T. K. and Chan, K. K. (2006). Rough set theory for distilling construction safety measures, *Construction Management and Economics* **24**(11): 1199−1206. https://doi.org/10.1080/01446190600879091.

Tianyi, Z., Yong-Cheol, L., Michele, S. and Aurelio, U. (2018). A Supervised Machine Learning-Based Sound Identification for Construction Activity Monitoring and Performance Evaluation, *Construction Research Congress 2018*, Proceedings, pp. 358−366. https://doi.org/10.1061/9780784481264.035.

Tixier, A. J., Hallowell, M. R., Rajagopalan, B. and Bowman, D. (2016). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports, *Automation in Construction* **62**(2016): 45−56. http://dx.doi.org/10.1016/j.autcon.2015.11.001.

Tranfield, D., Denyer, D. and Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review, *British Journal of Management* **14**: 207−222.

Turban, E., Sharda, R., Aronson, J. and King, D. (2009). *Business Intelligence: Um enfoque gerencial para a inteligência do negócio*, Vol. 7, Bookman, Porto Alegre.

Viljamaa, E. and Peltomaa, I. (2014). Intensified construction process control using information integration, *Automation in Construction* **39**: 126−133. http://dx.doi.org/10.1016/j.autcon.2013.08.015.

Vuori, V. (2007). Business intelligence activities in construction companies in Finland-A series of case studies, *Proceedings of the European Conference on Knowledge Management, ECKM*, number November, pp. 1086−1092.

Wang, H. J., Zhang, J. P., Chau, K. W. and Anson, M. (2004). 4D dynamic management for construction planning and resource utilization, *Automation in Construction* **13**(5 SPEC. ISS.): 575−589. https://doi.org.br/10.1016/j.autcon.2004.04.003.

Wang, Q., Xi, L. and Gao, K. (2009). Application of Business Intelligence in the information development of Construction Enterprise, *5th International Conference on Natural Computation, ICNC 2009*, Vol. 6, pp. 212−215. https://doi.org/10.1109/ICNC.2009.674.

Webster, J. and Watson, R. T. (2002). Analyzing the past to prepare for the future : Writing a literature review R, *MIS Quarterly* **26**(2): 13−23. http://dx.doi.org/10.2307/4132319.

Williams, T. P. and Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers, *Automation in Construction* **43**: 23−29. http://dx.doi.org/10.1016/j.autcon.2014.02.014.

Xiaopeng, D., Qiming, L., Dezhi, L. and Zhang, E. (2004). Application of Data Mining in Risk Management of Construction Projects, *Proceedings of the 2004 Internacional Conference on Construction & Real Estate Management*. Available at http://www.paper.edu.cn/scholar/showpdf/NUz2EN2IMTTOQxeQh.

You, Z. and Wu, C. (2019). A framework for data-driven informatization of the construction company, *Advanced Engineering Informatics* **39**(February): 269−277. https://doi.org/10.1016/j.aei.2019.02.002.

Zhang, M., Chen, S., Zhao, X. and Yang, Z. (2018). Research on construction workers' activity recognition based on smartphone, *Sensors (Switzerland)* **18**(8): 1−18. https://doi.org/10.3390/s18082667.

Zhigang, J. and Yajing, L. (2009). The application of RBF neural network on construction cost forecasting, *Proceedings - 2009 2nd International Workshop on Knowledge Discovery and Data Mining, WKKD 2009*, number x, pp. 32−35. http://doi.org/10.1109/WKDD.2009.53.

Zhou, Y. and Ding, L. (2006). Research and application of data mining technology on construction project cost control system, *Proceedings of CRIOCM 2006 International Research Symposium on Advancement of Construction Management and Real Estate*, Vol. 1, pp. 471−476. Available at https://www.irbnet.de/daten/iconda/CIB5697.pdf.