

**UNIVERSIDADE NOVE DE JULHO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DE PROJETOS - PPGP**

**PROPOSIÇÃO DE UM MODELO PARA ESTRUTURAÇÃO DE PROJETOS DE
BIG DATA**

GUSTAVO GRANDER

**São Paulo
2021**

Gustavo Grander

**PROPOSIÇÃO DE UM MODELO PARA ESTRUTURAÇÃO DE PROJETOS DE BIG
DATA**

PROPOSED MODEL FOR BIG DATA PROJECT STRUCTURING

Tese apresentada ao Programa de Pós-Graduação
em Gestão de Projetos da Universidade Nove de
Julho – UNINOVE, como requisito parcial para
obtenção do grau de Doutor em Administração.

Orientador: Prof. Dr. Luciano Ferreira da Silva
Co-Orientador: Prof. Dr. Ernesto Del Rosário
Santibañez Gonzalez

São Paulo

2021



DEFESA DE TESE DE DOUTORADO

GUSTAVO GRANDER

Tese apresentada ao Programa de Pós-Graduação em Gestão de Projetos da Universidade Nove de Julho – UNINOVE, como requisito parcial para obtenção do grau de **Doutor em Administração**, pela Banca Examinadora, formada por:

São Paulo, 08 de outubro de 2021.

Luciano Ferreira da Silva

Presidente: Prof. Dr. Luciano Ferreira da Silva- Orientador

Membro: Prof. Dr. Ernesto Del Rosario Santibañez González - Co-orientador

Membro: Prof. Dr. Renato Penha (UNINOVE)

Membro: Prof. Dr. Flavio Santino Bizarrias (UNINOVE)

Assinado de forma digital por
PAULO SERGIO GONCALVES DE
OLIVEIRA11632733846
Data: 2021.10.17 11:56:51 -03:00

Membro: Prof. Dr. Paulo Sergio Gonçalves de Oliveira - (ANHEMBI MORUMBI)

DocuSigned by:

Delci Grapegia Dal Vesco

9596BD608B54415...

Membro: Profa. Dra. Delci Grapegia Dal Vesco – (UNIOESTE)

Grander, Gustavo.

Proposição de um modelo para estruturação de projetos de Big Data. / Gustavo Grander. 2021.

111 f.

Tese (Doutorado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2021.

Orientador (a): Prof. Dr. Luciano Ferreira da Silva.

Coorientador (a): Prof. Dr. Ernesto Del Rosário Santibañez González.

1. Projeto de Big Data. 2. Gerenciamento de projetos. 3. Sistemas de suporte à decisão. 4. Patente. 5. Modelo de utilidade.

I. Silva, Luciano Ferreira da. II. González, Ernesto Del Rosário Santibañez. III. Título.

CDU 658.012.2

AGRADECIMENTOS

À Universidade Nove de Julho (UNINOVE) pela oportunidade que me foi concedida para ser aluno do programa de *strictu sensu* em Gestão de Projetos.

Ao meu orientador, professor Dr. Luciano Ferreira da Silva por ter confiado em mim e por ter me desafiado cada dia a ser uma pessoa melhor. Tenho uma admiração enorme ao senhor.

Ao meu coorientador, professor Dr. Ernesto Del Rosário Santibañez Gonzalez por toda a dedicação e paciência durante a pesquisa.

Aos professores Dr. Renato Penha e Dr. Paulo Sergio Gonçalves de Oliveira que além de terem participado das bancas de qualificação e defesa, também participaram diretamente na construção do modelo.

Ao professor Dr. Flavio Santino Bizarrias pelas contribuições valorosas ao longo das bancas de qualificação e defesa.

Ao professor Dr. Rodrigo Cunha da Silva pelos apontamentos e sugestões realizadas durante a banca de qualificação.

À professora Dra. Delci Grapegia Dal Vesco por ter aceito mais uma vez contribuir com minha formação acadêmica, agora fazendo parte da minha banca de defesa.

Ao corpo docente do Programa de Pós-Graduação em Gestão de Projetos – PPGP, por tanto conhecimento compartilhado.

Aos meus pais Rosane e Roberto, pelo constante apoio para que essa conquista fosse possível.

Aos colegas do programa que sempre me incentivaram para avançarmos nessa jornada.

A todos profissionais que aceitaram participar da pesquisa por meio das entrevistas e contribuíram para que esse resultado fosse possível.

