Revista Ibero-Americana de Estratégia
Iberoamerican Journal of Strategic Management

e-ISSN: 2176-0756
doi.org/10.5585/riae.v21i1.17689
Received: 16 July 2020 / **Approved:** 03 May 2022
Evaluation Process: Double Blind Review
Editor in Chief: Heidy Rodriguez Ramos
Co-editor: Ivano Ribeiro

COPE JM14473

Check for updates

HOW EXPERIMENTAL AND STRATEGIC ARE BUSINESS INTELLIGENCE (BI) AND DATA MINING APPLICATIONS?

Rodrigo Fontes Cruz[1] **Methanias Colaço Júnior**[2] **Victor Menezes Gois**[3]

[1] *Mestre em Ciência da Computação, Universidade Federal de Sergipe – UFS. São Cristóvão, Sergipe – Brasil.* rodrifontes@gmail.com
[2] *Doutor em Ciência da Computação – Pós Doutor em Gestão, Universidade Federal de Sergipe – UFS. São Cristóvão, Sergipe – Brasil.* mjrse@hotmail.com
[2] *Bacharel em Sistemas de Informação, Universidade Federal de Sergipe – UFS. Itabaiana, Sergipe – Brasil.* victormenezes.gois@gmail.com

Abstract

Objective: Identify and characterize the methodologies used for the experimental development of intelligent applications aligned with strategic planning.
Methodology: A systematic mapping was carried out to characterize the research in the area, considering the last ten years.
Originality: No scientific studies were found with the same research object of this article, to identify and characterize the methodologies for the experimental development of intelligent applications aligned with strategic planning, which increases the importance of the results presented here.
Main results: As a result, no studies were found that presented any complete approach to discipline strategic alignment and experimentation, providing clear compliance with strategic objectives and an experimental phase in the validation of results. However, some trials of parts of these characteristics could be mapped, such as experimentation found in 28,57% of the studies. Among the countries, China, the United States and Brazil led the ranking of publications on the subject. As for the medium of publication, Journal was the most used option for publication. In addition, the "IEEE International Conference on Advanced Communications, Control and Computing Technologies" and the journal "Expert Systems with Applications" stood out as major publishers.
Theoretical contributions: This research presents results relevant to academia and entrepreneurs, providing evidence that there is a gap in research on a formal method of BI and Data Mining applications experimental and strategy-driven development. In addition, this work is presented as a source of consultation to the existing method standards for the development of intelligent applications, as well as being replicable and extended by the applied systematization. Finally, there is a focus on research that proposes methods of creating experimental applications validated experimentally and aligned with strategy.

Keywords: Strategic Alignment. Business Intelligence. Data Mining. Data Science.

QUÃO EXPERIMENTAIS E ESTRATÉGICAS SÃO AS APLICAÇÕES DE BUSINESS INTELLIGENCE (BI) e DATA MINING?

Resumo

Objetivo do trabalho: Identificar e caracterizar as metodologias utilizadas para o desenvolvimento experimental de aplicações inteligentes alinhadas ao planejamento estratégico.
Metodologia: Um mapeamento sistemático foi realizado, para caracterizar a pesquisa na área, considerando os últimos dez anos.
Originalidade: Não foram encontrados trabalhos científicos com o mesmo objeto de pesquisa deste artigo, de identificar e caracterizar as metodologias para o desenvolvimento experimental de aplicações inteligentes alinhadas ao planejamento estratégico, o que aumenta a importância dos resultados aqui apresentados.
Principais resultados: Como resultados, não foram encontrados trabalhos que apresentassem alguma abordagem completa para disciplinar o alinhamento estratégico e a experimentação, prevendo atendimento claro aos objetivos estratégicos e uma fase experimental na validação dos resultados. No entanto, alguns ensaios de partes dessas características puderam ser mapeados, como, por exemplo, a experimentação, encontrada em 28,57% dos trabalhos. Entre os países, a China, os Estados Unidos e o Brasil lideraram o ranking de publicações sobre o tema. Quanto ao meio de publicação, o Journal foi a opção mais utilizada para publicação. Além disso, a conferência "IEEE International Conference on Advanced Communications, Control and Computing Technologies" e o periódico "Expert Systems with Applications", destacaram-se como maiores publicadores.
Contribuições teóricas: Esta pesquisa apresenta resultados relevantes à academia e aos empreendedores, fornecendo evidências de que há uma lacuna nas pesquisas sobre um método formal de desenvolvimento experimental e dirigido à estratégia de aplicações de BI e Data Mining. Além disso, este trabalho apresenta-se como uma fonte de consulta aos padrões de métodos existentes para o desenvolvimento de aplicações inteligentes, bem como pode ser replicado e estendido, pela sistematização aplicada. Por fim, há o direcionamento para pesquisas que proponham métodos de criação de aplicações inteligentes validadas experimentalmente e alinhadas à estratégia.

Palavras-chave: Alinhamento Estratégico. Business Intelligence. Data Mining. Mineração de Dados. Data Science. Ciência de Dados.

¿CUÁN EXPERIMENTALES Y ESTRATÉGICAS SON LAS APLICACIONES DE INTELIGENCIA EMPRESARIAL (BI) Y MINERÍA DE DATOS?

Resumen

Objetivo del trabajo: Identificar y caracterizar las metodologías utilizadas para el desarrollo experimental de aplicaciones inteligentes alineadas con la planificación estratégica.
Metodología: Se realizó un mapeo sistemático para caracterizar la investigación en el área, considerando los últimos diez años.
Originalidad: No se encontraron estudios científicos con el mismo objeto de investigación de este artículo, para identificar y caracterizar las metodologías para el desarrollo experimental de aplicaciones inteligentes alineadas con la planificación estratégica, lo que aumenta la importancia de los resultados presentados aquí.
Principales resultados: Como resultado, no se encontraron estudios que presentaran un enfoque completo para disciplinar la alineación estratégica y la experimentación, proporcionando un cumplimiento claro de los objetivos estratégicos y una fase experimental en la validación de los resultados. Sin embargo, algunos ensayos de partes de estas características podrían mapearse, como la experimentación encontrada en el 28,57% de los estudios. Entre los países, China, Estados Unidos y Brasil lideraron el ranking de publicaciones sobre el tema. En cuanto al medio de publicación, Journal fue la opción más utilizada para la publicación. Además, la "IEEE International Conference on Advanced Communications, Control and Computing Technologies" y la revista "Expert Systems with Applications" se destacaron como las principales editoriales.
Contribuciones teóricas: Esta investigación presenta resultados relevantes para la academia y los empresarios, y proporciona evidencia de que existe una brecha en la investigación de un método formal para el desarrollo experimental de aplicaciones de BI y minería de datos centradas en la planificación estratégica de una organización. Además, este trabajo se presenta como una fuente de consulta a los estándares de métodos existentes para el desarrollo de aplicaciones inteligentes, además de ser replicable y extendido por la sistematización aplicada. Finalmente, hay un enfoque en la investigación que propone métodos para crear aplicaciones experimentales validadas experimentalmente y alineadas con la estrategia.

Palabras-claves: Alineamiento Estratégico, Business Intelligence. Data Mining. Minería de datos. Procesamiento de datos. Data Science. Ciencia de los datos.

Cite as / Como citar

American Psychological Association (APA)
Cruz, R. F., Colaço Júnior, M., & Gois, V. M. (2022). How experimental and strategic are Business Intelligence (BI) and data mining applications? *Iberoamerican Journal of Strategic Management (IJSM), 21*, 1-34, e20638. https://doi.org/10.5585/riae.v21i1.17689.

(ABNT – NBR 6023/2018)
CRUZ, Rodrigo Fontes; COLAÇO JÚNIOR, Methanias; GOIS, Victor Menezes. How experimental and strategic are Business Intelligence (BI) and data mining applications? *Iberoamerican Journal of Strategic Management (IJSM),* v. 21, p. 1-34. 2022. https://doi.org/10.5585/riae.v21i1.17689.

1 Introduction

With the inevitable market mutation in the business scope, knowing how to obtain positive results amid these changes and developing adequate and effective processes to manage the transformations properly has become an obligation. Wrong decisions, whether strategic, tactical, or operational, can cost the company's future, as well as a correct one, defining its survival or expansion (Côrte-Real, Oliveira & Ruivo, 2017).

In this context, data emerge as a crucial source to winning competitive advantage (Kubina, Varmus & Kubinova, 2015). This advantage is based on the knowledge obtained from data analysis and has boosted areas such as Business Intelligence (BI) and Data Mining (DM). This associated with the current Big Data era, in which technological progress drives the creation of large volumes of data at high speed from a variety of sources, has further justified the investment in DM, whose power and automaticity have made it possible to deal with large amounts of data and extracting value (Shmueli et al., 2017).

However, before any attempt to extract this helpful knowledge is possible, a general approach that describes how to extricate knowledge needs to be established (Kurgan & Musilek, 2006). In this way, several models of Data Mining processes have been proposed by researchers and professionals. Author examples include Fayyad et al. (1996), Cabena et al. (1998), Cios et al. (2000), CRISP-DM (2003), Berry & Linoff (1997), Sharma, Osei-Bryson & Kasper (2012), and Ławrynowicz & Potoniec (2014).

Furthermore, as the organization's strategic objectives need to be translated to the lowest levels of the business, this methodology must be planned and executed from within the organization. In other words, it is necessary to bridge the gap between the business strategy and its implementation at the software project level (Basili et al., 2010; Basili et al., 2014). Along the same lines, Mandic et. al. (2010) and Astley et. al. (2017) emphasize that analyzing all decision-making relevant aspects is necessary to implement methods for integrating existing data with strategic goals.

Despite this common-sense among several researchers, a survey performed in Brazil (Lima et al., 2017) showed that 72% of companies do not use a specific method for developing BI applications in line with the organization's strategic planning. Another piece of evidence found by the survey highlights that 67.50% of respondents do not use a formal methodology for the development of BI applications. In this case, it is worth mentioning that despite the literature and practice conceptually separating the areas of Data Mining and BI, there is a strong convergence and integration of these, since the "I" or, Intelligence of BI, can only be achieved with the application of techniques of Data Mining, which include Machine Learning. It indicates that the absence of these alignment methodologies should also affect Data Mining projects since there are BI projects without Data Mining, Data Mining without BI, and, in the best case, which will be considered in this proposal, a complete BI, which used integration of analytical data - BI -, databases, statistics and artificial intelligence - Data Mining -.

Given these data, the objective of this work is to identify, gather, organize and synthesize the enormous possible number of scientific evidence on methodologies for the experimental development of intelligent applications aligned with strategic planning, which allows characterizing both the area of study and forming a basic theory on the topic. In addition, the characterization result can help identify new research opportunities, besides directing efforts towards more appropriate strategies in the study area.

