



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

 **Rodrigo Fontes Cruz**<sup>1</sup>  **Methanias Colaço Júnior**<sup>2</sup>  **Victor Menezes Gois**<sup>3</sup>

<sup>1</sup> Mestre em Ciência da Computação, Universidade Federal de Sergipe – UFS. São Cristóvão, Sergipe – Brasil. [rodrifontes@gmail.com](mailto:rodrifontes@gmail.com)

<sup>2</sup> Doutor em Ciência da Computação – Pós Doutor em Gestão, Universidade Federal de Sergipe – UFS. São Cristóvão, Sergipe – Brasil. [mjrse@hotmail.com](mailto:mjrse@hotmail.com)

<sup>3</sup> Bacharel em Sistemas de Informação, Universidade Federal de Sergipe – UFS. Itabaiana, Sergipe – Brasil. [vicormenezes.gois@gmail.com](mailto:vicormenezes.gois@gmail.com)

## 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.

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

### 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.

## ¿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). Quão experimentais e estratégicas são as aplicações de Business Intelligence (BI) e data mining? *Iberoamerican Journal of Strategic Management (IJSM)*, 21, 1-36, e20638. <https://doi.org/10.5585/riae.v21i1.17689>.

(ABNT – NBR 6023/2018)

CRUZ, Rodrigo Fontes; COLAÇO JÚNIOR, Methanias; GOIS, Victor Menezes. Quão experimentais e estratégicas são as aplicações de Business Intelligence (BI) e data mining? *Iberoamerican Journal of Strategic Management (IJSM)*, v. 21, p. 1-36. 2022. <https://doi.org/10.5585/riae.v21i1.17689>.

## 1 Introdução

Com a inevitável mutação dos mercados, no âmbito empresarial, saber como obter resultados positivos em meio a essas mudanças, desenvolvendo processos adequados e eficazes para gerir adequadamente as transformações, tornou-se uma obrigação. Decisões erradas, sejam estratégicas, táticas ou operacionais, podem custar o futuro da empresa, assim como uma correta, definir sua sobrevivência ou sua expansão (Côrte-Real, Oliveira & Ruivo, 2017).

Neste contexto, os dados surgem como uma importante fonte para obter vantagem competitiva (Kubina, Varmus & Kubinova, 2015). Vantagem que se baseia no conhecimento obtido com a análise de dados e tem impulsionado áreas como a de *Business Intelligence* (BI) e *Data Mining* (DM). Isto associado à atual era Big Data, na qual o progresso tecnológico impulsiona a criação de grandes volumes de dados em alta velocidade, a partir de uma variedade de fontes, tem justificado ainda mais o investimento em DM, cujo poder e automaticidade têm possibilitado lidar com grandes quantidades de dados e extrair valor (Shmueli et al., 2017).

No entanto, antes que qualquer tentativa possa ser feita para realizar a extração desse conhecimento útil, uma abordagem geral que descreva como extrair conhecimento precisa ser estabelecida (Kurgan & Musilek, 2006). Desta forma, vários modelos de processos de *Data Mining* foram propostos por pesquisadores e profissionais. Os exemplos de autores incluem Fayyad, et al. (1996), Cabena et al. (1998), Cios et al. (2000), CRISP-DM (2003), Berry & Linoff (1997), Sharma, Osei-Bryson & Kasper (2012) e Ławrynowicz & Potoniec (2014).

Além disso, como os objetivos estratégicos da organização precisam ser traduzidos para os níveis mais baixos do negócio, esta metodologia deve ser planejada e executada a partir de dentro da organização. Em outras palavras, é preciso preencher a lacuna entre a estratégia de negócios e sua implementação ao nível do projeto de software (Basili et al., 2010; Basili et al., 2014). Nesta mesma linha, Mandic et. al. (2010) e Astley et. al. (2017) enfatizam que para se analisar todos os aspectos relevantes para a tomada de decisão, é necessário implantar métodos para a integração dos dados existentes com as metas estratégicas.

Apesar deste senso comum entre diversos pesquisadores, um *survey* realizado no Brasil (Lima et al., 2017) evidenciou que 72% das empresas não utilizam um método específico para o desenvolvimento de aplicações de BI alinhado ao planejamento estratégico da organização. Outra evidência encontrada pelo *survey* destaca que 67,50% dos entrevistados não utilizam uma metodologia formal para o desenvolvimento de aplicações de BI. Neste caso, vale ressaltar que apesar da literatura e a prática separarem conceitualmente as áreas de *Data Mining* e BI, há uma forte convergência e integração destas, pois o “I”, ou Inteligência do BI, só pode ser concretizado com a aplicação de técnicas de *Data Mining*, as quais englobam *Machine Learning*. Isto indica que a ausência destas metodologias de alinhamento também deve atingir os projetos de Data Mining, uma vez que existem projetos de BI sem *Data Mining*, *Data Mining* sem BI e, no melhor caso, o qual será considerado nesta proposta, um

BI completo, que utilizada integração de dados analíticos - BI -, bancos de dados, estatística e inteligência artificial - *Data Mining*.

Diante destes dados, o objetivo deste trabalho é identificar, reunir, organizar e sintetizar o maior número possível de evidências científicas sobre metodologias para o desenvolvimento experimental de aplicações inteligentes alinhadas ao planejamento estratégico, as quais permitam caracterizar tanto a área de estudo quanto formar uma base teórica sobre o tema. Além disso, o resultado dessa caracterização pode auxiliar na identificação de novas oportunidades de pesquisas, além de direcionar os esforços em estratégias mais adequadas na área de estudo.

Neste contexto, foi planejado e executado um estudo secundário, do tipo Mapeamento Sistemático. Cabe ressaltar que, até o momento, nenhum mapeamento sobre o assunto mencionado foi publicado ou se encontra disponível para pesquisas acadêmicas.

Após responder às questões de pesquisa, identificou-se a não existência de trabalhos que apresentassem alguma abordagem para disciplinar o alinhamento estratégico no desenvolvimento de aplicações de BI e *Data Mining*, bem como não foi possível identificar nenhuma metodologia que fosse dirigida à experimentação, ou seja, que prevê uma fase experimental na validação dos resultados. Em relação aos tipos de estudos, destacaram-se o "Aplicação Prática", com 47,61% das publicações e o "Experimento Controlado", com 28,57%. 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.

O restante deste artigo está organizado da seguinte forma. Na seção 2, os trabalhos relacionados sobre o tema são apresentados. Na seção 3, temos a base conceitual para o entendimento deste trabalho. Na seção 4, o método adotado neste trabalho é abordado. A seção 5 apresenta a condução do mapeamento. Na seção 6, temos os resultados alcançados. A seção 7 analisa criticamente os resultados e discute lições aprendidas. Na seção 8, as ameaças à validade encontradas são detalhadas. E, finalmente, na seção 9, a conclusão é apresentada.

## 2 Trabalhos relacionados

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. No entanto, alguns trabalhos mencionam a importância desse alinhamento estratégico no desenvolvimento de tais aplicações.

Com relação aos trabalhos avaliados por este mapeamento, ou seja, trabalhos que lidaram especificamente com *Business Intelligence* e *Data Mining*, alguns trabalhos pontuaram a importância

do alinhamento estratégico, tais como os trabalhos de Sharma, Osei-Bryson & Kasper (2012) e Kohavi et al. (2013). Além disso, muitos trabalhos na literatura destacam a importância do alinhamento estratégico de aplicações de tecnologia da informação em geral. São exemplos os trabalhos de: Isaca (2018), Weber & Klein (2013), Medeiros Júnior et al. (2017) e Araújo & Dornelas (2017). A seguir, serão sintetizados outros trabalhos relevantes à temática específica aqui abordada.

Mola et al. (2015) fizeram um estudo exploratório que analisa os efeitos das características técnicas e organizacionais dos sistemas de BI nos processos de compartilhamento de conhecimento, colaboração e tomada de decisão. Em média, as características técnicas e organizacionais dos Sistemas de BI estão positivamente associadas a um aumento no compartilhamento de conhecimento, levando a uma melhoria na colaboração interna e na qualidade de tomada de decisão. Estas melhorias dependem da forma como o BI é projetado.

Em um *survey* realizado no Brasil por Lima et al. (2017), foi constatado que 67,50% das empresas não utilizam uma metodologia experimentada para o desenvolvimento de BI, o que contribui para que os projetos não obtenham sucesso. Associado a este resultado, um percentual de 72,00% das empresas não utiliza uma metodologia de alinhamento estratégico. A ausência de uma metodologia alinhada à estratégia da empresa evidencia que os gestores podem estar tomando decisões com base em informações não relevantes à instituição ou desalinhadas às estratégias de negócio.

Em outro *survey* similar sobre *Business Intelligence*, Duan & Xu (2012) apresentam uma introdução sobre BI, com ênfase em algoritmos fundamentais para o uso de BI em ambientes empresariais, destacando os desafios e oportunidades encontrados nesses ambientes.

Colaço Júnior et al. (2019) apresentaram um processo que mescla a abordagem GQM+Strategies com uma metodologia de desenvolvimento ágil de aplicações de *Business Intelligence* proposta pelo autor, visando garantir o alinhamento estratégico. O processo proposto foi avaliado por meio de um estudo de caso, em uma empresa multinacional latino-americana do mercado de varejo, no qual foi evidenciado que é possível integrar a abordagem de alinhamento estratégico adotada com uma metodologia de desenvolvimento de aplicações de BI. Com as boas evidências iniciais, os pesquisadores poderão evoluir o processo e prever o uso de experimentação para validação dos modelos inteligentes que poderão ser criados com técnicas de *Data Mining* e IA.

### 3 Base conceitual

Nesta seção, são introduzidos alguns conceitos necessários para o entendimento deste trabalho.