Thus, a secondary study of the Systematic Mapping type was planned and executed. It should be noted that, to date, no mapping on the mentioned subject has been published or is available for academic research.

After answering the research questions, it identified that there were no works that presented any approach to discipline the strategic alignment in the development of BI and Data Mining applications, as well as it was not possible to identify any methodology that aimed at experimentation, in other words, which provides for an experimental phase in the validation of results. Regarding the types of studies, "Practical Application" stood out, with 47.61% of publications and "Controlled Experiment" with 28.57%. Among the countries, China, the United States and Brazil led the ranking of publications on the subject. As for the means of publication, the Journal was the most used option for publication. In addition, the "IEEE International Conference on Advanced Communications, Control and Computing Technologies" and the journal "Expert Systems with Applications" stood out as the best publishers.

The rest of this article is structured as follows. In section 2, related works on the topic are presented. In section 3, we have the conceptual basis for understanding this work. In section 4, the method adopted in this work is discussed. Section 5 presents the mapping process. In section 6, we have the results achieved. Section 7 critically analyzes the results and discusses lessons learned. In section 8, the validity threats encountered are detailed. And finally, in section 9, the conclusion is presented.

2 Related works

No scientific works were found with the same research purpose of this article to identify and characterize the methodologies for the experimental development of intelligent applications aligned with strategic planning, which increases the result importance presented here. However, some works mention the strategic alignment importance in application development.

Regarding the works evaluated by this mapping, in other words, articles that dealt mainly with Business Intelligence and Data Mining, some papers highlighted the importance of strategic alignments, such as the works of Sharma, Osei-Bryson & Kasper (2012) and Kohavi et al. al. (2013). In addition, many papers in the literature highlight the importance of strategic alignment of information technology applications in general. Examples are the works of: Isaca (2018), Weber & Klein (2013), Medeiros Júnior et al. (2017) and Araújo & Dornelas (2017). Next, other relevant works to the specific theme discussed here will be summarized.

Mola et al. (2015) carried out an exploratory study that analyzes the effects of technical and organizational characteristics of BI systems on knowledge sharing, collaboration, and decision-making processes. On average, the technical and organizational features of BI Systems are positively associated with an increase in knowledge sharing, leading to an improvement in internal collaboration and decision-making quality. These improvements depend on how BI is designed.

In a survey carried out in Brazil by Lima et al. (2017), it was found that 67.50% of the companies do not use an experienced methodology for BI development, which contributes to the project's unsuccessfulness. Associated with this result, 72.00% of the companies do not use a strategic alignment methodology. The methodology's absence aligned with the company's strategy shows that managers may be making decisions based on information that is not relevant to the institution or misaligned with business strategies.

In another similar survey on Business Intelligence, Duan & Xu (2012) present an introduction to BI, with fundamental algorithms for BI use in business environments, highlighting the challenges and opportunities discovered in these environments.

Colaço Júnior et al. (2019) presented a process that merges the GQM+Strategies approach with an agile development methodology for Business Intelligence applications proposed by the author, aiming to ensure strategic alignment. The proposed procedure was evaluated through a case study in a Latin American multinational company in the retail market, in which it was shown that it is possible to integrate the strategic alignment approach adopted with a methodology for developing BI applications. With good initial evidence, researchers will be able to evolve the process and predict the experimentation usage to validate intelligent models created with Data Mining and AI techniques.

3 Conceptual basis

In this section, some concepts necessary for the understanding of this work are introduced.

3.1 Strategic alignment

In the literature, many concepts detail the meaning of strategic alignment. However, we look for those most closely related to the subject of our proposal. (1) Alignment between the strategic business plan (SBP) and the strategic information technology plan (SITP) is achieved when the set of systems strategies (objectives, obligations, and strategies) is derived from the organizational strategic set (mission, goals, and strategies) (King, 1988); (2) The SBP-SITP link corresponds to the degree to which the IT mission, objectives, and plans reflect, support and are supported by the mission, objectives and business plans (Reich & Benbasat, 1996); (3) It is the way business and IT work together to achieve the common goal (Campbell, 2005) and (4) The alignment between SBP-SITP is the adequacy of the strategic orientation of the business with the IT (Chan et al. al., 1997).

For Lima (2017), the problem for organizations today is that companies are not always able to declare their strategic objectives explicitly or enough so that it can be verified if such aims have achieved the goals aligned with IT. This challenge is not having the IT area as a support, but as part of a business platform, serving as an essential element of the business strategy. This new vision must link the strategic alignment of IT to the organization's business. These two elements need to be related to each other in search of continuous improvement and the organization's success.

Today, one of the most crucial IT strategies is the adoption of Business Intelligence and Data Mining, whose main characteristic is the ability to handle large amounts of data and extract value. A question that can be asked is: how is this alignment made between Strategic Planning and the development of this type of application? Next, we will detail the concept of BI and Data Mining and their relationship with Strategic Planning.

3.2 Business intelligence

Business Intelligence – BI – can be understood as a set of methodologies, processes, architectures, and technologies used to support the collection, analysis, presentation, and dissemination of business information to enable more effective strategic, tactical and operational decision-making (Hans et al. ., 2013; Dedić & Stanier, 2017).

BI helps companies think better about the competition through a better understanding of the customer base (Brannon, 2010), which can lead to the creation of a closer and stronger relationship with customers and increase revenue (Brannon, 2010). Alexander, 2014). In addition, it plays a critical role for business in terms of organizational development, providing a competitive advantage, in the context of achieving positive information asymmetry (Thamir & Poulis, 2015), and contributes to optimizing business processes and resources, maximizing profits, and improve proactively, as well as for strategic decision making (Dedić & Stanier, 2016

In this way, Business Intelligence in organizations is understood as a strategic advantage (Chaudhuri et al., 2011; Kohtamäki & Farmer, 2017), regardless of the area in which the organization operates, whether private or not, because, at present, the organizations that use systems of this type find it easy to acquire specific knowledge about the various factors that influence it, being able to later apply such knowledge, identifying the market potential and, with this, the directing it in its strategy, vision, and goals to be achieved.

In addition to its strategic and tactical advantages, Business Intelligence is also used at the operational level to enable various types of users to identify emerging trends, make faster decisions, take action and deal with organizational issues as they arise. Its purpose is to help stakeholders better understand their organization's operations and make wiser and more informed business decisions, as well as manage operational performance (AICPA, 2015).

Thus, Business Intelligence refers to the act of providing businesses with the necessary support for decision-making through the use of a set of techniques and tools. Data mining can also compose a BI solution.

3.3 Data mining

Data mining (DM) is the data analysis process to extract information and knowledge that is not clearly visualized in common systems. It seeks to withdraw information and knowledge through relationships between data that allow inferences about what may occur (predictive analysis) or correlations between what has already happened(Barbieri, 2011).

Technically speaking, it consists of a process that uses statistical, mathematical, and artificial intelligence techniques to extract and identify useful information and, consequently, knowledge (or patterns) from large volumes of data, and these patterns can be presented as trends, rules of business, correlations or predictive models. These models and standards can be used to guide the decision-making process, as well as predict the effect of these choices (Laudon, Laudon & Marques, 2004).

In data mining, we have the tasks and techniques that guide what we want and how to obtain such information. The tasks indicate the objectives, i.e., what information is intended to be obtained. Mining techniques point to the methods used to obtain desired patterns (Camilo & Silva, 2009). Duan & Xu (2012) classify mining techniques into two broad groups: supervised learning and unsupervised learning. Supervised learning methods build models to predict an unknown attribute according to observed attributes, while unsupervised learning methods extract patterns such as clusters, process graphs, and correlations between the data.

Supervised methods require a dataset that has a pre-defined main target variable and the records are categorized concerning this, i.e., the precisely programmed rule is applied. In unsupervised methods, there is no need to pre-categorize the records, which means that an attribute is not needed as the main target (Filho et al., 2007).

The most common supervised learning techniques are classification (which can also be unsupervised) and regression (Mccue, 2007). However, in data mining, several techniques may be applied or even that there is a combination, aiming at the search for the best-expected result (Costa et al., 2012). Evaluating these techniques or any software product experimentally is a process to support quality assurance, as can be seen in the next section.

3.4 Software experimentation

Quickly delivering value to customers is a top priority for software companies (Fagerholm et al., 2017). With this goal in mind, companies often develop their development practices. Initially, they inherit the agile principles in the organization development part (Martin, 2002) and expand them to other departments (Olsson, Alahyari & Bosch, 2012). Next, companies focus on various lean concepts, such as eliminating waste, removing constraints in the development pipeline (Goldratt & Cox, 2016), and moving towards continuous integration (Dittrich et al., 2018) and deployment of software functionality (Rodríguez et al., 2017). However, continuous deployment is characterized by a two-way channel that allows companies not only to send data to their customers for rapid prototyping (Singh, 2016) but also to receive feedback data from products in the field.

In this context, software development companies' intuition about customer preferences can be wrong up to 90% of the time (Clancy, 1995; Castellion, 2008; Manzi, 2019). To mitigate this, current product usage data has the potential to make the prioritization process in the development of new products more accurate, as it allows focusing on what customers do rather than what they say (Bosch-Sijtsema & Bosch, 2015). In this sense, experimentation is becoming the norm in advanced software companies for reliable evaluation of ideas with customers to correctly prioritize product development activities (Olsson & Bosch, 2014; Kohavi & Longbotham, 2017).

Thus, there is a growing understanding in the community that empirical studies are needed to develop or improve processes, methods, and tools for software development and maintenance (Basili, 1996; Endres & Rombach, 2003; Fagerholm et al., 2017).

Empirical software engineering research should aim to acquire general knowledge about which technology (process, method, technique, language, or tool) is helpful, for whom it is useful, in which tasks (software engineering), and in which environments. Therefore, this research focuses on the type of technology being studied in the investigated experiments (which reflects the topics of the experiments), the subjects who participated, the tasks they performed, the type of application systems on which these tasks were performed, and the environments in which the experiments were conducted. In addition, it should also include data on replication of experiments and the extent to which internal and external validity is discussed (Sjøberg et al., 2005).

An important category of the empirical study is the controlled experiment, whose conduct is governed by the classical scientific method, to identify cause-effect relationships. In a controlled experiment, users are randomly divided between variants (for example, the two different designs of a product interface) in a persistent manner (a user receives the same experience multiple times). User interactions with the product are instrumented and key metrics are computed (Kohavi & Longbotham, 2017).

One of the main challenges with metrics is deciding what to include in an Overall Evaluation Criteria (OEC). An OEC is a quantitative measure of the controlled experiment objective (Roy, 2001) and guides the business development direction. In controlled experimentation, it is intuitive to measure the short-term effect, i.e., the impact observed during the experiment (Hohnhold, O'brien & Tang, 2014). Giving more weight to advertising metrics, for example, makes companies more profitable in the short term. However, the short-term effect is not always predictive of the long-term effect and therefore should not be the only component of an OEC (Kohavi et al., 2014). Defining an OEC is not trivial and must be done with great care. Kohavi et al. (2009, 2014, 2017), in their work, present common pitfalls in the process of establishing a controlled experimentation system and guidelines on how to reliably define an OEC. This work identified BI and Data Mining applications that were experimentally validated.