#### 3.1 Alinhamento estratégico

Na literatura, existem muitos conceitos que detalham o significado de alinhamento estratégico, no entanto, buscamos aqueles que possuem maior vínculo com o assunto de nossa proposta. (1) O alinhamento entre o plano estratégico de negócio (PEN) e o plano estratégico de tecnologia de

informação (PETI) é alcançado quando o conjunto de estratégias de sistemas (objetivos, obrigações e estratégias) é derivado do conjunto estratégico organizacional (missão, objetivos e estratégias) (King, 1988); (2) O elo entre PEN-PETI corresponde ao grau no qual a missão, os objetivos e os planos de TI refletem, suportam e são suportados pela missão, pelos objetivos e pelos planos de negócio (Reich & Benbasat, 1996); (3) É a forma como os negócios e a TI trabalham em conjunto para alcançar o objetivo comum (Campbell, 2005) e (4) O alinhamento entre PEN-PETI é a adequação da orientação estratégica do negócio com a de TI (Chan et al., 1997).

Para Lima (2017), o problema das organizações, atualmente, é que nem sempre as empresas conseguem declarar os objetivos estratégicos de forma explícita ou suficientemente clara, para que se possa verificar se tais objetivos têm realmente alcançado as metas e estão alinhados à TI. Este desafio não é ter a área de TI como um suporte, mas sim como parte de uma plataforma de negócio, servindo como elemento essencial à estratégia de negócio. Essa nova visão deve vincular o alinhamento estratégico da TI ao negócio da organização. Estes dois elementos precisam relacionar-se entre si, em busca da melhoria contínua e do sucesso da organização.

Hoje, uma das estratégias de TI mais importantes é a adoção de *Business Intelligence* e *Data Mining*, que tem como principal característica a capacidade de lidar com grandes quantidades de dados e extrair valor. Uma pergunta que pode ser feita é: como é feito este alinhamento entre o Planejamento Estratégico e o desenvolvimento desse tipo de aplicação? A seguir, detalharemos o conceito de BI e Data Mining e sua relação com o Planejamento Estratégico.

### 3.2 Business intelligence

*Business Intelligence* – BI – pode ser entendido como um conjunto de metodologias, processos, arquiteturas e tecnologias usadas para apoiar a coleta, análise, apresentação e disseminação de informações de negócios para permitir tomadas de decisões estratégicas, táticas e operacionais mais efetivas (Hans et al., 2013; Dedić & Stanier, 2017).

O BI ajuda as empresas a pensar melhor sobre a concorrência por meio de um melhor entendimento da base de clientes (Brannon, 2010), o que pode levar à criação de um relacionamento mais próximo e mais forte com os clientes e ao aumento da receita (Alexander, 2014). Além disso, desempenha um papel crítico para os negócios em termos de desenvolvimento organizacional, fornecendo vantagem competitiva, no contexto de alcançar assimetria positiva de informações (Thamir & Poulis, 2015), e contribui para otimizar processos e recursos de negócios, maximizar lucros e melhorar proativamente, bem como para tomada de decisão estratégica (Dedić & Stanier, 2016).

Dessa forma, o *Business Intelligence* nas organizações é entendido como uma vantagem estratégica (Chaudhuri et al., 2011; Kohtamäki & Farmer, 2017), independentemente da área em que a organização atue, seja ela privada ou não, pois, na atualidade, as organizações que utilizam sistemas deste tipo têm facilidade em adquirir conhecimento específico sobre os diversos fatores que a



influenciam, podendo posteriormente aplicar tal conhecimento, identificando o potencial de mercado e, com isto, o direcionar na sua estratégia, visão e metas a atingir.

Além de suas vantagens estratégicas e táticas, o *Business Intelligence* também é usado no nível operacional, de forma a permitir que vários tipos de usuários identifiquem tendências emergentes, tomem decisões mais rápidas, tomem ações e lidem com os problemas organizacionais assim que surgirem. Seu objetivo é ajudar as partes interessadas a entender melhor as operações de sua organização, tomar decisões de negócios mais sábias e informadas, bem como gerenciar o desempenho operacional (AICPA, 2015).

Assim, *Business Intelligence* refere-se ao ato de proporcionar aos negócios o apoio necessário para a tomada de decisão, por meio do uso de um conjunto de técnicas e ferramentas. A mineração de dados também pode compor uma solução de BI.

### 3.3 Data mining

A mineração de dados ou *Data Mining* (DM) é o processo de análise de dados para extrair informações e conhecimentos que não são visualizados claramente em sistemas comuns. Busca extrair informação e conhecimento, por meio de relações entre os dados, que permitam inferências sobre o que pode ocorrer (análise preditiva) ou correlações entre o que já ocorreu (Barbieri, 2011).

Tecnicamente falando, consiste em um processo que utiliza técnicas de estatística, matemática e inteligência artificial para extrair e identificar informações úteis e conseqüentemente, conhecimento (ou padrões) a partir de grandes volumes de dados, sendo que estes padrões podem ser apresentados como tendências, regras de negócio, correlações ou modelos preditivos. Esses modelos e padrões podem ser utilizados para guiar o processo decisório, bem como prever o efeito dessas escolhas (Laudon, Laudon & Marques, 2004).

Em mineração de dados temos as tarefas e as técnicas que norteiam o que queremos e como obtermos tais informações. As tarefas apontam quais os objetivos, ou seja, quais as informações pretende-se obter. As técnicas de mineração apontam para os métodos empregados para se obter padrões desejados (Camilo & Silva, 2009). Duan & Xu (2012) classificam as técnicas de mineração em dois abrangentes grupos: aprendizado supervisionado e aprendizado não supervisionado. Os métodos de aprendizado supervisionado constroem modelos para prever um atributo não conhecido de acordo com atributos observados, enquanto os métodos de aprendizado não supervisionados extraem padrões, como agrupamentos, gráficos de processo e correlações entre os dados.

Os métodos supervisionados necessitam de um conjunto de dados que possuem uma variável alvo principal pré-definida e os registros são categorizados em relação a esta, ou seja, aplica-se a regra programada com precisão. Nos métodos não-supervisionados não há a necessidade de uma pré-categorização para os registros, isso significa que não é preciso um atributo como alvo principal (Filho et al., 2007).

As técnicas mais comuns de aprendizado supervisionado são a classificação (que também pode ser não-supervisionada) e a regressão (Mccue, 2007). Porém, em mineração de dados, permite-se que várias técnicas sejam aplicadas ou até mesmo que haja uma combinação, visando a busca de melhor resultado esperado (Costa et al., 2012). Avaliar estas técnicas ou qualquer produto de software experimentalmente é um processo de apoio à garantia de qualidade, como pode ser verificado na próxima seção.

### 3.4 Experimentação em software

A rápida entrega de valor aos clientes é uma das principais prioridades das empresas de software (Fagerholm et al., 2017). Com esse objetivo em mente, as empresas geralmente desenvolvem suas práticas de desenvolvimento. Inicialmente, estas herdam os princípios Agile na parte de desenvolvimento da organização (Martin, 2002) e os expandem para outros departamentos (Olsson, Alahyari & Bosch, 2012). Em seguida, as empresas se concentram em vários conceitos enxutos, tais como eliminar o desperdício, remover restrições no pipeline de desenvolvimento (Goldratt & Cox, 2016) e avançar para a integração (Dittrich et al., 2018) e a implantação contínuas da funcionalidade do software (Rodríguez et al., 2017). No entanto, a implantação contínua é caracterizada por um canal bidirecional que permite às empresas não apenas enviar dados aos seus clientes para prototipar rapidamente (Singh, 2016), mas também receber dados de feedback dos produtos em campo.

Neste contexto, a intuição das empresas de desenvolvimento de software sobre as preferências do cliente pode estar errada em até 90% das vezes (Clancy, 1995; Castellion, 2008; Manzi, 2019). Para mitigar isso, os dados atuais de uso de um produto têm o potencial de tornar o processo de priorização no desenvolvimento de novos produtos mais preciso, pois permitem a concentração no que os clientes fazem e não no que dizem (Bosch-Sijtsema & Bosch, 2015). Neste sentido, a experimentação está se tornando a norma nas empresas de software avançadas, para a avaliação confiável de ideias com os clientes, a fim de priorizar corretamente as atividades de desenvolvimento de produtos (Olsson & Bosch, 2014; Kohavi & Longbotham, 2017).

Desta forma, existe uma crescente compreensão na comunidade de que estudos empíricos são necessários para desenvolver ou melhorar processos, métodos e ferramentas para desenvolvimento e manutenção de software (Basili, 1996; Endres & Rombach, 2003; Fagerholm et al., 2017).

A pesquisa em engenharia empírica de software deve ter como objetivo adquirir conhecimentos gerais sobre qual tecnologia (processo, método, técnica, linguagem ou ferramenta) é útil, para quem é útil, na realização de quais tarefas (engenharia de software), em quais ambientes. Portanto, essa pesquisa se concentra no tipo de tecnologia que está sendo estudada nos experimentos investigados (que reflete os tópicos dos experimentos), nos sujeitos que participaram, nas tarefas que eles executaram, no tipo de sistemas de aplicativos nos quais essas tarefas foram executadas e os ambientes em que os experimentos

foram conduzidos. Além disso, também deve incluir dados sobre replicação de experimentos e até que ponto a validade interna e externa é discutida (Sjøberg et al., 2005).

Uma categoria importante de estudo empírico é a do experimento controlado, cuja condução é regida pelo método científico clássico, para identificar relações de causa-efeito. Em um experimento controlado, os usuários são divididos aleatoriamente entre as variantes (por exemplo, os dois designs diferentes de uma interface de produto) de maneira persistente (um usuário recebe a mesma experiência várias vezes). As interações dos usuários com o produto são instrumentadas e as principais métricas são computadas (Kohavi & Longbotham, 2017).