4 Method

To identify and characterize the methodologies for the Business Intelligence Development and Data Mining applications aimed at strategy or which anticipate experimental evaluation, Literature Systematic Mapping Methodology was adopted for this work. According to Kitchenham (2004), Petersen et al. (2008), and Petersen et al. (2015), a mapping consists of a systematic protocol for the search and selection of relevant studies, intending to extract data and map the results to a specific research problem.

This method was adopted because it is based on the concepts of Evidence-based Medicine, a mature area in the process of systematic reviews, and because it proposes a paradigm shift in how research in the Software area should be conducted. According to Kitchenham et al. (2009), research in medicine has changed drastically with the evidence-based paradigm, enabling a more effective organization of medical research and supporting the clinical judgment of experts. The success of this new paradigm strongly influenced the adoption of the evidence-based approach in other areas of knowledge such as psychology, nursing, social sciences, education, and computing.

The following sections detail the steps for replicating this mapping, encompassing the strings and commands used, the bases searched, and the criteria for selecting articles and extracting data. For extraction, it was explicitly described how the works were classified as experimental or strategic, as well as the variables and their values were identified. The summarized phases of the method are listed below.

1. Planning: the research objectives are listed and the review protocol is defined;
2. Conduct: in this activity, the sources for the mapping are selected, and the primary studies are identified, selected, and evaluated according to the inclusion, exclusion, and quality criteria established during the mapping protocol;
3. Analysis and Publication of Results: data from studies are extracted and synthesized for publication.

4.1. Systematic mapping planning

4.1.1. Objective

As mentioned, this mapping aims to identify and characterize the methodologies for developing Business Intelligence and Data Mining applications aimed at strategy or that provide Experimental evaluation.

4.1.2. Research questions

The research questions were designed based on the PICO approach (Bergin & Wraight, 2006; Costa et al., 2007). This model, which emerged in the execution of clinical studies, structures the research in four elementary elements: Population, Intervention, Comparison (or control), and "Outcomes" (Results). Despite having emerged in medicine, this strategy can be adapted to other areas, as demonstrated by the Kitchenham work(2004), utilized as a reference for the production of Table 1. The table also presents some control articles, which were chosen with care. A preliminary consultation in the research bases, with the selection of the first two found articles, aligned with the intervention of interest. This exploratory search served as a basis to find other crucial keywords, present in true positive articles that will be the final results of the mapping, and refine the search string. A control article may already be known to researchers and is an element to help create and validate the string. In this sense, a control article, which is effectively a comparison, was also used, i.e., it did not entirely fit the interest intervention and is not indexed in the researched bases but presents a proposal for strategic alignment. Control articles of this type also help establish a more assertive search string.

Table 1

PICO model for research question compliance

Acronym	Definition	Description
P	Population	Publications by researchers and developers, with a view to developing Business Intelligence applications with or without Data Mining support.
I	Intervention	Strategy-driven BI and Data Mining application development methods addressed to strategy or that used experimental evaluation. Strategic direction presupposes procedures that start from the strategic objective identification that will be leveraged by the final product, avoiding the solution creation that is not aligned with the organizational strategy and does not directly help the outlined objective achievement.
C	Control	Methods of BI applications creation or knowledge discovery, such as the Data Mining process model: Cross Industry Standard Process for Data Mining (CRISP-DM). Control article Articles that fit the intervention:: • *Evaluation of an integrated knowledge discovery and data mining process model*; • *Pattern based feature construction in semantic data mining.* Comparison article: • Proposta e Avaliação de um Processo para o Desenvolvimento de Aplicações de Business Intelligence Dirigido à Estratégia
O	Result	BI and Data Mining methodologies validate their conclusions through controlled experiments or explicitly define the application's direction toward one or more strategic objectives.

Source: Produced by the authors.

Thus, based on the definition of PICO, the following research questions were elaborated:

- Q1: What are the strategy-oriented methodologies used in the BI and Data Mining application development?
- Q2: How is this alignment between Strategic Planning and the development of BI and Data Mining applications?
- Q3: Do BI and Data Mining development methodologies foresee an experimental evaluation phase with conclusions validation through appropriate statistical tests?
- Q4: Which countries have more researchers who is publishing on this topic?
- Q5: Which years had the most publications in this area?
- Q6: What are the most important journals and conferences on the topic?
- Q7: What are the publishing ways?
- Q8: What study types?

These questions were formulated based on the guidelines of the Systematic Literature Mapping protocol (Kitchenham, 2004; Petersen et al., 2008; Petersen et al., 2015). It is noteworthy that the differences between a systematic literature mapping and a systematic literature review are fundamentally made explicit in the research question formulation. According to Petersen et al. (2008), systematic mapping has an exploratory research question and requires less depth in data extraction. In addition to the overview, it was intended to verify if the methodologies used explicitly related the application requirements with strategic objectives and if they had a phase to validate the solutions experimentally.

4.1.3. Search and selection strategy

To perform the search, the following databases were consulted: ACM Digital Library (ACM), IEEE Xplore (IEEE), and SCOPUS. The searches were performed using the filtering tools available in each database mentioned above, considering in the search: title, résumé/abstract, and keywords of the respective articles. Regarding the language, only works in English were selected. Regarding the area, only works related to Computer Science were chosen. And regarding the time of publication, only works published after 2008 were selected.

A search string defined with the use of English terms and several synonyms to accomplish the research in the digital databases, associated with the assumption that the studies would be in the areas of computing that deal with methodologies for developing BI and Data Mining applications. Such terms were identified with the help of the PICO model control articles, described in the previous section (Table 1), and later refined and adapted for better use of the string. Table 2 shows the terms, before refining them, that were selected.

Table 2

PICO model categories and terms identified for bibliographic research before refining them

Category	Description
Population	Development, Implementation, Construction, Deployment, Creation, Business Intelligence, Data Mining.
Intervention	Methodology, Method, Approach, Process, Experiment.
Control	Métodos de criação de aplicações de BI ou de descoberta de conhecimento, como, por exemplo, o CRISP-DM (sem *strings*).
Result	Strategy Oriented, Strategy Driven, Strategic Alignment, Hypothesis Testing, Statistical Validation, Statistical Analysis, Control Experiment, Controlled Experiment, Experimental Analysis, Experimental Evaluation, Statistical Test, Formal Experiment, Null Hypothesis, Primary Hypothesis, Statistical Significance.

Source: Produced by the authors.

After refinement, the adjusted terms were used to build the search string which is described in Table 3.

Table 3

Strings chosen after refinement

Search string terms		
Development, Implementation, Construction, Deployment, Creation, Business Intelligence, Data Mining.	Methodology, Method, Approach, Process, Experiment.	Strategy Oriented, Strategy Driven, Strategic Alignment, Hypothesis Testing, Statistical Validation, Statistical Analysis, Control Experiment, Controlled Experiment, Experimental Analysis, Experimental Evaluation, Statistical Test, Formal Experiment, Null Hypothesis, Primary Hypothesis, Statistical Significance.

Source: Produced by the authors.

- The search string generated with the terms highlighted above was:

TITLE-ABS-KEY(("Method" OR "Approach" OR "Process" OR "Experiment*") AND ("Development" OR "Implementation" OR "Construction" OR "Deployment" OR "Creation") AND ("Business Intelligence" OR "Mining") AND ("Strategy Oriented" OR "Strategy-Oriented" OR "Strategy Driven" OR "Strategy-Driven" OR "Strategic Alignment" OR "Hypothes* Test*" OR "Statistic* Valid*" OR "Statistic* Analy*" OR "Contro* Experiment*" OR "Experiment* Analy*" OR "Experimen* Evaluation" OR "Statisti* Test*" OR "Formal Experiment*" OR "Null Hypothes*" OR "Primary Hypothes*" OR "Statisti* Significan*")) AND SUBJAREA(COMP) AND (PUBYEAR > 2008).*

The search for research in computing was considered based on control articles that covered more than one area, since computing is a means for management and decision making, and, in addition, it was considered that the objective was to explore the area that develops the technical part of Information Technology, which should already have strategic alignment and experimentation as a standard, in other words: Is this already a reality?

4.1.4 Source selection criteria

To filter the articles relevant to this mapping, inclusion, and exclusion criteria were established. After all the phases, the study counted the papers that focused on the use of some methodology for the BI and DM applications development centered on strategic planning or that used experimental evaluation, using the following preliminary inclusion criteria:

1. The result must contain the subject of this study, already automatically limited by the string, in the title, abstract, or keywords;
2. The result must date between the years 2009 to 2019;
3. The result must present an experimental evaluation or explore some methodology, method, process, or approach for developing BI or Data Mining applications. It was done to manage the risk of not having any strategy-driven or experimental approach and at least listing works that addressed BI or Data Mining development methodologies or used some experimental evaluation;
4. The result must be available for online consultation.

The inclusion criteria confirmation was given after analyzing the abstract of each of the articles found, in the first filtering, followed by a complete reading for the second filtering. Before the first reading of each paper, an analysis was performed regarding the exclusion criteria. Were deleted:

1. Secondary studies, as they deal with third-party approaches;
2. Duplicate publications;
3. Short Papers;
4. Surveys.

4.1.5 Information extraction strategy

The data extraction strategy is delineated to gather the necessary information and answer the research questions, assessing the quality of the work. That said, after the selection stage is over, the defined works will be read in full, then a form will be answered, as shown in Table 4, containing information about the content covered in each work. Therefore, this form allows us to evaluate the works in a more detailed and accurately, performing the classification based on the inclusion and exclusion criteria.

Regarding strategic alignment identification, it was not enough to quote the term or highlight its importance at some point in the article. Thus, to be classified as a work that used some alignment approach, it was necessary for the solution proposed by the authors to consider as a premise or basis at least one explicit strategic objective.