Um dos principais desafios das métricas é decidir sobre o que incluir em um Critério de Avaliação Geral (Overall Evaluation Criteria - OEC). Um OEC é uma medida quantitativa do objetivo de um experimento controlado (Roy, 2001) e orienta a direção do desenvolvimento de negócios. Na experimentação controlada, é intuitivo medir o efeito a curto prazo, ou seja, o impacto observado durante o experimento (Hohnhold, O'brien & Tang, 2014). Fornecer mais peso às métricas de publicidade, por exemplo, torna as empresas mais lucrativas no curto prazo. No entanto, o efeito a curto prazo nem sempre é preditivo do efeito a longo prazo e, conseqüentemente, não deve ser o único componente de uma OEC (Kohavi et al., 2014). Definir um OEC não é trivial e deve ser realizado com muito cuidado. Kohavi et al. (2009, 2014, 2017), em seus trabalhos, apresentam armadilhas comuns no processo de estabelecimento de um sistema de experimentação controlado e orientações sobre como definir de forma confiável um OEC. Este trabalho identificou aplicações de BI e Data Mining que foram validadas experimentalmente.

#### 4 Método

Com o objetivo de identificar e caracterizar as metodologias de desenvolvimento de aplicações de *Business Intelligence* e *Data Mining* dirigidas à estratégia e/ou que preveem avaliação experimental, foi adotada para este trabalho a metodologia de Mapeamento Sistemático da Literatura. Segundo Kitchenham (2004), Petersen et al. (2008) e Petersen et al. (2015), um mapeamento consiste em um protocolo sistemático de busca e seleção de estudos relevantes, com o objetivo de extrair dados e mapear os resultados para um problema de pesquisa específico.

Este método foi adotado por se pautar nos conceitos da Medicina baseada em Evidência (*Evidence-based Medicine*), área madura no processo de revisões sistemáticas, e por propor uma mudança de paradigma em como as pesquisas na área de Software deveriam ser conduzidas. Segundo Kitchenham et al. (2009), a pesquisa em Medicina mudou de forma drástica com o paradigma baseado em evidências, possibilitando uma organização mais efetiva da pesquisa médica e embasando o julgamento clínico de especialistas. O sucesso deste novo paradigma influenciou fortemente a adoção da abordagem baseada em evidências em outras áreas do conhecimento tais como psicologia, enfermagem, ciências sociais, educação e computação.



As próximas seções detalham os passos para replicação deste mapeamento, englobando as *strings* e comandos utilizados, as bases pesquisadas, bem como os critérios de seleção dos artigos e extração dos dados. Para extração, foi explicitamente descrito como os trabalhos foram classificados como experimentais e/ou estratégicos, bem como as variáveis a serem identificadas e seus valores. A seguir, são listadas as fases resumidas do método.

1. Planejamento: os objetivos da pesquisa são listados e o protocolo da revisão é definido;
2. Condução: nesta atividade, as fontes para o mapeamento são selecionadas, os estudos primários são além de identificados, selecionados, avaliados de acordo com os critérios de inclusão, exclusão e de qualidade estabelecidos durante o protocolo do mapeamento;
3. Análise e Publicação dos Resultados: os dados dos estudos são extraídos e sintetizados para serem publicados.

#### 4.1 Planejamento do mapeamento sistemático

##### 4.1.1. Objetivo

Como dito anteriormente, este mapeamento tem como objetivo identificar e caracterizar as metodologias de desenvolvimento de aplicações de *Business Intelligence* e *Data Mining* dirigidas à estratégia e/ou que preveem avaliação Experimental.

##### 4.1.2. Questões de pesquisa

As questões de pesquisa foram elaboradas tendo como base a abordagem PICO (Bergin & Wraight, 2006; Costa et al., 2007). Este modelo, que surgiu na execução de estudos clínicos, estrutura a pesquisa em quatro elementos básicos: População, Intervenção, Comparação (ou controle) e “*Outcomes*” (Resultados). Apesar de ter surgido na medicina, essa estratégia pode ser adaptada para outras áreas, como demonstra o trabalho de Kitchenham (2004), usado como referência para a produção da Tabela 1. A tabela também apresenta alguns artigos de controle, os quais foram selecionados com uma consulta preliminar nas bases de pesquisa, com a seleção dos dois primeiros artigos encontrados, alinhados com a intervenção de interesse. Essa busca preliminar serviu de base para encontrar outras palavras-chave importantes, presentes em artigos verdadeiros positivos que serão resultados finais do mapeamento, e refinar a *string* de busca. Um artigo de controle pode inclusive já ser de conhecimento dos pesquisadores e é um elemento para ajudar a criar e validar a *string*. Neste sentido, foi também usado um artigo de controle que é efetivamente de comparação, ou seja, não se enquadrava totalmente na intervenção de interesse e não está indexado nas bases pesquisadas, mas apresenta uma proposta de alinhamento estratégico. Artigos de controle deste tipo também ajudam a estabelecer uma *string* de busca mais assertiva.

**Tabela 1**

*Modelo PICO para conformidade das questões de pesquisa*

| Acrônimo | Definição   | Descrição   |
|----------|-------------|---|
| <b>P</b> | População   | Publicações de pesquisadores e desenvolvedores, tendo em vista o desenvolvimento de aplicações de <i>Business Intelligence</i> com ou sem o apoio de <i>Data Mining</i> .   |
| <b>I</b> | Intervenção | Métodos de desenvolvimento de aplicações de BI e <i>Data Mining</i> dirigidos à estratégia e/ou que usaram avaliação experimental. A direção estratégica pressupõe métodos que partem da identificação de objetivo(s) estratégico(s) que será(ão) alavancado(s) pelo produto final, evitando a criação de soluções que não estão alinhadas com a estratégia organizacional e não ajudam diretamente a realização dos objetivos traçados.  |
| <b>C</b> | Controle    | Métodos de criação de aplicações de BI ou de descoberta de conhecimento, como, por exemplo, o modelo de processo de <i>Data Mining: Cross Industry Standard Process for Data Mining (CRISP-DM)</i> .<br>Artigos de Controle<br>Artigos que se enquadram na intervenção: <ul style="list-style-type: none"> <li>• <i>Evaluation of an integrated knowledge discovery and data mining process model;</i></li> <li>• <i>Pattern based feature construction in semantic data mining.</i></li> </ul> Artigo de comparação: <ul style="list-style-type: none"> <li>• Proposta e Avaliação de um Processo para o Desenvolvimento de Aplicações de Business Intelligence Dirigido à Estratégia</li> </ul> |
| <b>O</b> | Resultado   | Metodologias de BI e <i>Data Mining</i> que validam suas conclusões por meio de experimentos controlados e/ou definem explicitamente a direção da aplicação para um ou mais objetivos estratégicos.   |

**Source:** Produzida pelos próprios autores.

Assim, a partir da definição do PICO, as seguintes questões de pesquisa foram elaboradas:

- Q1: Quais as metodologias dirigidas à estratégia utilizadas no desenvolvimento de aplicações de BI e *Data Mining*?
- Q2: Como é feito este alinhamento entre o Planejamento Estratégico e o desenvolvimento de aplicações de BI e *Data Mining*?
- Q3: As metodologias de desenvolvimento de BI e *Data Mining* preveem uma fase de avaliação experimental com validação de conclusões por meio de testes estatísticos apropriados?
- Q4: Quais países possuem mais pesquisadores publicando sobre esse tema?
- Q5: Quais os anos que tiveram mais publicações nessa área?
- Q6: Quais os principais periódicos e conferências sobre o tema?
- Q7: Quais os meios de publicações?
- Q8: Quais os tipos de estudos?

Estas questões foram formuladas tendo como base as diretrizes do protocolo de Mapeamento Sistemático da Literatura (Kitchenham, 2004; Petersen et al., 2008; Petersen et al., 2015). Destaca-se que as diferenças entre um mapeamento sistemático da literatura e uma revisão sistemática da literatura

são, fundamentalmente, explicitadas na formulação da questão de pesquisa. De acordo com Petersen et al. (2008), um mapeamento sistemático possui uma pergunta de pesquisa do tipo exploratória e exige menos profundidade na extração dos dados. Além da visão geral, pretendeu-se averiguar se as metodologias usadas relacionavam explicitamente os requisitos da aplicação com objetivos estratégicos e se possuíam uma fase para validar as soluções experimentalmente.

#### 4.1.3. Estratégia de busca e de seleção

Para a execução da busca, foram consultadas as bases de dados: ACM Digital Library (ACM), IEEE Xplore (IEEE) e SCOPUS. As buscas foram realizadas utilizando as ferramentas de filtragem disponibilizadas em cada base de dados citada anteriormente, considerando nas buscas: título, resumo/abstract e palavras-chave dos respectivos artigos. A respeito do idioma, foram selecionados apenas trabalhos em inglês. A respeito da área, foram selecionados apenas trabalhos referentes à Ciência da Computação. E a respeito do tempo de publicação, foram selecionados apenas trabalhos publicados a partir de 2008.

Para realizar a pesquisa nas bases digitais, foi definida uma *string* de busca com a utilização de termos em inglês e do uso de vários sinônimos, associados ao pressuposto de que os estudos estariam contidos nas áreas da computação que lidam com metodologias de desenvolvimento de aplicações de BI e *Data Mining*. Tais termos foram identificados com auxílio dos artigos de controles do modelo PICO, descritos na seção anterior (Tabela 1), e posteriormente, refinados e adaptados para o maior aproveitamento da *string*. A Tabela 2 mostra os termos, antes de refiná-los, que foram selecionados.