Table 4

Extraction form

1.	What is the methodology used?	
2.	How is strategic alignment done?	
3.	Was the article referring to Business Intelligence?	[Yes, No]
4.	Was the article referring to Data Mining?	[Yes, No]
5.	What type of study was used?	[Practical Application, Case Study,...]
6.	Was the study strategy-oriented?	[Yes, No]
7.	Does the study have any experimental evaluation?	[Yes, No]
8.	Was the sample size calculated?	[Yes, No]
9.	Was the normality test done?	[Yes, No]
10.	Was a hypothesis formally stated?	[Yes, No]
11.	Was the confidence interval calculated?	[Yes, No]
12.	Have threats to validity been declared?	[Yes, No]

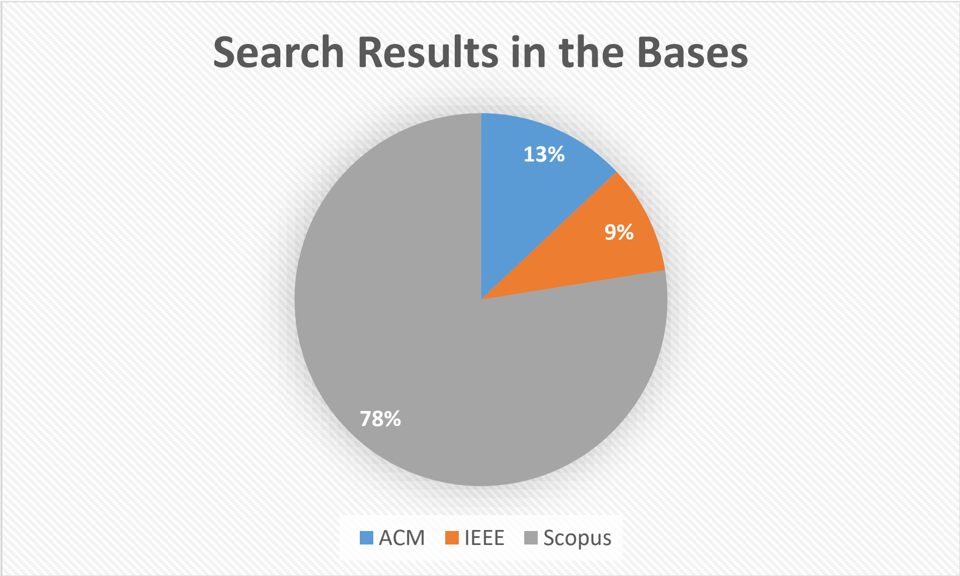
Source: Produced by the authors.

5 Conducting the systematic mapping

The systematic mapping planning was prepared between February and March 2019, while the execution occurred in April of the same year. It was necessary to form the search string from combinations of keywords in English to obtain the primary studies. Thus, the defined base string in the Scopus search engine, refined, and, when judged that the string was adequate, it was translated for the ACM and IEEE search engines. Soon after, searches were performed. In total, 841 works were returned, 652 (78%) from Scopus, 109 (13%) from ACM, and 80 (10%) from IEEE, as shown in figure 1.

Figure 1

Execution result of search strings in digital libraries



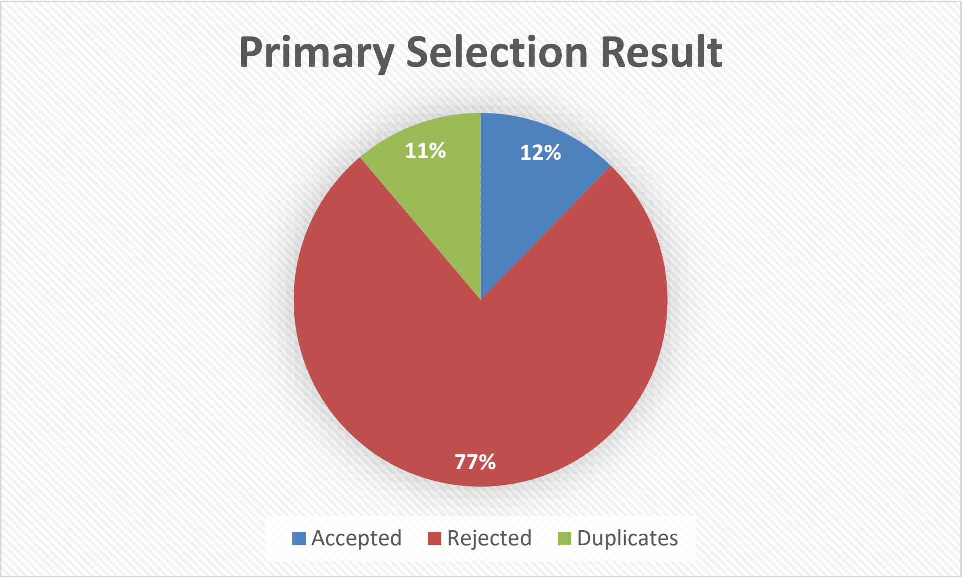
Source: Produced by the authors.

With the end of the search, the filtering of the articles found began, based on the selection criteria defined in section 4.1.4. At this stage, the works were classified as Accepted, Rejected, and Duplicate.

In this sense, of the total of 841 publications analyzed, 94 (11% of the total) were duplicates and, consequently, the equivalent ended up being eliminated. Thus, 747 works were designated for a peripheral evaluation, in which all titles and abstracts were read, applying the inclusion and exclusion criteria previously defined in the protocol. At this stage end, 644 (77%) works were identified that were outside the scope of this mapping and were therefore rejected. These works were short articles, literature reviews, surveys, or non-criteria articles. Those accepted mentioned experimenting performance or the use or proposal of an approach, process, methodology, or method. As a final result, 103 (12%) accepted works were evaluated precisely. Figure 2 illustrates the results of this first selection.

Figure 2

The first stage result of work selection



Source: Produced by the authors.

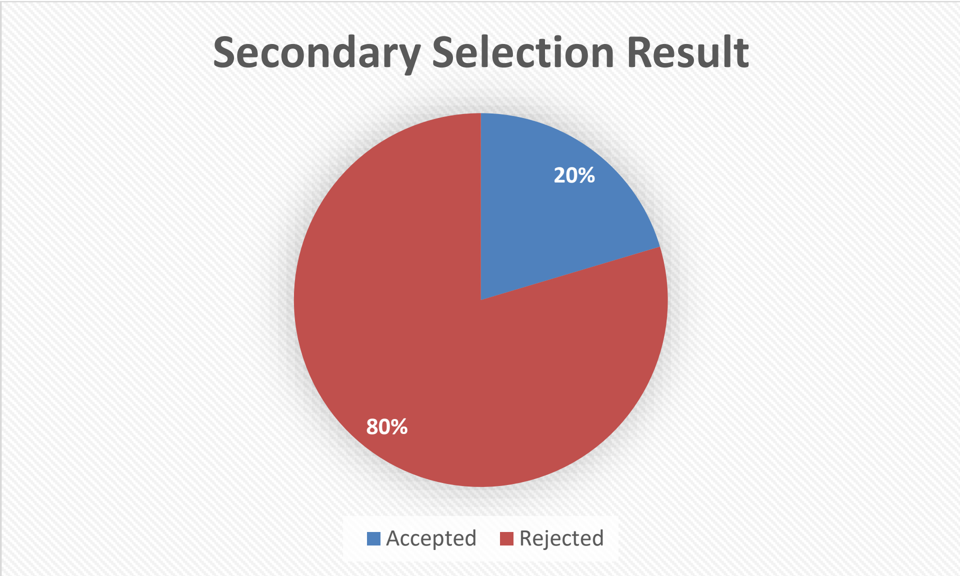
After this first selection stage, a precise and detailed assessment is performed based on the complete work reading to confirm the indication of strategic alignment or experimental evaluation. However, it was not possible to access all publications, even after contacting the authors by email. From the 103 articles that should have been assessed at this stage, 100 (97.09%) were retrieved. Thus, it was not possible to access the full text of 3 (2.91%) articles, which consequently ended up being rejected.

To classify the studies as controlled experiments and organize the experimental evaluation presence, it was considered the studies in which the verified scientific method basis, that is, those that formalized the hypotheses definition and carried out the statistical validation necessary for testing.

During this stage, it was identified that 82 (80%) works were outside of this systematic mapping scope and therefore rejected. All articles accepted again confirmed the experiment's performance. Apart from these, only articles in which the method, methodology, process, or approach mentioned related to the BI or Data Mining applications design were accepted. In other words, BI or Data Mining was not a transversal subject of the article. As a final result, 21 (20%) works were accepted. Finally, for these articles, a new reading guided the completion of the extraction form, described in section 4.1.5. Figure 3 illustrates the final result of this second phase of data selection and extraction.

Figure 3

The second stage result of work selection



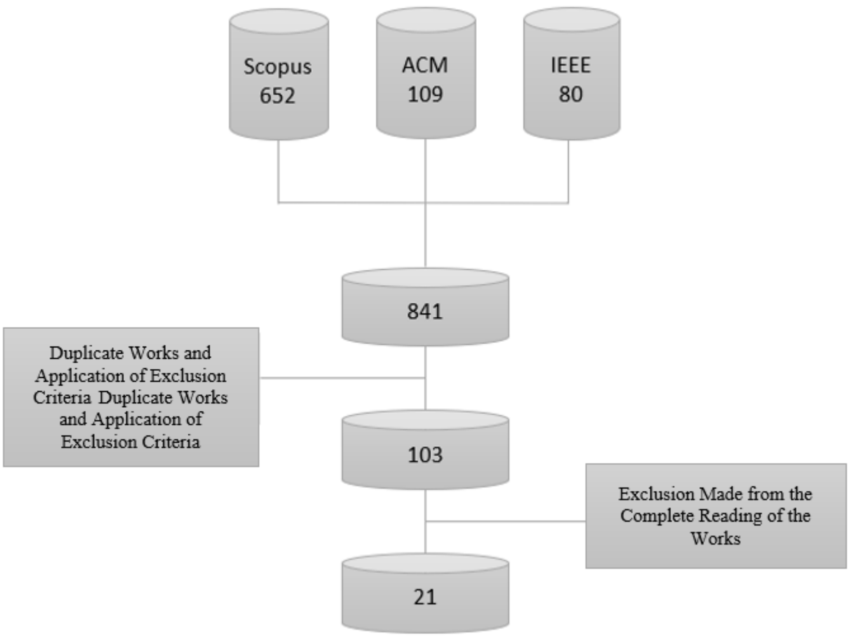
Source: Produced by the authors.

6 Data summary and result presentation

In this section, the systematic mapping results are presented. Figure 4 shows a summary of the number of works obtained in each step of the search process and then the research questions are answered according to the extracted data.

Figure 4

Results obtained during the search process



Source: Produced by the authors.

6.1 What strategy-driven methodologies are used in the bi and data mining applications development?

No works were found that presented any approach to discipline strategic alignment. However, some works mention the alignment importance or propose some methodology for the efficient development of this application type, such as Sharma, Osei-Bryson & Kasper (2012), Lin et al. (2017), Cheng et al. (2009), Ju et al. (2018), Kohavi et al. (2013), Manigandan et al. (2019), Ławrynowicz & Potoniec (2014) and Wang & Sun (2013).

6.2 How is this alignment between strategic planning and the development of bi and data mining applications done?

As previously stated, it was not possible to identify development methods for both applications, BI and Data Mining, with a forecast of strategic alignment. However, we can highlight some works that propose some methodology for the efficient development of this application type.