**Tabela 2**

*Categorias do modelo PICO e termos identificados para pesquisa bibliográfica antes de refiná-los*

| <b>Categoria</b>   | <b>Descrição</b>  |
|--------------------|---|
| <b>População</b>   | Development, Implementation, Construction, Deployment, Creation, Business Intelligence, Data Mining.  |
| <b>Intervenção</b> | Methodology, Method, Approach, Process, Experiment.   |
| <b>Controle</b>    | Métodos de criação de aplicações de BI ou de descoberta de conhecimento, como, por exemplo, o CRISP-DM (sem <i>strings</i> ).   |
| <b>Resultado</b>   | 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:** Produzida pelos próprios autores.

Após o refinamento, os termos ajustados foram utilizados para construir a *string* de busca, os quais estão descritos na Tabela 3.

**Tabela 3**

Strings *eleitas após o refinamento*

| Termos da <i>string</i> de busca   |   |   |
|--|---|---|
| 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:** Produzida pelos próprios autores.

- A string de pesquisa gerada com os termos evidenciados acima foi:

*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).*

A busca por pesquisas em computação foi considerada com base em artigos de controle que abrangiam mais de uma área, uma vez que a computação é um meio para a gestão e tomada de decisão, e, além disso, considerou-se que o objetivo foi explorar a área que desenvolve a parte técnica da Tecnologia da Informação, a qual já deveria ter como padrão o alinhamento estratégico e a experimentação, ou seja: Isso já é uma realidade?

#### 4.1.4 Critérios de seleção de fontes

Para filtrar os artigos relevantes para este mapeamento, foram estabelecidos os critérios de inclusão e exclusão dos mesmos. Depois de todas as fases, o estudo contabilizou os artigos que focassem no uso de alguma metodologia para o desenvolvimento de aplicações de BI e DM voltada ao planejamento estratégico ou que usaram avaliação experimental, utilizando os seguintes critérios preliminares de inclusão:

1. Resultado deve conter o tema deste estudo, já automaticamente limitado pela *string*, no título, resumo ou palavras-chave;
2. O resultado deve datar entre os anos de 2009 a 2019;

3. O resultado deve apresentar uma avaliação experimental ou explorar alguma metodologia, método, processo ou abordagem de desenvolvimento de aplicações de BI ou *Data Mining*. Isso foi feito para gerir o risco de não haver metodologias dirigidas à estratégia ou experimentais e pelo menos serem listados trabalhos que abordaram metodologias de desenvolvimento de BI ou *Data Mining* ou usaram alguma avaliação experimental;
4. O resultado deve estar disponível para consulta online.

A confirmação dos critérios de inclusão foi dada após análise do resumo de cada um dos artigos encontrados, na primeira filtragem, seguida de uma leitura completa, para segunda filtragem. Antes da primeira leitura de cada artigo, foi realizada a análise quanto aos critérios de exclusão. Foram eliminados:

1. Estudos secundários, pois eles tratam de abordagens de terceiros;
2. Publicações duplicadas;
3. Artigos Curtos (*Short Papers*);
4. *Surveys*.

#### 4.1.5. Estratégia de extração de informações

A estratégia de extração de dados é projetada para reunir as informações necessárias e responder às questões de pesquisa, avaliando a qualidade do trabalho. Isto posto, após o término da etapa de seleção, os trabalhos definidos serão lidos na íntegra, logo em seguida, um formulário será respondido, conforme Tabela 4, contendo informações sobre o conteúdo abordado em cada trabalho. Sendo assim, este formulário nos permite avaliar os trabalhos de forma mais detalhada e precisa, realizando a classificação baseado nos critérios de inclusão e exclusão.

No que concerne à identificação de alinhamento estratégico, não foi suficiente a citação do termo e/ou o destaque para sua importância em algum ponto do artigo. Desta forma, para ser classificado como um trabalho que usou alguma abordagem de alinhamento, foi necessário que a solução proposta pelos autores considerasse como premissa ou base pelo menos um objetivo estratégico explícito.



**Tabela 4**

*Formulário de extração*

|     |  |  |
|-----|--|--|
| 1.  | Qual a metodologia utilizada?                  |  |
| 2.  | Como é feito o alinhamento estratégico?        |  |
| 3.  | O artigo se referia a Business Intelligence?   | [Sim, Não]                               |
| 4.  | O artigo se referia a Data Mining?             | [Sim, Não]                               |
| 5.  | Qual o tipo de estudo utilizado?               | [Aplicação Prática, Estudo de Caso, ...] |
| 6.  | O estudo foi dirigido à estratégia?            | [Sim, Não]                               |
| 7.  | O estudo possui alguma avaliação experimental? | [Sim, Não]                               |
| 8.  | Foi feito o cálculo do tamanho da amostra?     | [Sim, Não]                               |
| 9.  | Foi feito o teste de normalidade?              | [Sim, Não]                               |
| 10. | Foi declarada hipótese formalmente?            | [Sim, Não]                               |
| 11. | Foi calculado o intervalo de confiança?        | [Sim, Não]                               |
| 12. | Foram declaradas as ameaças à validade?        | [Sim, Não]                               |

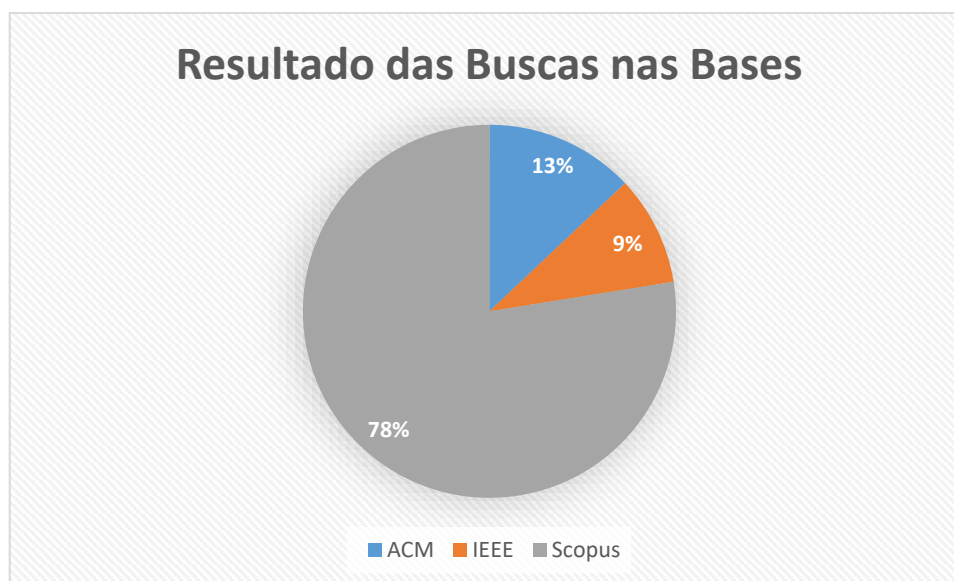
**Source:** Produzida pelos próprios autores.

## 5 Condução do mapeamento sistemático

O planejamento do mapeamento sistemático foi elaborado entre os meses de fevereiro a março de 2019, já a execução, ocorreu em abril do mesmo ano. Para a obtenção dos estudos primários, foi necessária a formação da *string* de busca a partir das combinações das palavras-chave em inglês. Assim, a *string* base foi definida na máquina de busca Scopus, refinada e, quando julgado que a *string* era adequada, foi traduzida para as máquinas de busca ACM e IEEE. Logo em seguida, as buscas foram realizadas. No total, foram retornados 841 trabalhos, sendo 652 (78%) do Scopus, 109 (13%) da ACM e 80 (10%) do IEEE, como mostra a Figura 1.

**Figura 1**

*Resultado da execução das strings de busca nas bibliotecas digitais*



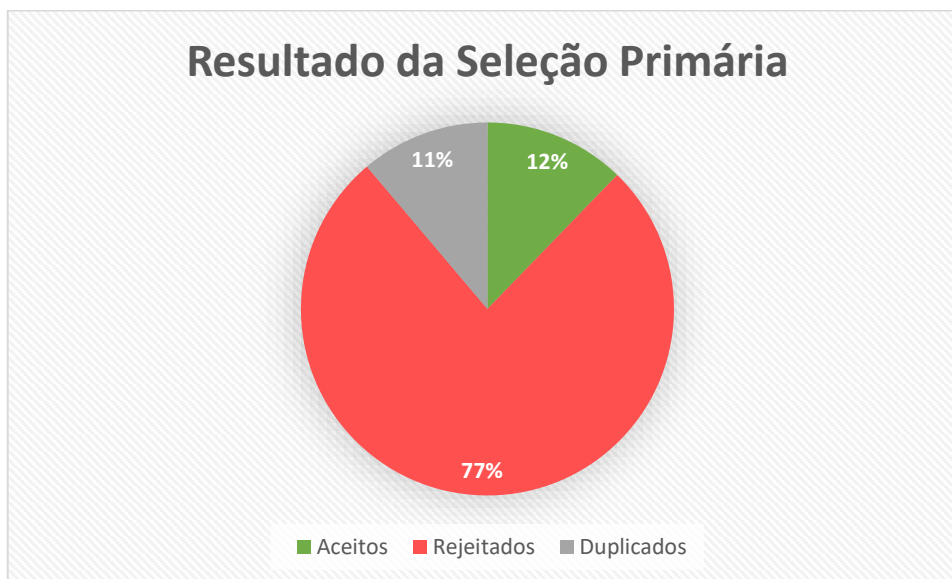
**Source:** Produzida pelos próprios autores.

Com a finalização da busca, houve o início do processo de filtragem dos artigos encontrados, com base nos critérios de seleção, definidos na seção 4.1.4. Nesta fase, os trabalhos foram classificados em Aceito, Rejeitado e Duplicado.

Neste sentido, do total de 841 publicações analisadas, 94 (11% do total) eram duplicadas e, conseqüentemente, as duplicatas acabaram sendo eliminadas. Assim, 747 trabalhos foram selecionados, para uma avaliação superficial, na qual foi realizada a leitura de todos os títulos e *abstract*, aplicando os critérios de inclusão e exclusão, definidos previamente no protocolo. Ao final desta etapa, foram identificados 644 (77%) trabalhos que estavam fora do escopo deste mapeamento e por isso foram rejeitados. Estes trabalhos eram artigos curtos, revisões de literatura, *surveys* ou artigos fora dos critérios. Os aceitos mencionavam a realização de um experimento ou mencionavam o uso ou proposta de uma abordagem, processo, metodologia ou método. Como resultado final, foram aceitos 103 (12%) trabalhos para serem avaliados de forma mais precisa e detalhada. A Figura 2 ilustra os resultados dessa primeira seleção.

**Figura 2**

*Resultado da primeira etapa de seleção dos trabalhos*



**Source:** Produzida pelos próprios autores.

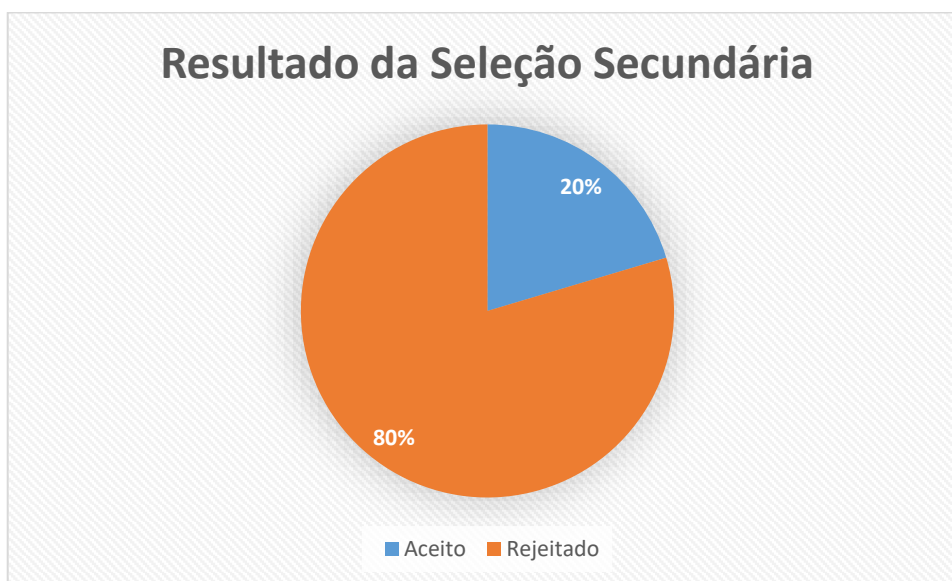
Após essa primeira etapa de seleção, foi feita uma avaliação precisa e detalhada a partir da leitura completa dos trabalhos, de forma a confirmar o indício da presença de alinhamento estratégico ou de avaliação experimental. No entanto, não foi possível acessar todas as publicações, mesmo após contatar os autores por e-mail. Dos 103 artigos que deveriam ter sido avaliados nessa etapa, foram recuperados 100 (97,09%). Dessa forma, não foi possível acessar o texto completo de 3 (2,91%) artigos, que consequentemente acabaram sendo rejeitados.

Para enquadrar os estudos como experimentos controlados e classificar a presença de uma avaliação experimental, foram considerados os trabalhos nos quais a base do método científico foi verificada, ou seja, aqueles que formalizaram a definição das hipóteses e efetuaram a validação estatística necessária para teste.

Durante esta etapa, foi identificado que 82 (80%) trabalhos estavam fora do escopo deste mapeamento sistemático e por isso foram rejeitados. Todos os artigos que confirmaram a realização de um experimento foram novamente aceitos. Fora estes, só foram aceitos artigos que o método, metodologia, processo ou abordagem mencionados estavam relacionados à concepção de aplicações de BI ou *Data Mining*, ou seja, BI ou *Data Mining* não eram assuntos transversais do artigo. Como resultado final, foram aceitos 21 (20%) trabalhos. Para estes artigos, finalmente, uma nova leitura norteou o preenchimento do formulário de extração, descrito na seção 4.1.5. A Figura 3 ilustra o resultado final dessa segunda fase de seleção e extração dos dados.

Figura 3

Resultado da segunda etapa de seleção dos trabalhos



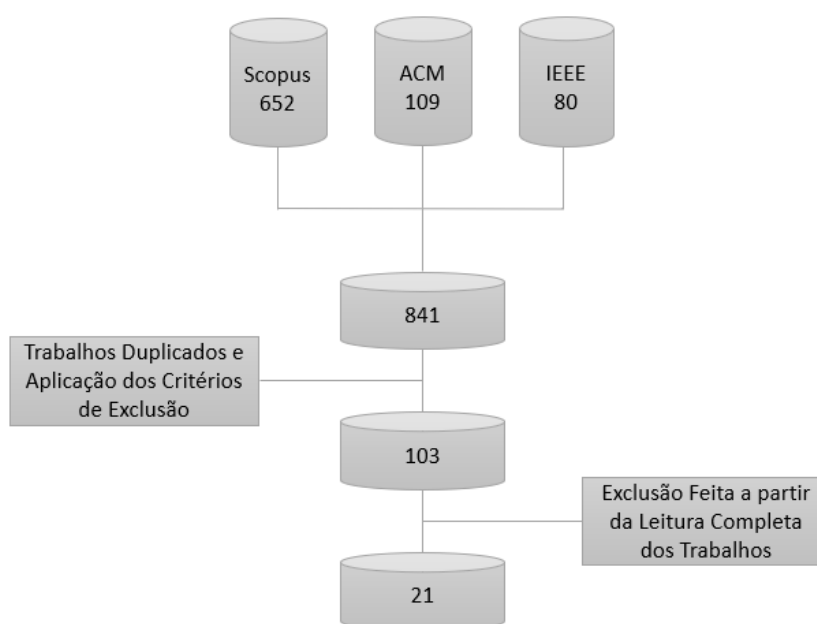
Source: Produzida pelos próprios autores.

## 6 Síntese dos dados e apresentação dos resultados

Nesta seção, os resultados do mapeamento sistemático são apresentados. A Figura 4 mostra um resumo do número de trabalhos obtidos em cada etapa do processo de busca e em seguida as questões de pesquisa são respondidas de acordo com os dados extraídos.

Figura 4

Resultados obtidos durante o processo de busca



Source: Produzida pelos próprios autores.

### 6.1 Quais as metodologias dirigidas à estratégia utilizadas no desenvolvimento de aplicações de bi e de data mining?

Não foram encontrados trabalhos que apresentassem alguma abordagem para disciplinar o alinhamento estratégico. No entanto, alguns trabalhos mencionam a importância deste alinhamento ou propõem alguma metodologia para o desenvolvimento eficiente deste tipo de aplicação, a exemplo de 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) e Wang & Sun (2013).

### 6.2 Como é feito este alinhamento entre o planejamento estratégico e o desenvolvimento de aplicações de bi e data mining?

Conforme dito anteriormente, não foi possível identificar métodos de desenvolvimento de ambas as aplicações, BI e Data Mining, com previsão de alinhamento estratégico. No entanto, podemos destacar alguns trabalhos que propõem alguma metodologia para o desenvolvimento eficiente deste tipo de aplicação.

Sharma, Osei-Bryson & Kasper (2012) abordam em seu trabalho as várias limitações identificadas nos modelos de processos existentes de *Data Mining* e propõem resolvê-las por meio da proposta de um novo modelo melhorado, denominado, Modelo Integrado de Descoberta de Conhecimento e *Data Mining* (IKDDM), que apresenta uma visão integrada do processo KDDM (*Knowledge Discovery and Data Mining* – Descoberta de Conhecimento e Mineração de Dados) e fornece suporte explícito para a execução de cada uma das tarefas descritas no modelo. Também foi avaliada a eficácia e a eficiência oferecidas pelo modelo IKDDM contra o CRISP-DM, um modelo líder no processo de KDDM. Os resultados dos testes estatísticos indicaram que o modelo IKDDM supera o modelo CRISP-DM em termos de eficiência e eficácia. O modelo IKDDM também superou o CRISP-DM em termos de qualidade do próprio modelo de processo.

Cheng et al. (2009) apresentam uma abordagem baseada em ontologias para aplicações de BI, especificamente em Análise Estatística e Data Mining. Implementando a abordagem em um Sistema de Gestão do Conhecimento Financeiro (FKMS), que é capaz de: (i) extração, transformação e carregamento de dados, (ii) criação e recuperação de cubos de dados, (iii) análise estatística e mineração de dados, (iv) gerenciamento de metadados de experimentos; (v) recuperação de experimentos para nova resolução de problemas. O conhecimento resultante de cada experimento, definido como um conjunto de conhecimento que consiste em sequências de dados, modelo, parâmetros e relatórios, é armazenado, compartilhado, disseminado e, portanto, útil para apoiar a tomada de decisões.

Assim como Cheng et al. (2009), Ławrynowicz & Potoniec (2014) propõem uma abordagem de Data Mining, na qual as ontologias de domínio são usadas como conhecimento de fundo. Ao invés de usarem apenas dados puramente empíricos, os autores também desenvolveram uma ferramenta que implementa essa abordagem. Desta forma, foi conduzida uma avaliação experimental, comparando o método proposto com abordagens de ponta para a classificação de dados semânticos.