Sharma, Osei-Bryson & Kasper (2012) address in their work the various limitations identified in existing Data Mining process models and propose to solve them through a new improved model proposal, called Integrated Model of Knowledge Discovery and Data Mining (IKDDM), which presents an integrated view of the KDDM (Knowledge Discovery and Data Mining) process and provides explicit support for performing each of the tasks described in the model. The effectiveness and efficiency offered by the IKDDM model against CRISP-DM, a leading model in the KDDM process, were also evaluated. The results of the statistical tests indicated that the IKDDM model outperforms the CRISP-DM model in terms of efficiency and effectiveness. The IKDM model also outperformed CRISP-DM in terms of the quality of the process model itself.

Cheng et al. (2009) present an ontology-based approach to BI applications, specifically in Statistical Analysis and Data Mining. The implemented approach in a Financial Knowledge Management System (FKMS) is capable of: (i) extraction, transformation, and loading of data, (ii) creation and retrieval of data cubes, (iii) statistical analysis and data mining data, (iv) experiment metadata management; (v) recovery of experiments for new problem-solving. The knowledge resulting from each experiment, defined as a set of understanding consisting of data sequences, models, parameters, and reports, is stored, shared, disseminated, and, therefore, useful to support decision making.

As Cheng et al. (2009), Ławrynowicz & Potoniec (2014) propose a Data Mining approach in which domain ontologies are used as background knowledge. Rather than just using purely empirical data, the authors have also developed a tool that implements this approach. This way, an experimental evaluation was performed, comparing the proposed method with cutting-edge perspectives for semantic data classification.

Ju et al. (2018) propose a framework for using citizen-centric big data analytics to drive governance intelligence in smart cities through two perspectives: urban governance issues and data analysis algorithms. The framework consists of three layers: 1) The data merge layer, which builds the citizen-centric panoramic data, for each citizen, by merging citizen-related data from various sources in collaborative urban governance through similarity and resolution calculations of conflicts; 2) a knowledge discovery layer, which profiles the citizen, at an individual and group level, in terms of urban public service delivery and citizen participation through simple statistical analysis techniques, machine learning, and econometric methods; and 3) a decision-making layer, which uses ontology models to standardize attributes related to governance, people and associations to support decision-making governance through data mining and Bayesian Network techniques. The proposed framework in a case study on blood donation governance in China is validated.

Manigandan et al. (2019) propose the M-Clustering algorithm, which provides a solution for data mining using clusters. The proposed algorithm was evaluated by comparing the experimental data concerning K-Means processing efficiency.

Wang & Sun (2013) propose, based on service-oriented architecture and cloud computing, a Geographic Information System platform for water resources and electricity. The platform's objective is to manage the diverse and massive data efficiently, based on the construction of the fundamental large research data structure, design, construction, environment, immigration, equipment, and supplies.

6.3 Do the bi and data mining development methodologies predict an experimental evaluation phase with validation of conclusions through appropriate statistical tests?

Table 5 presents a work summary with some experimental evaluation types.

Table 5

Experimental work evaluation

Question	Number of Articles	Works
Does the study have any experimental evaluation?	06	Bock et al. (2018); Costa et al. (2015); Costa et al. (2016); Ławrynowicz & Potoniec (2014); Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)
Was the sample size calculated?	0	-
Was the normality test done?	04	Costa et al. (2016); Costa et al. (2015); Ławrynowicz & Potoniec (2014); Santos et al. (2017);
Was a hypothesis formally stated?	06	Bock et al. (2018); Costa et al. (2015); Costa et al. (2016); Ławrynowicz & Potoniec (2014); Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)
Was the confidence interval calculated?	02	Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)
Have threats to validity been declared?	04	Costa et al. (2015); Costa et al. (2016); Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)

Source: Produced by the authors.

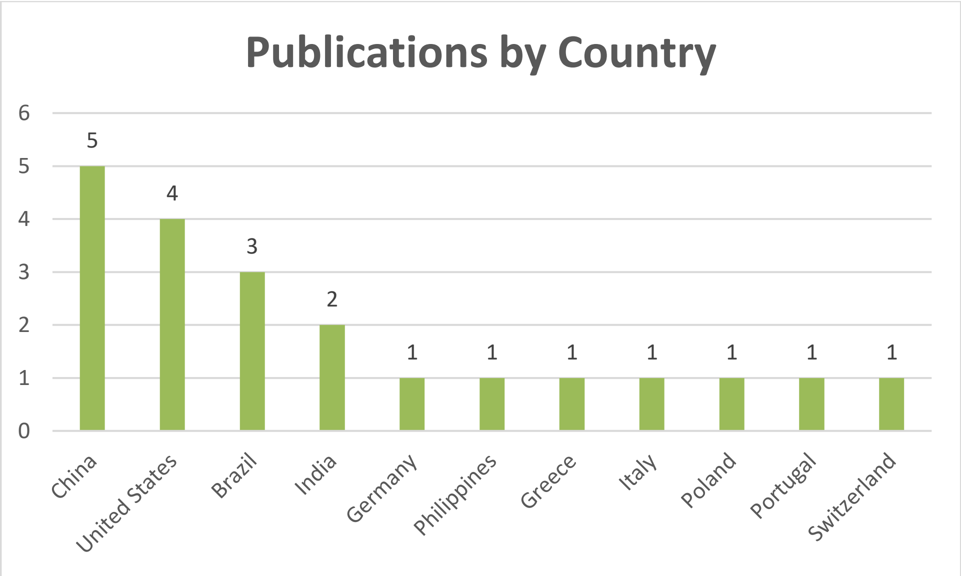
Of the 21 selected works, only 6 (28.5%) were validated experimentally. Thus, it is evident that research involving methodologies for designing BI and Data Mining applications has not prioritized the experimental process conduction. In addition, the analyzed work approaches were not aimed at experimentation, that is, with an exploratory phase prediction in the result validation. The analyzed work experimentation used to evaluate the proposed methodology or approach does not configure a step of the procedure itself.

6.4 Which countries have more researchers who are publishing on this topic?

Figure 5 shows the works selected by country, among which China predominated as the country with the most researchers on this topic, followed by the United States and Brazil.

Figure 5

Main conferences on the theme



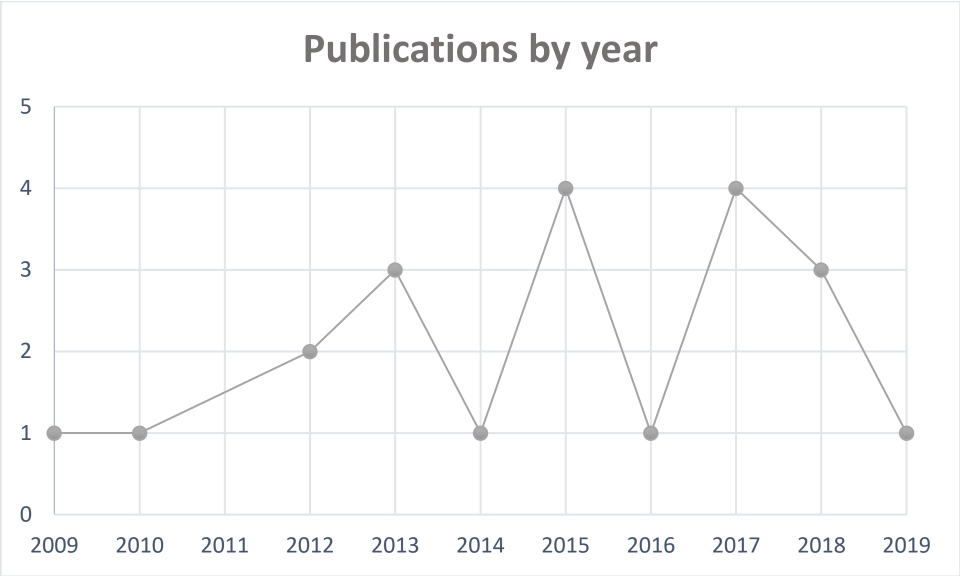
Source: Produced by the authors.

6.5 Which years had the most publications in this area?

A Figure 6 shows selected articles by year of publication. In 2019, the selected articles were from January until March (the month in which the first mapping stage was completed). It can be observed that most amounts of the study numbers were published in 2015 and 2017.

Figure 6

Articles selected by publication year



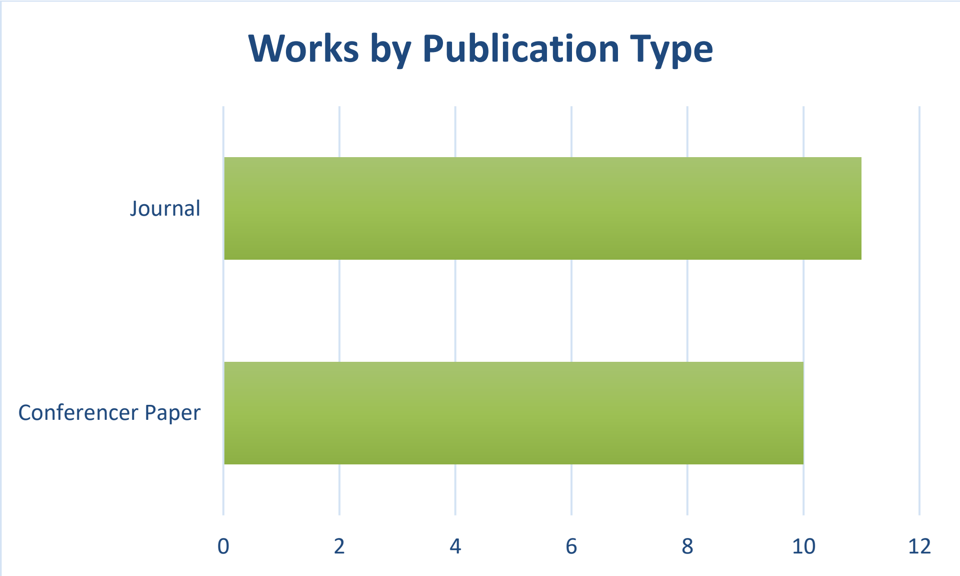
Source: Produced by the authors.

6.6 What are the most popular publishing way?

Figure 7 shows the work numbers selected by publication type.