Ju et al. (2018) propõem uma estrutura para o uso de análise de big data centrada no cidadão, para impulsionar a inteligência de governança em cidades inteligentes, por meio de duas perspectivas: questões de governança urbana e algoritmos de análise de dados. A estrutura consiste em três camadas: 1) A camada de mesclagem de dados, que constrói os dados panorâmicos centrados no cidadão, para cada cidadão, mesclando dados relacionados a cidadãos de várias fontes na governança urbana colaborativa, por meio de cálculos de similaridade e resolução de conflitos; 2) uma camada de descoberta de conhecimento, que traça o perfil do cidadão, em nível individual e de grupo, em termos de prestação de serviços públicos urbanos e participação do cidadão por meio de técnicas simples de análise estatística, aprendizado de máquina e métodos econométricos; e 3) uma camada de tomada de decisão, que usa modelos de ontologia para padronizar atributos relacionados à governança, pessoas e associações para apoiar a governança de tomada de decisão, por meio de mineração de dados e técnicas de Rede Bayesiana. A estrutura proposta é validada em um estudo de caso sobre a governança da doação de sangue na China

Manigandan et al. (2019) propõem o algoritmo M-Clustering, o qual fornece uma solução para a mineração de dados usando clusters. O algoritmo proposto foi avaliado, comparando a eficiência de processamento de dados experimentais em relação ao K-Means.

Wang & Sun (2013) propõem, com base na arquitetura orientada a serviços e computação na nuvem, uma plataforma de Sistema de Informação Geográfica de recursos hídricos e de energia elétrica. O objetivo da plataforma é gerenciar os diversos e massivos dados de forma eficiente, com base na construção da estrutura fundamental de big data de pesquisa, projeto, construção, ambiente, imigração, equipamentos e suprimentos.

### *6.3 As metodologias de desenvolvimento de bi e data mining preveem uma fase de avaliação experimental com validação de conclusões por meio de testes estatísticos apropriados?*

A Tabela 5 apresenta uma síntese dos trabalhos que apresentaram algum tipo de avaliação experimental.

**Tabela 5**

*Avaliação experimental dos trabalhos*

| <i>Questão</i>                                 | <i>Número de Artigos</i> | <i>Trabalhos</i>   |
|--|--------------------------|--|
| O estudo possui alguma avaliação experimental? | 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) |
| Foi feito o cálculo do tamanho da amostra?     | 0                        | -  |
| Foi feito o teste de normalidade?              | 04                       | Costa et al. (2016); Costa et al. (2015); Ławrynowicz & Potoniec (2014); Santos et al. (2017);   |
| Foi declarada hipótese formalmente?            | 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) |
| Foi calculado o intervalo de confiança?        | 02                       | Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)  |
| Foram declaradas as ameaças à validade?        | 04                       | Costa et al. (2015); Costa et al. (2016); Santos et al. (2017); Sharma, Osei-Bryson & Kasper (2012)  |

**Source:** Produzida pelos próprios autores.

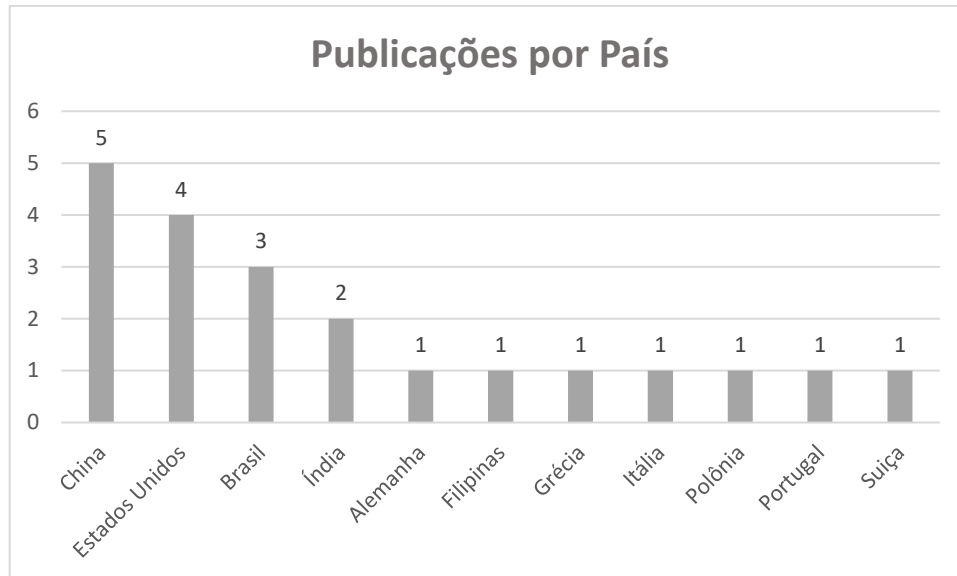
Dos 21 trabalhos selecionados, apenas 6 (28,5%) eram validados experimentalmente. Assim, fica evidenciado que as pesquisas que envolvem metodologias para concepção de aplicações de BI e *Data Mining* não têm priorizado a condução de um processo experimental. Além disso, as metodologias observadas nos trabalhos não eram dirigidas à experimentação, ou seja, com previsão de uma fase experimental na validação dos resultados. A experimentação, nos trabalhos analisados, foi usada para avaliar a metodologia ou abordagem proposta, não configura um passo da própria metodologia.

#### *6.4 Quais países possuem mais pesquisadores publicando sobre esse tema?*

A Figura 5 mostra os trabalhos selecionados por país, entre os quais predominou a China como o país que possui mais pesquisadores sobre esse tema, seguida por Estados Unidos e Brasil.

**Figura 5**

*Publicações por país*



**Source:** Produzida pelos próprios autores.

#### *6.5 Quais os anos que tiveram mais publicações nessa área?*

A Figura 6 mostra os artigos selecionados por ano de publicação. No ano de 2019, os artigos foram selecionados até o mês de março (mês em que a primeira etapa da execução do mapeamento foi encerrada). Pode-se observar que os maiores números de estudos foram publicados nos anos de 2015 e 2017.

**Figura 6**

*Artigos selecionados por ano de publicação*



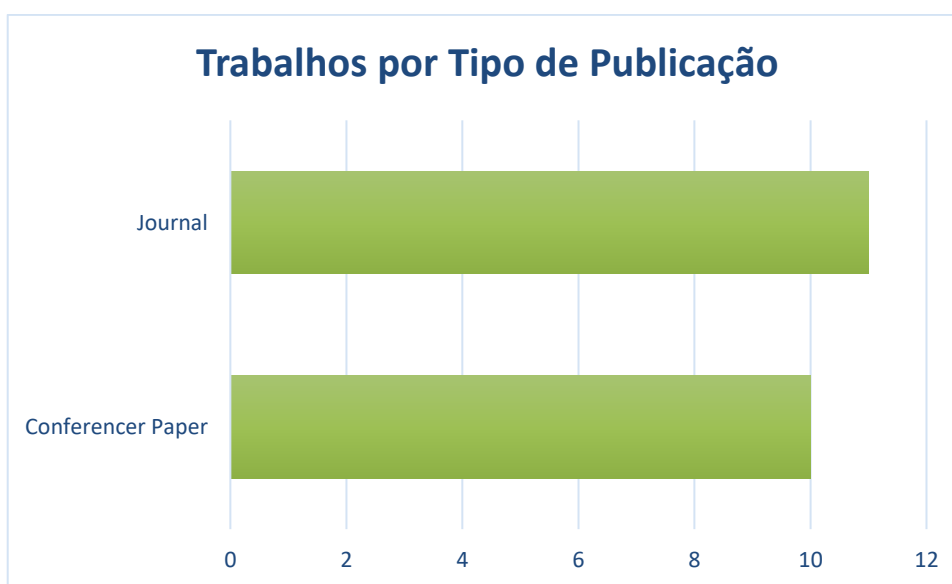
Source: Produzida pelos próprios autores.

#### 6.6 Quais os meios de publicações mais populares?

A Figura 7 apresenta o quantitativo dos trabalhos selecionados por tipo de publicação.

**Figura 7**

*Trabalhos selecionados por tipo de publicação*



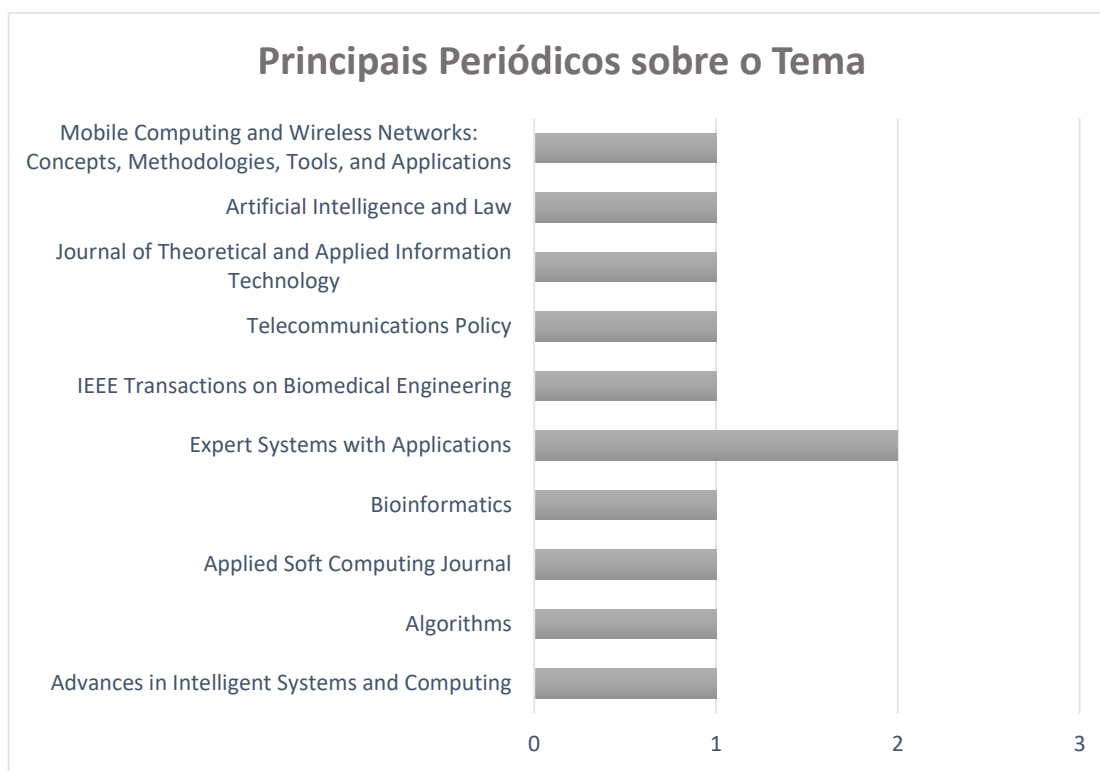
Source: Produzida pelos próprios autores.