Figure 7

Works selected by publication type



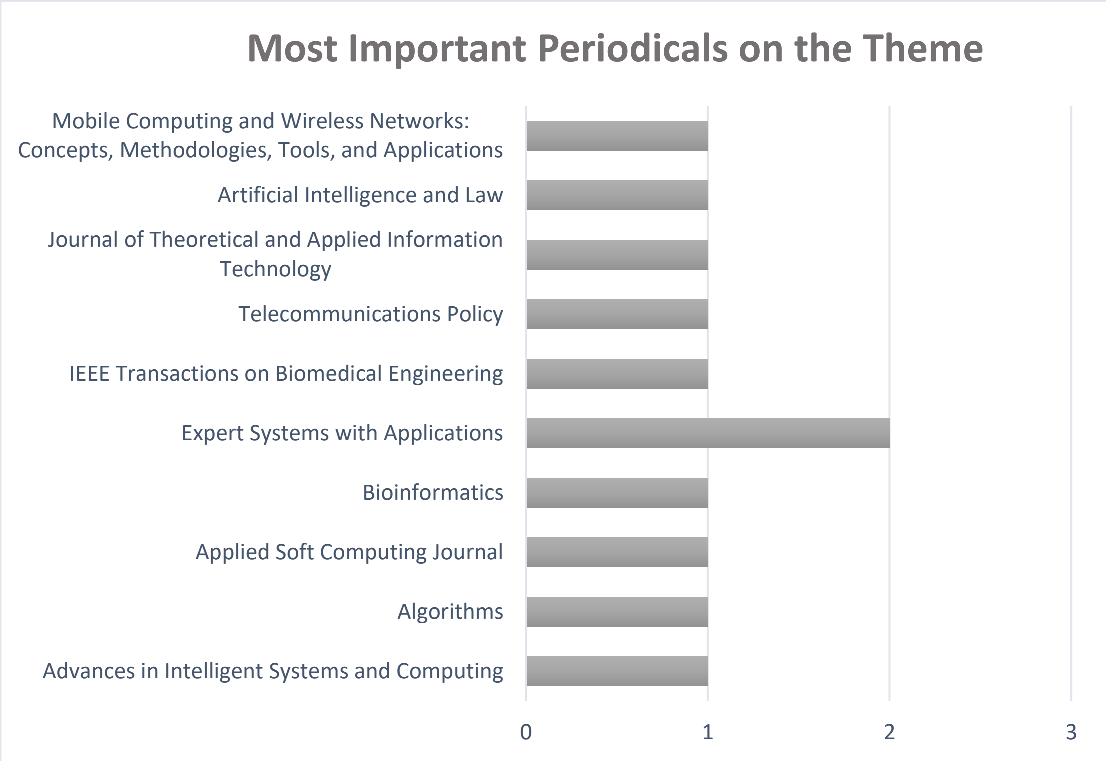
Source: Produced by the authors.

6.7 What are the most important journals and conferences on the topic?

Figure 8 presents the main journals on the subject.

Figure 8

Most important periodicals on the theme

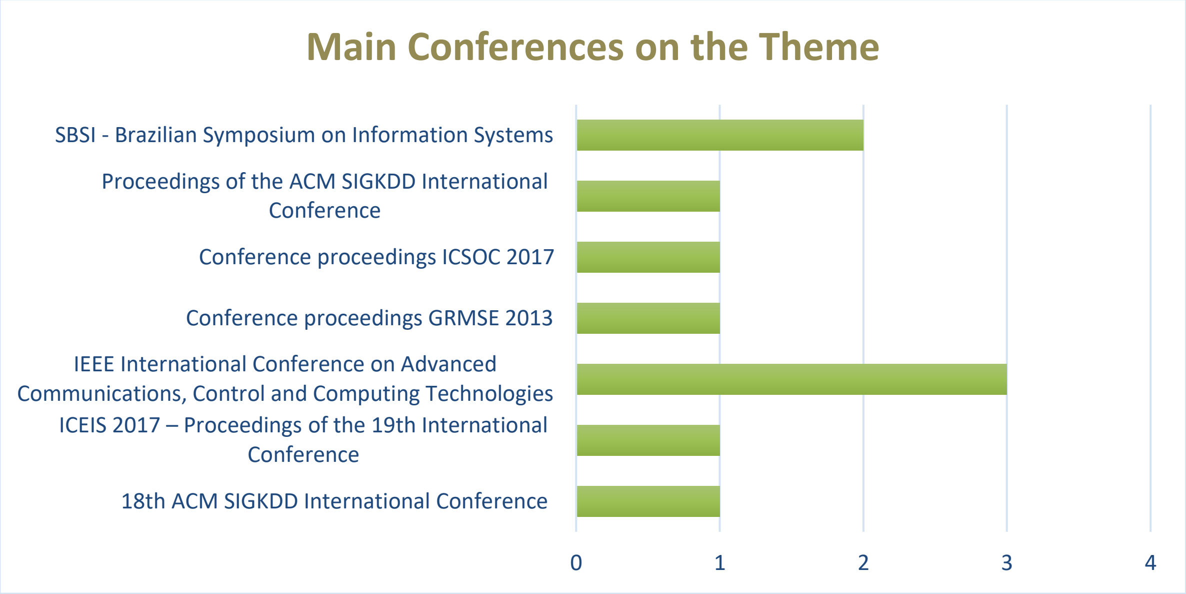


Source: Produced by the authors.

In Figure 9, you can view the most important conferences. Among these "IEEE International Conference on Advanced Communications, Control and Computing Technologies" was the prominent one, with three consecutive publications, in 2013, 2014, and 2015.

Figure 9

Main conferences on the theme



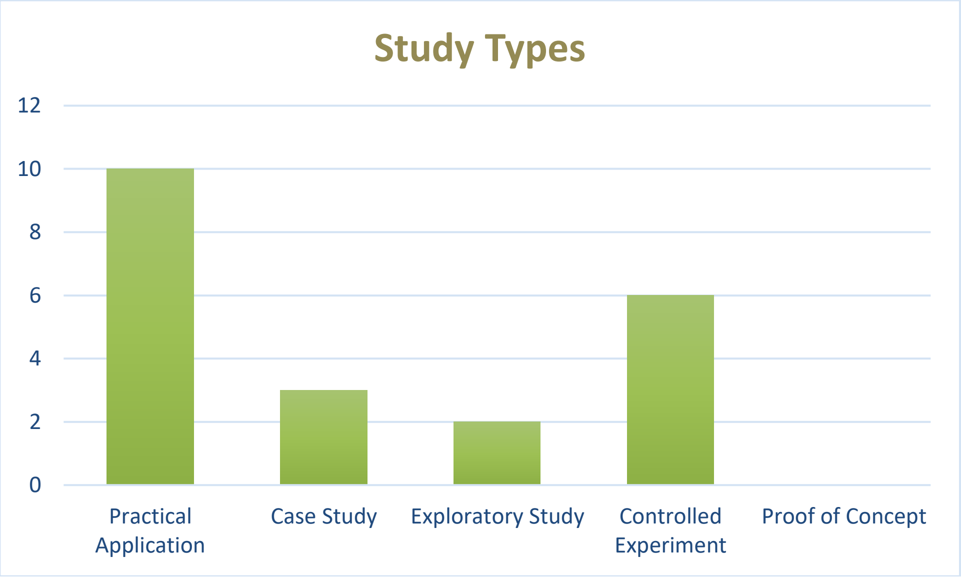
Source: Produced by the authors.

6.8 What performed study types?

Figure 10 shows the study types performed in primary articles, with "Practical Application" being the most used study type.

Figure 10

Study types



Source: Produced by the authors.

For this classification, the following definitions were used:

- Controlled Experiment: An experimental study form in which the investigator has control over the main study aspects and the independent variables studied. This study type is characterized by the variable systematic control and the process, aiming to confirm theories, and conventional knowledge, explore relationships, and evaluate the model predictions or validate measures. In addition, it involves the hypotheses formulation, which concerning the results obtained need to be verified (Delamaro, Jino, & Maldonado, 2017; Wohlin et al., 2012).
- Case Study: It is based on one or more qualitative methods or does not follow a rigid line of investigation. It usually consists of an in-depth study of a single "case" or of "related cases" being carried out under typical conditions, for example, from some common representative projects (Read, 2003).
- Practical Application: We consider this classification when we find similar works to a case study, however, with some gaps in the evaluation method In other words, applications that may not have used real data, or have not run in an actual environment. In addition, there may not have been a qualitative assessment at all.
- Proof of Concept: Term used to name a practical model that can carry out an essay on a (theoretical) concept established by research or technical article. It can also be considered an implementation, generally summarized or incomplete, of a method or an idea, carried out to verify that the concept or theory in question is susceptible to being explored usefully (Farias et al., 2019).
- Exploratory Study: Characterized by the flexibility, creativity, and informality that this study type allows the researcher, in the quest, obtain better knowledge about a particular topic or research problem. Many authors consider exploratory studies as a preliminary stage in the research process as a whole, serving to collect data and information (Hall & Rist, 1999).

Finally, the table below summarizes the main characteristics of the 21 works mapped, including the comparison article used at the beginning of the research. The column that maps the methodology considers "Yes" as the work that presented a procedure, not just a phase of the process of designing an application.

Table 6

Result summary

Article	Approached BI	Approached DM	Study Type	Did it present any Development Methodology?	Strategic Alignment
Cheng et al. (2009)	X	X	**Case Study**	**Yes**	
Ruggieri, Pedreschi & Turini (2010)		X	**Practical Application**	**Yes**	
Sharma, Osei-Bryson & Kasper (2012)		X	**Controlled Experiment**	**Yes**	
Yu et al. (2012)	X	X	**Practical Application**	**No**	
Han et al. (2013)		X	**Practical Application**	**No**	
Kohavi et al. (2013)	X		**Exploratory Study**	**Yes**	
Wang & Sun (2013)	X	X	**Exploratory Study**	**Yes**	
Ławrynowicz & Potoniec (2014)		X	**Controlled Experiment**	**Yes**	
More (2014)		X	**Case Study**	**No**	
Costa et al. (2015)	X		**Controlled Experiment**	**No**	
Puppala et al. (2015)	X	X	**Practical Application**	**No**	
Vitt & Xiong (2015)	X	X	**Practical Application**	**No**	
Costa et al. (2016)	X		**Experimento Controlado**	**No**	
Kanavos et al. (2017)		X	**Practical Application**	**No**	
Lin et al. (2017)		X	**Practical Application**	**Yes**	
Santos et al. (2017)	X		**Controlled Experiment**	**No**	
Sun et al. (2017)		X	**Practical Application**	**Yes**	
Bautista (2018)		X	**Practical Application**	**Yes**	
Bock et al. (2018)		X	**Controlled Experiment**	**No**	
Ju et al. (2018)	X	X	**Case Study**	**Yes**	
Manigandan et al. (2019)		X	**Practical Application**	**Yes**	
Colaço Júnior et al. (2019) – Comparison Article	X		**Case Study**	**Yes**	X

Source: Produced by the authors.

7 Narrative summary

In this section, in addition to the discussion on experimentation, strategy, and raw results carried out in the previous sections, a discussion of the main aspects and lessons learned about possible improvements in the process of Business Intelligence Experimental Development and Data Mining applications is presented.

Regarding strategic alignment, as a limited exception, it was possible to present only the process published by Colaço Júnior et al. (2019), a control article of this study, which merges the GQM+Strategies approach (see Table 6) with an agile development methodology for Business Intelligence applications proposed by the author, aiming to ensure strategic alignment. Despite the advance in disciplining strategic alignment, the process still does not cover Data Mining solutions and does not foresee experimental validation of knowledge models, which are the main components of the objective of this work.

Improvements in the developing intelligent applications process have been proposed since the 1950s, when the term Business Intelligence was first used by Hans Peter Luhn, an IBM researcher, in the article entitled "A Business Intelligence System" (LUHN, 1958). During this period, several knowledge discovery process models were proposed by researchers and professionals. Examples include Fayyad, et al. (1996), Berry & Linoff (1997), Cabena et al. (1998), Cios et al. (2000), CRISP-DM (2003), IBM (2005), SAS (2005), Sharma, Osei-Bryson & Kasper (2012) and Ławrynowicz & Potoniec (2014).

Specifically on the coupling of Data Mining to BI applications, besides the seminal KDD (Knowledge Discovery in Databases) (Fayaad et al., 1996), despite not having been developed by the scientific community, CRISP-DM (Wirth & Hipp, 2000); Kurgan & Musilek, 2006) and SEMMA (Sample, Explore, Modify, Model, Assess) (Mariscal, Marban & Fernandez, 2020; Matignon, 2007) are options, however, they also do not present a definition of the integration of the concepts of BI and Data Mining, as well as they do not cover aspects related to the strategic planning of the organization. An alternative, not found in this mapping for Data Mining, would be the use of auxiliary methods for strategic alignments, such as those presented in Table 7.

Table 7

Auxiliary methods for strategic alignment

Method	Description	Source
COBIT	The Control Objectives for Information and Related Technology (COBIT) is defined as an audit-based guidelines set for ICT processes, practices, and controls aimed at reducing risk, seeking integrity, reliability, and information security.	(Cobit, 2016)
BSC	Balanced Scorecard (BSC) is a method derived from corporate governance, which performs the measurement role well, but does not include good practices. Its concepts have been incorporated into the Information Technology strategic plan process.	(Tonelli, et al., 2014)
GQM+Strategies	The systematic approach that integrates business objectives, adapting them to software process models, products, and quality perspectives, based on the project-specific needs.	(Basili, et al., 2014)

Source: Produced by the authors.

Regarding the use of an experimental approach, this is an alternative to meet the assumptions of Data Science (Bock et al., 2018; Costa et al., 2015; Costa et al., 2016; Ławrynowicz & Potoniec, 2014; Santos et al., 2017; Sharma, Osei-Bryson & Kasper, 2012), since the application of a rigorous scientific method is consistent with the attempt to make data analysis a science, with principles that reduce threats to the validity of knowledge generated.

Application being the most used research method (47.61%). These numbers show the need to increase the use of the scientific method in this area, with study repetition that will allow us to assess whether other researchers will independently reach the same results. Furthermore, even those confirmed did not follow or propose a BI or Data Mining methodology aimed at experimentation, i.e., it predicts an experimental phase in the result validation.

In summary, one of the crucial lessons learned here is that any proposed approach, methodology, or new process should focus on meeting strategic organizational objectives and validating its usefulness in well-executed experiments or case studies. It is far from proposing a new knowledge model and performing some feasibility studies. Researchers and practitioners should focus on answering questions such as: Can my approach be utilized in the real world?; How do my results generalize to other organizations?. Otherwise, there may be a problem with the external validity of the proposed approach, and it will be arduous to move from the art state to the practice state.

Finally, the difficulty in classifying the discovered approaches as strategically aligned or experimental. Even with the possibility of using intelligent text mining and robot use based on this technology, plenty of human intervention is necessary to define a study as exploratory and verify strategic alignment. It was essential to read the entire text and meticulously interpret the data, with some non-automatic associations which would have to be implemented in an innovative extraction tool. For example, the presence of a strategic objective does not imply that the methodology used anticipated the use of this objective or met it.

8 Threats to validity

Threats to validity can limit the ability to interpret or describe results from the data obtained. Therefore, there is no way to disregard the following threats found in this study.

Build Validity: The search string and research questions used may not cover the area of BI and Data Mining methodologies addressed tostrategy-driven and experimentally assessed. To mitigate this threat, an attempt to develop a string as comprehensive as possible was made, regarding the terms that could be used in the area, using several synonyms. Such terms were identified and refined, with the help of control articles guided by the PICO model, using works that were of interest to the research (intervention) and false positives, to calibrate the search string. In addition, the three researchers' opinions were considered.

Internal Validity: Three researchers were responsible for extracting and classifying data from each publication. Therefore, biases or problems in data extraction can threaten the data characterization validity. (Selection Bias): Initially, the articles were included or excluded according to the own researchers' judgment. Consequently, some studies may have been categorized incorrectly. Thus, to mitigate these threats, selection and extraction reviews were made by all researchers involved and any disagreements found were resolved in a final vote.

External Validity: The use of the English language may have contributed to the non-inclusion of possible relevant documents in other languages. In addition, for this specific area, there may be unpublished industry guides and best practices.

9 Conclusion

In this work, a systematic mapping was accomplished, aiming to identify and characterize the methodologies for the Business Intelligence Development and Data Mining applications aimed at strategy or that provide for Experimental evaluation, evaluating, and highlighting of the most relevant works in the area. In this way, the mapping included works published in the period from 2009 to 2019. Therefore, the initial search returned 841 studies, of which, after being evaluated, according to the inclusion and exclusion criteria established in the selection protocol, only 21 works were accepted.

As a result, no works were found that presented a complete approach to discipline strategic alignment and experimentation, providing clear compliance with strategic objectives and an experimental phase in the validation of results. However, We could map some parts of these characteristics which were tested, such as, for example, experimentation, found in 28.57% of the works. Among the countries, China, the United States, and Brazil led the ranking of publications on the subject. As for the means of publication, the Journal was the most used option for publication. In addition, the "IEEE International Conference on Advanced Communications, Control and Computing Technologies" and the journal "Expert Systems with Applications" stood out as the best publishers.

Thus, it is believed that this research presents relevant results to academia and entrepreneurs, providing evidence that there is a gap in research on a formal method of experimental development of BI and Data Mining applications aimed at the organization's strategic planning.

Finally, this work presents itself as a consultation source to the standards of existing methods for intelligent applications development, as well as it can be replicated and extended. In future work, methods for creating intelligent applications aligned with the strategy and with experimental validation can be proposed.

References

Alexander, A. (2014). Case studies: Business Intelligence. *Accounting Today*. (June), 32.

Araújo, M. V. M., & Dornelas, J. S. (2017). Mapeamento perceptual da associação entre sucesso de projetos de TI e fatores promotores do alinhamento estratégico. *EnANPAD*.

Astley, W. G., Axelsson, R., Butler, R. J., Hickson, D. J., & Wilson, D. C. (2017). Complexity and cleavage: dual explanations of strategic decision-making. *In The Bradford studies of strategic decision making* (pp. 47-65). *Ashgate*.

Barbieri, C. (2011). BI2-*Business intelligence: Modelagem & Qualidade*. Elsevier Editora.

Basili, V. R. (1996, March). The role of experimentation in software engineering: past, current, and future. In *Proceedings of IEEE 18th International Conference on Software Engineering* (pp. 442-449). IEEE.

Basili, V. R., Lindvall, M., Regardie, M., Seaman, C., Heidrich, J., Münch, J., ... & Trendowicz, A. (2010). *Linking software development and business strategy through measurement*. *Computer*, *43*(4), 57-65.

Basili, V. R., Trendowicz, M. Kowalczyk, J. Heidrich, C. Seaman, J. Münch, D. Rombach. (2014). *Aligning Organizations Through Measurement: The GQM+Strategies Approach*. *Springer Publishing Company, Incorporated.*

Bautista, R. M. (2018). Clustering failed courses of engineering students using association rule mining. *Journal of Theoretical & Applied Information Technology*, v. 96(4).

Bergin, S., & Wraight, P. (2006). Silver based wound dressings and topical agents for treating diabetic foot ulcers. *Cochrane Database of Systematic Reviews*, (1).

Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Bock, C., Gumbsch, T., Moor, M., Rieck, B., Roqueiro, D., & Borgwardt, K. (2018). Association mapping in biomedical time series via statistically significant shapelet mining. *Bioinformatics*, 34(13), i438-i446.

Bosch-Sijtsema, P., & Bosch, J. (2015). User involvement throughout the innovation process in high-tech industries. *Journal of Product Innovation Management*, 32(5), 793-807.

Brannon, N. (2010). Business intelligence and E-discovery. *Intellectual Property & Technology Law Journal*, 22(7), 1.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.

Camilo, C. O., & Silva, J. C. D. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFC)*, 1(1), 1-29.

Campbell, B. R. (2005). Alignment: Resolving ambiguity within bounded choices. *In Pacific Asia Conference on Information Systems*. University of Hong King.

Castellion, G. (2008). Do it wrong quickly: how the web changes the old marketing rules by Mike Moran. *Journal of Product Innovation Management*, v. 25, n. 6, p. 633-635.

Chan, Y. E., Huff, S. L., Barclay, D. W., & Copeland, D. G. (1997). Business strategic orientation, information systems strategic orientation, and strategic alignment. *Information systems research*, 8(2), 125-150.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.

Cheng, H., Lu, Y. C., & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications*, *36*(2), 3614-3622.

Cios, K. J., Teresinska, A., Konieczna, S., Potocka, J., & Sharma, S. (2000). Diagnosing myocardial perfusion from PECT bull's-eye maps-A knowledge discovery approach. *IEEE Engineering in Medicine and Biology Magazine*, *19*(4), 17-25.

Clancy, T. (1995). The standish group report. *Chaos report*.

Cobit. (2016). *What is Cobit 5? It's the leading framework for the governance and management of enterprise IT. Information Systems Audit and Control Foundation (ISACA)*. [Online] 20 de Junho de 2019. http://www.isaca.org/COBIT/Pages/default.aspx

Colaço Júnior, M., Cruz, R. F. & Lima, A. S. (2019). Proposal and Evaluation of a Strategy-Driven Business Intelligence Applications Development Process. In: *International Conference on Information Systems and Technology Management*, 2019, São Paulo. CONTECSI.

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, *70*, 379-390.

Costa, E., Baker, R. S., Amorim, L., Magalhães, J., & Marinho, T. (2013). Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1), 1-29.

Costa, J. K. G., Santos, I. P. O., Nascimento, A. V. R., & Colaço Júnior, M. (2015, May). Experimentation at Industrial Setting to Improve the Effectiveness of the ETL Procedures Implementation in a Business Intelligence Environment. In *SBSI* (pp. 459-466).

Costa, J. K., Santos, I. P., Colaço Junior, M.., & Nascimento, A. V. (2016, May). An Experiment in an Industrial Business Intelligence environment to improve data loads maintenance. In *Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era-Volume 1* (pp. 534-541).