#### 6.7 Quais os principais periódicos e conferências sobre o tema?

A Figura 8 apresenta os principais periódicos sobre o tema.

Figura 8

Principais periódicos sobre o tema



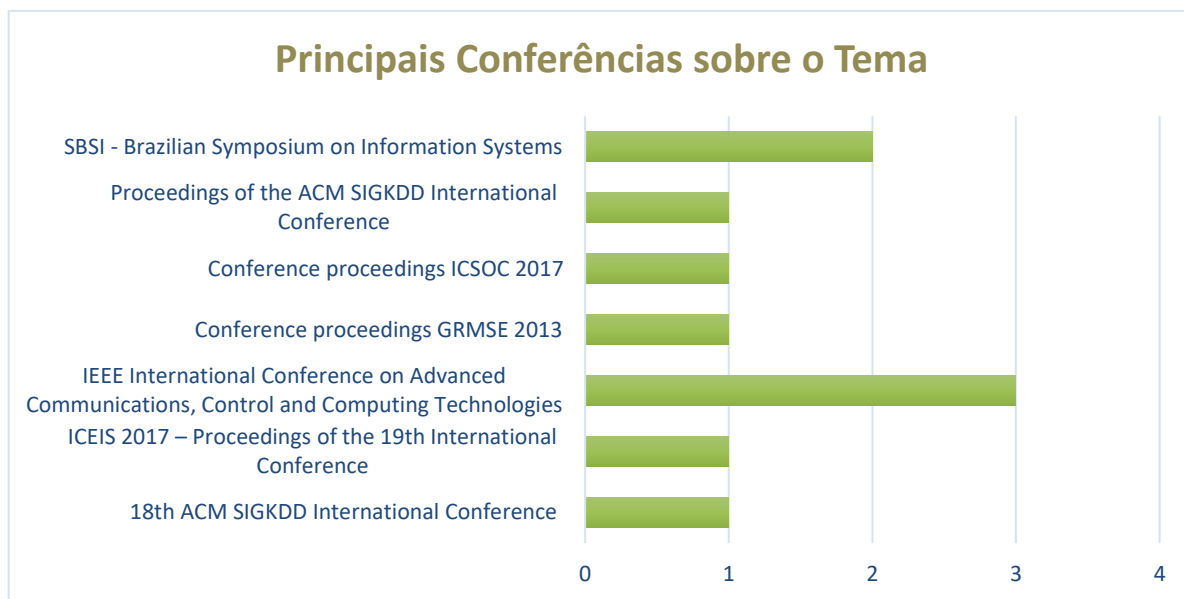
Source: Produzida pelos próprios autores.

Na Figura 9, é possível visualizar as principais conferências. Entre as quais a “*IEEE International Conference on Advanced Communications, Control and Computing Technologies*” foi a proeminente, com 3 publicações consecutivas, nos anos 2013, 2014 e 2015.



**Figura 9**

*Principais conferências sobre o tema*



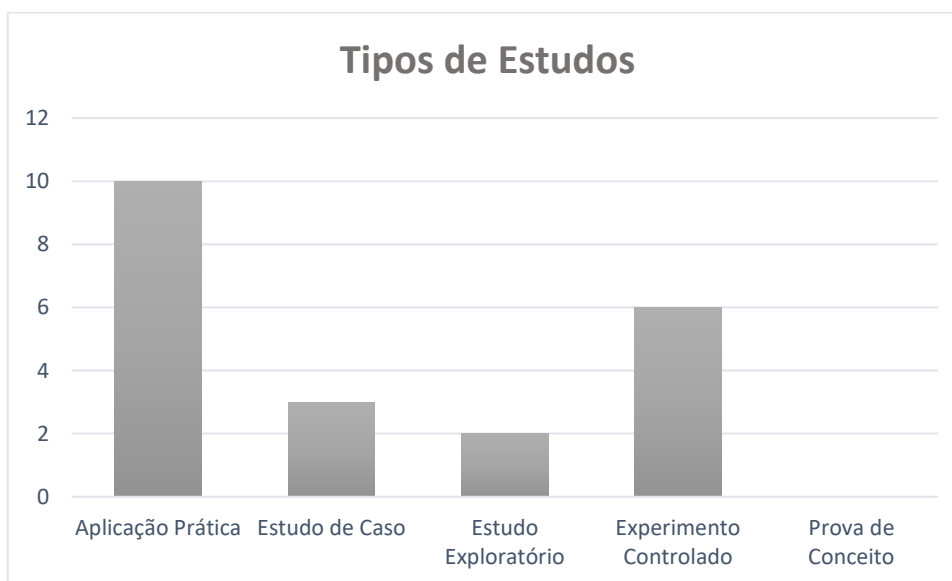
**Source:** Produzida pelos próprios autores.

### 6.8 Quais os tipos de estudos executados?

A Figura 10 apresenta os tipos de estudos executados nos artigos primários, sendo que a “Aplicação Prática” foi o tipo de estudo mais utilizado.

**Figura 10**

*Tipos de estudos*



**Source:** Produzida pelos próprios autores.

Para esta classificação, foram utilizadas as definições:

- **Experimento Controlado:** Forma de estudo experimental na qual o investigador tem controle sobre os principais aspectos do estudo e as variáveis independentes que estão sendo estudadas. Este tipo de estudo é caracterizado pelo controle sistemático das variáveis e do processo, tendo como objetivo confirmar teorias, conhecimento convencional, explorar relacionamentos, avaliar a predição de modelos ou validar medidas. Além disso, envolve a formulação de hipóteses, que precisam ser verificadas em relação aos resultados obtidos (Delamaro, Jino, & Maldonado, 2017; Wohlin et al., 2012).
- **Estudo de Caso:** Baseia-se na utilização de um ou mais métodos qualitativos ou não segue uma linha rígida de investigação. Consiste geralmente no estudo aprofundado de um único “caso” ou de “casos relacionados”, sendo executado em condições típicas, por exemplo, a partir de alguns projetos típicos representativos (Read, 2003).
- **Aplicação Prática:** Consideramos esta classificação ao encontrarmos trabalhos que são semelhantes a um estudo de caso, no entanto, com alguma lacuna no método de avaliação. Em outras palavras, aplicações que podem não ter usado dados reais, ou não foram executadas em ambiente real. Além disso, pode não ter existido uma avaliação, minimamente, qualitativa.
- **Prova de Conceito:** Termo utilizado para denominar um modelo prático que possa realizar um ensaio sobre um conceito (teórico) estabelecido por uma pesquisa ou artigo técnico. Pode ser considerado também uma implementação, em geral resumida ou incompleta, de um método ou de uma ideia, realizada com o propósito de verificar que o conceito ou teoria em questão é suscetível de ser explorado de uma maneira útil (Farias et al., 2019).
- **Estudo Exploratório:** Caracterizado pela flexibilidade, criatividade e informalidade que esse tipo de estudo permite ao pesquisador na busca em obter um maior conhecimento sobre um determinado tema ou problema de pesquisa. Muitos autores consideram os estudos exploratórios como um estágio preliminar no processo de pesquisa como um todo, servindo para coletar dados e informações (Hall & Rist, 1999).

Por fim, a tabela abaixo resume as principais características dos 21 trabalhos mapeados, incluindo o artigo de comparação usado no início da pesquisa. A coluna que mapeia a metodologia considera como “Sim” os trabalhos que apresentaram de fato uma metodologia, não apenas uma fase do processo de concepção de uma aplicação.

**Tabela 6**

*Síntese dos resultados*

| Artigo   | Abordou |    | Tipo de Estudo         | Apresentou alguma Metodologia de Desenvolvimento? | Alinhamento Estratégico |
|--|---------|----|------------------------|---|-------------------------|
|  | BI      | DM |                        |   |                         |
| Cheng et al. (2009)                                | X       | X  | Estudo de Caso         | Sim   |                         |
| Ruggieri, Pedreschi & Turini (2010)                |         | X  | Aplicação Prática      | Sim   |                         |
| Sharma, Osei-Bryson & Kasper (2012)                |         | X  | Experimento Controlado | Sim   |                         |
| Yu et al. (2012)                                   | X       | X  | Aplicação Prática      | Não   |                         |
| Han et al. (2013)                                  |         | X  | Aplicação Prática      | Não   |                         |
| Kohavi et al. (2013)                               | X       |    | Estudo Exploratório    | Sim   |                         |
| Wang & Sun (2013)                                  | X       | X  | Estudo Exploratório    | Sim   |                         |
| Ławrynowicz & Potoniec (2014)                      |         | X  | Experimento Controlado | Sim   |                         |
| More (2014)  |         | X  | Estudo de Caso         | Não   |                         |
| Costa et al. (2015)                                | X       |    | Experimento Controlado | Não   |                         |
| Puppala et al. (2015)                              | X       | X  | Aplicação Prática      | Não   |                         |
| Vitt & Xiong (2015)                                | X       | X  | Aplicação Prática      | Não   |                         |
| Costa et al. (2016)                                | X       |    | Experimento Controlado | Não   |                         |
| Kanavos et al. (2017)                              |         | X  | Aplicação Prática      | Não   |                         |
| Lin et al. (2017)                                  |         | X  | Aplicação Prática      | Sim   |                         |
| Santos et al. (2017)                               | X       |    | Experimento Controlado | Não   |                         |
| Sun et al. (2017)                                  |         | X  | Aplicação Prática      | Sim   |                         |
| Bautista (2018)                                    |         | X  | Aplicação Prática      | Sim   |                         |
| Bock et al. (2018)                                 |         | X  | Experimento Controlado | Não   |                         |
| Ju et al. (2018)                                   | X       | X  | Estudo de Caso         | Sim   |                         |
| Manigandan et al. (2019)                           |         | X  | Aplicação Prática      | Sim   |                         |
| Colaço Júnior et al. (2019) - Artigo de Comparação | X       |    | Estudo de Caso         | Sim   | X                       |

**Source:** Produzida pelos próprios autores.

## 7 Síntese narrativa

Nesta seção, além da discussão sobre experimentação, estratégia e resultados brutos, realizada nas seções anteriores, uma discussão dos principais aspectos e lições aprendidas sobre possíveis melhorias no processo de desenvolvimento Experimental de aplicações de *Business Intelligence* e *Data Mining* é apresentada.