Costa, S. C. M., de Mattos Pimenta, C. A., & Nobre, M. R. C. (2007). A estratégia PICO para a construção da pergunta de pesquisa e busca de evidências. *Revista Latino-Americana de Enfermagem*, 15(3).

CRISP-DM. (2003). *Cross Industry Standard Process for Data Mining 1.0: Step by Step Data Mining Guide*. [Online] 20 de Junho de 2019. http://www.crisp-dm.org/

Dedić, N., & Stanier, C. (2016). *An evaluation of the challenges of multilingualism in data warehouse development*.

Dedić, N., & Stanier, C. (2017). Measuring the success of changes to Business Intelligence solutions to improve Business Intelligence reporting. *Journal of Management Analytics*, 4(2), 130-144.

Dittrich, Y., Nørbjerg, J., Tell, P., & Bendix, L. (2018, May). Researching cooperation and communication in continuous software engineering. In *2018 IEEE/ACM 11th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)* (pp. 87-90). IEEE.

Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: A survey. *IEEE Transactions on Industrial Informatics*, 8(3), 679-687.

Endres, A., & Rombach, H. D. (2003). *A handbook of software and systems engineering: Empirical observations, laws, and theories*. Pearson Education.

Fagerholm, F., Guinea, A. S., Mäenpää, H., & Münch, J. (2017). The RIGHT model for continuous experimentation. *Journal of Systems and Software*, 123, 292-305.

Farias, M. A., Xisto, R., Santos, M. S., Fontes, R. S., Colaço Júnior, M., Spínola, R., & Mendonça, M. (2019, May). Identifying technical debt through a code comment mining tool. In *Proceedings of the XV Brazilian Symposium on Information Systems* (pp. 1-8).

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD* (Vol. 96, pp. 82-88).

Goldratt, E. M., & Cox, J. (2016). *The goal: a process of ongoing improvement*. Routledge.

Hall, A. L., & Rist, R. C. (1999). Integrating multiple qualitative research methods (or avoiding the precariousness of a one-legged stool). *Psychology & Marketing*, *16*(4), 291-304.

Han, R., Nie, L., Ghanem, M. M., & Guo, Y. (2013, October). Elastic algorithms for guaranteeing quality monotonicity in big data mining. In *2013 IEEE International Conference on Big Data* (pp. 45-50). IEEE.

Hans, R. T., & Mnkandla, E. (2013, September). Modeling software engineering projects as a business: A business intelligence perspective. In *2013 Africon* (pp. 1-5). IEEE.

Hohnhold, H., O'Brien, D., & Tang, D. (2015, August). Focusing on the long-term: It's good for users and business. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1849-1858).

IBM. (2005). *Analytics solutions unified method*. ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf

Isaca. (2018). *COBIT® 2019 Framework: Introduction & Methodology*. *Information Systems Audit and Control Foundation* (ISACA).

Ju, J., Liu, L., & Feng, Y. (2018). Citizen-centered big data analysis-driven governance intelligence framework for smart cities. *Telecommunications Policy*, 42(10), 881-896.

Kanavos, A., Nodarakis, N., Sioutas, S., Tsakalidis, A., Tsolis, D., & Tzimas, G. (2017). Large scale implementations for twitter sentiment classification. *Algorithms*, *10*(1), 33.

King, W. R. (1988). How effective is your information systems planning?. *Long range planning*, 21(5), 103-112.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*(2004), 1-26.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1), 7-15.

Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the web: survey and practical guide. *Data mining and knowledge discovery*, *18*(1), 140-181.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013, August). Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1168-1176).

Kohavi, R., Deng, A., Longbotham, R., & Xu, Y. (2014, August). Seven rules of thumb for web site experimenters. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1857-1866).

Kohavi, R., & Longbotham, R. (2017). Online Controlled Experiments and A/B Testing. *Encyclopedia of machine learning and data mining*, *7*(8), 922-929.

Kohtamäki, M., & Farmer, D. (2017). Strategic Agility—Integrating Business Intelligence with Strategy. In *Real-time Strategy and Business Intelligence* (pp. 11-36). Palgrave Macmillan, Cham.

Kubina, M., Varmus, M., & Kubinova, I. (2015). Use of big data for competitive advantage of company. *Procedia Economics and Finance*, *26*, 561-565.

Kurgan, L. A., & Musilek, P. (2006). A survey of knowledge discovery and data mining process models. *Knowledge Engineering Review*, *21*(1), 1-24.

Laudon, K. C., Laudon, J. P., & Marques, A. S. (2004). *Sistemas de informação gerenciais*. Pearson Educación.

Ławrynowicz, A., & Potoniec, J. (2014). Pattern based feature construction in semantic data mining. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *10*(1), 27-65.

Lima, Adriano, Colaço Júnior, Methanias, Nascimento & Andre Vinicius RP. (2017). Um Survey com Empresas Brasileiras acerca da Utilização de Business Intelligence (BI) e um diagnóstico sobre a infraestrutura e metodologias associadas. *Conferência Ibero-Americana de Engenharia de Software* – Trilha de Engenharia de Software Experimental.

Lin, Y. F., Huang, C. F., & Tseng, V. S. (2017). A novel methodology for stock investment using high utility episode mining and genetic algorithm. *Applied Soft Computing*, *59*, 303-315.

Mandić, V., Basili, V., Harjumaa, L., Oivo, M., & Markkula, J. (2010, September). Utilizing GQM+

Strategies for business value analysis: An approach for evaluating business goals. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 1-10).

Manigandan, E., Shanthi, V., & Kasthuri, M. (2019). Parallel clustering for data mining in CRM. In *Data Management, Analytics and Innovation* (pp. 117-127). Springer, Singapore.

Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, *25*(2), 137.

Matignon, R. (2007). *Data mining using SAS enterprise miner* (Vol. 638). John Wiley & Sons.

McCue, C. (2014). *Data mining and predictive analysis: Intelligence gathering and crime analysis*. Butterworth-Heinemann.

Medeiros Júnior, J. V., de Sousa Neto, M. V., Añez, M. E. M., & de Moraes, E. A. (2017). Identifying mechanisms to develop information technology capabilities. *Revista Ibero-Americana de Estratégia*, 16(4), 37-49.

Mola, L., Rossignoli, C., Carugati, A., & Giangreco, A. (2015). Business intelligence system design and its consequences for knowledge sharing, collaboration, and decision-making: an exploratory study. *International Journal of Technology and Human Interaction (IJTHI)*, *11*(4), 1-25.

More, S. (2014, May). Modified path traversal for an efficient web navigation mining. In *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* (pp. 940-945). IEEE.

Olsson, H. H., Alahyari, H., & Bosch, J. (2012, September). Climbing the" Stairway to Heaven"--A Mulitiple-Case Study Exploring Barriers in the Transition from Agile Development towards Continuous Deployment of Software. In *2012 38th euromicro conference on software engineering and advanced applications* (pp. 392-399). IEEE.

Olsson, H. H., & Bosch, J. (2014). The HYPEX model: from opinions to data-driven software development. In *Continuous software engineering* (pp. 155-164). Springer, Cham.

Olszak, C. M., & Ziemba, E. (2012). Critical success factors for implementing business intelligence systems in small and medium enterprises on the example of upper Silesia, Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*, *7*(2), 129-150.

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)* 12 (pp. 1-10).

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1-18.

Puppala, M., He, T., Chen, S., Ogunti, R., Yu, X., Li, F., ... & Wong, S. T. (2015). METEOR: an enterprise health informatics environment to support evidence-based medicine. *IEEE Transactions on Biomedical Engineering*, *62*(12), 2776-2786.

Reich, B. H., & Benbasat, I. (1996). Measuring the linkage between business and information technology objectives. *MIS quarterly*, 55-81.

Rodríguez, P., Haghighatkhah, A., Lwakatare, L. E., Teppola, S., Suomalainen, T., Eskeli, J., ... &

Oivo, M. (2017). Continuous deployment of software intensive products and services: A systematic mapping study. *Journal of Systems and Software*, 123, 263-291.

Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Integrating induction and deduction for finding evidence of discrimination. *Artificial Intelligence and Law, 18*(1), 1-43.

Santos, I. P. O., Costa, J. K. G., Colaço Júnior, M., & Nascimento, A. V. R. (2017, April). Experimental Evaluation of Automatic Tests Cases in Data Analytics Applications Loading Procedures. In *ICEIS (1)* (pp. 304-311).

SAS. (2005). *Semma data mining methodology*. http://www.sas.com/technologies/analytics/datamining/miner/semma.html

Sharma, S., Osei-Bryson, K. M., & Kasper, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. In *Expert Systems with Applications*, *39*(13), 11335-11348.

Shi, Y., & Lu, X. (2010, November). The role of business intelligence in business performance management. In *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering* (Vol. 4, pp. 184-186). IEEE.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Silva Filho, R. L. L., Motejunas, P. R., Hipólito, O., & Lobo, M. B. D. C. M. (2007). A evasão no ensino superior brasileiro. *Cadernos de pesquisa*, 37, 641-659.

Singh, B. (2016). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Bangalore* Vol. 11, Ed. 2.

Sjøberg, D. I., Hannay, J. E., Hansen, O., Kampenes, V. B., Karahasanovic, A., Liborg, N. K., & Rekdal, A. C. (2005). A survey of controlled experiments in software engineering. *IEEE transactions on software engineering*, 31(9), 733-753.

Sun, Y., Bauer, B., & Weidlich, M. (2017, November). Compound trace clustering to generate accurate and simple sub-process models. In *International Conference on Service-Oriented Computing* (pp. 175-190). Springer, Cham.

Thamir, A., & Poulis, E. (2015). Business intelligence capabilities and implementation strategies. *International Journal of Global Business*, 8(1), 34.

Tonelli, A. O., Bermejo, P. H. D. S., & Zambalde, A. L. (2014). Using the bsc for strategic planning of it (information technology) in brazilian organizations. *JISTEM-Journal of Information Systems and Technology Management*, *11*(2), 361-378.

Vitt, C. A., & Xiong, H. (2015, November). The impact of patent activities on stock dynamics in the high-tech sector. In *2015 IEEE International Conference on Data Mining* (pp. 399-408). IEEE.

Yu, L., Zheng, J., Shen, W. C., Wu, B., Wang, B., Qian, L., & Zhang, B. R. (2012, August). BC-PDM: data mining, social network analysis and text mining system based on cloud computing. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1496-1499).

Wang, X., & Sun, Z. (2013, November). The design of water resources and hydropower cloud GIS platform based on big data. In *International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem* (pp. 313-322). Springer, Berlin, Heidelberg.

Weber, M., & Klein, A. Z. (2013). Gestão estratégica em empresas de tecnologia da informação: um estudo de caso. *Revista Ibero Americana de Estratégia*, 12(3), 37-65.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1). London, UK: Springer-Verlag.