Com relação ao alinhamento estratégico, como exceção limitada, foi possível apresentar apenas o processo publicado por Colaço Júnior et al. (2019), um artigo de controle deste estudo, o qual mescla a abordagem GQM+Strategies (vide Tabela 6) com uma metodologia de desenvolvimento ágil de aplicações de *Business Intelligence* proposta pelo autor, visando garantir o alinhamento estratégico. Apesar do avanço de disciplinar o alinhamento estratégico, o processo ainda não abrange soluções de *Data Mining* e não prevê uma validação experimental dos modelos de conhecimento, partes integrantes principais do objetivo deste trabalho.

Melhorias no processo de desenvolvimento de aplicações inteligentes são propostas desde a década de 50, quando o termo *Business Intelligence* foi utilizado pela primeira vez por Hans Peter Luhn, um pesquisador da IBM, no artigo intitulado “A Business Intelligence System” (LUHN, 1958). Nesse período, vários modelos de processo de descoberta de conhecimento foram propostos por pesquisadores e profissionais. Exemplos incluem: 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) e Ławrynowicz & Potoniec (2014).

Especificamente sobre o acoplamento de Data Mining às aplicações de BI, além da seminal KDD (*Knowledge Discovery in Databases*) (Fayyad et al., 1996), apesar de não terem sido desenvolvidas pela comunidade científica, CRISP-DM (Wirth & Hipp, 2000; Kurgan & Musilek, 2006) e SEMMA (Sample, Explore, Modify, Model, Assess) (Mariscal, Marban & Fernandez, 2020; Matignon, 2007) são opções, todavia, também não apresentam uma definição da integração dos conceitos de BI e *Data Mining*, bem como não dão cobertura aos aspectos relacionados ao planejamento estratégico da organização. Uma alternativa, não encontrada neste mapeamento para *Data Mining*, seria o uso de métodos auxiliares para o alinhamento estratégico, tais como os apresentados na Tabela 7.

**Tabela 7**

*Métodos auxiliares para o alinhamento estratégico*

| Método                | Descrição   | Fonte                   |
|-----------------------|---|-------------------------|
| <b>COBIT</b>          | O Control Objectives for Information and Related Technology (COBIT) é definido como um conjunto de diretrizes baseadas em auditoria para processos, práticas e controles de TIC voltadas à redução de riscos, busca pela integridade, confiabilidade e segurança da informação. | (Cobit, 2016)           |
| <b>BSC</b>            | Balanced Scorecard (BSC) é um método oriundo da governança corporativa, o qual exerce bem o papel de medição, mas não contempla boas práticas. Seus conceitos vêm sendo incorporados ao processo de plano estratégico de Tecnologia da Informação.                              | (Tonelli, et al., 2014) |
| <b>GQM+Strategies</b> | Abordagem sistemática que integra os objetivos de negócio, adaptando-os aos modelos de processos de software, produtos e perspectivas de interesse de qualidade, com base nas necessidades específicas do projeto.  | (Basili, et al., 2014)  |

**Source:** Produzida pelos próprios autores.

Com relação ao uso de uma abordagem experimental, esta é uma alternativa para atender os pressupostos da Ciência de Dados (Bock et al., 2018; Costa et al., 2015; Costa et al., 2016; Ławrynowicz & Potoniec, 2014; Santos et al., 2017; Sharma, Osei-Bryson & Kasper, 2012), uma vez que a aplicação de um método científico rigoroso coaduna com a tentativa de tornar a análise de dados uma ciência, com princípios que diminuem as ameaças à validade do conhecimento gerado.

No entanto, este trabalho evidenciou que ainda é tímido o uso do método científico no desenvolvimento de aplicações inteligentes. Apenas 28,57% dos trabalhos encontrados validaram suas soluções por meio de Experimentos Controlados, sendo a Aplicação Prática o método de pesquisa mais utilizado (47,61%). Esses números mostram a necessidade de aumentar o uso do método científico nessa área, com repetições de estudos que permitirão avaliar se outros pesquisadores chegarão, independentemente, aos mesmos resultados. Além disso, mesmo os que validaram, não seguiram ou propuseram uma metodologia de BI ou de Data Mining dirigida à experimentação, ou seja, que prevê uma fase experimental na validação dos resultados.

Em resumo, uma das principais lições aqui aprendidas é que qualquer abordagem, metodologia ou novo processo proposto deve se concentrar em atender aos objetivos estratégicos organizacionais e validar sua utilidade em experimentos ou estudos de caso bem executados. Isso está longe de propor um novo modelo de conhecimento e de executar alguns estudos de viabilidade sobre este. Pesquisadores e profissionais devem se concentrar em responder a perguntas tais como: Minha abordagem pode ser usada no mundo real?; Como meus resultados se generalizam para outras organizações?. Caso contrário, poderá existir um problema com a validade externa da abordagem proposta e será difícil passar do estado da arte para o estado da prática.

Por fim, destaca-se a dificuldade para classificação das abordagens encontradas como alinhadas estrategicamente ou como experimentais. Mesmo com a possibilidade do uso da mineração inteligente de textos e da utilização de robôs baseados nesta tecnologia, muita intervenção humana faz-se necessária para definir um estudo como experimental e averiguar alinhamento estratégico. Foi preciso ler os textos inteiros e interpretar minuciosamente os dados, com algumas associações não automáticas, as quais teriam que ser implementadas em uma inovadora ferramenta de extração. Por exemplo, a presença de um objetivo estratégico não implica que a metodologia usada previu o uso deste objetivo ou atendeu a este.

## 8 Ameaças à validade

As ameaças à validade podem limitar a habilidade de interpretar e/ou descrever resultados dos dados obtidos. Portanto, não há como desconsiderar as seguintes ameaças encontradas nesse estudo.

Validade de Construção: A *string* de busca e as questões de pesquisa utilizadas podem não cobrir a área de metodologias de BI e *Data Mining* dirigidas à estratégia e avaliadas experimentalmente. Para mitigar essa ameaça, tentou-se elaborar uma *string* mais abrangente possível, quanto aos termos que

pudessem ser usados na área, utilizando vários sinônimos. Tais termos foram identificados e refinados, com auxílio dos artigos de controle norteados pelo modelo PICO, utilizando trabalhos que interessavam à pesquisa (intervenção) e falsos positivos, com o objetivo de calibrar a *string* de busca. Além disso, foram consideradas as opiniões de três pesquisadores.

Validade Interna: (Extração de dados): Três pesquisadores foram responsáveis por extrair e classificar os dados de cada publicação. Logo, vieses ou problemas na extração dos dados podem ameaçar a validade da caracterização dos dados. (Viés de Seleção): Inicialmente, os artigos foram incluídos ou excluídos de acordo com julgamento dos próprios pesquisadores. Consequentemente, alguns estudos podem ter sido categorizados incorretamente. Para mitigar estas ameaças, as revisões da seleção e extração foram feitas por todos os pesquisadores envolvidos e as discordâncias encontradas foram resolvidas em uma votação final.

Validade Externa: O uso da língua inglesa pode ter contribuído para a não inclusão de possíveis documentos relevantes em outras línguas. Além disso, para esta área específica, podem existir guias e boas práticas de indústria não publicadas.

## 9 Conclusão

Neste trabalho, foi realizada um mapeamento sistemático, visando identificar e caracterizar as metodologias de desenvolvimento de aplicações de *Business Intelligence* e *Data Mining* dirigidas à estratégia e/ou que preveem avaliação Experimental, avaliando e evidenciando os trabalhos mais relevantes na área. Nesse caminho, o mapeamento compreendeu trabalhos publicados no período de 2009 a 2019. Sendo assim, a busca inicial retornou 841 estudos, dos quais, após serem avaliados, segundo critérios de inclusão e exclusão estabelecidos no protocolo de seleção, foram aceitos apenas 21 trabalhos.

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.

Assim, acredita-se que esta pesquisa apresenta resultados relevantes à academia e aos empreendedores, fornecendo evidências de que há uma lacuna sobre pesquisas de um método formal de desenvolvimento experimental de aplicações de BI e *Data Mining* direcionadas ao planejamento estratégico de uma organização.



Por fim, 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. Como trabalhos futuros, podem ser propostos métodos para criação de aplicações inteligentes alinhadas à estratégia e com validação experimental.

### Referências

- 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 (UFG)*, 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 Multiple-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., Haghghatkah, 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). *Semina 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.