Tudo vale a pena, se a alma não é pequena.
Fernando Pessoa

RESUMO

Dados são ativos estratégicos que podem gerar resultados valiosos para as organizações quando tratados de forma adequada. Cada vez mais as organizações estão tomando decisões orientadas a dados e cada vez menos orientadas a intuição. Essa situação faz com que o sucesso na implantação de projetos de Big Data seja fundamental para o progresso das organizações. Com base nesse contexto, esta tese tem como objetivo propor um modelo gerencial para estruturação de projetos de Big Data. Para tanto, a pesquisa foi composta por seis fases que compuseram três artigos e uma reivindicação de modelo de utilidade. No primeiro estudo foi analisado como os Sistemas de Suporte à Decisão gerenciam Big Data para obter valor, estudo que resultou no primeiro artigo. No segundo estudo explorou-se a estrutura e o contexto de um projeto de Big Data. No terceiro estudo buscou-se compreender as características de patentes relacionadas a projetos de Big Data, estudo que resultou no segundo artigo. Por fim, no quarto estudo, foi validado um modelo gerencial para estruturação de um projeto de Big Data, levando em consideração as fases estratégicas, tecnológica e analítica, este que, junto com o segundo estudo, resultaram no terceiro artigo. O modelo proposto nesta tese, pelo aspecto gerencial, pode ser aplicado para projetos de diferentes áreas e setores. Como contribuição teórica, destaca-se as técnicas de aprendizado de máquina, e tecnologias como computação em nuvem, que vem sendo utilizadas em Sistemas de Suporte de Decisão, e também a constatação de grupos de patentes direcionadas à computação em nuvem, otimização de soluções e estruturas de armazenamento e compartilhamento de dados. Como contribuição prática, destaca-se a reivindicação de propriedade intelectual do modelo de utilidade, modelo que poderá ser utilizado por profissionais e acadêmicos que estão buscando, constantemente, soluções baseadas na utilização de dados para problemas cada vez mais significativos em nossa sociedade.

Palavras-chave: Projeto de Big Data; Gerenciamento de projetos; Sistemas de suporte à decisão; Patente; Modelo de utilidade.

ABSTRACT

Data are strategic assets that can generate valuable results for organizations when properly treated. More and more organizations are making decisions driven by data and less and less by intuition. This situation makes the successful implementation of Big Data projects essential for the progress of organizations. Based on this context, this thesis aims to propose a management model for structuring Big Data projects. Therefore, the research consisted of four studies comprising three articles and a utility model claim. The first study analyzed how Decision Support Systems manage Big Data to obtain value and resulted in the first article. The second study explored the structure and context of a Big Data project. The third study sought to understand the characteristics of patents related to Big Data projects resulting in the second article. Finally, in the fourth study, a managerial model for structuring a Big Data project was validated, considering the strategic, technological and analytical phases. This latter study, together with the second, constitutes the third article. Due to its managerial aspect, the model proposed in this thesis can be applied to projects from various areas and sectors. As a theoretical contribution, machine learning techniques and technologies, such as cloud computing, have been used in Decision Support Systems and for identifying patent groups directed to cloud computing, optimization of solutions, and structures data storage and sharing. As a practical contribution, the intellectual property claim of the utility model stands out since our model can be employed by professionals and academics who seek solutions to increasingly significant problems in our society through the use of data.

Keywords: Big Data Project; Project management; Decision support systems; Patent; Utility model.

LISTA DE TABELAS

Tabela 1 - Fatores de sucesso de projetos de Big Data	21
Tabela 2 - Fluxo do filtro da RSL	30
Tabela 3 - Quantidade de artigos de acordo com os anos	30
Tabela 4 - Descrição dos entrevistados	32
Tabela 5 - Patentes eliminadas e repetidas da amostra.....	37
Tabela 6 - Resumo da análise do corpus	37
Tabela 7 – Entrevistas com grupos focais	39
Tabela 8 - Sistemas de Suporte à Decisão gerenciam Big Data para obter valor por meio de quais técnicas e tecnologias?	40
Tabela 9 - Sistemas de Suporte à Decisão têm sido aplicados para resolver quais tipos de problemas?.....	42
Tabela 10 - Fluxo de projetos de Big Data.....	46
Tabela 11 - Mapa de intensidade do número de registros em países/regiões relacionados a projetos de Big Data	49
Tabela 12 - Requerentes com mais de uma patente registrada.....	56

LISTA DE FIGURAS

Figura 1 - Fluxo de trabalho do projeto de Big Data.....	24
Figura 2 - Estrutura para a implementação de projetos de Big Data em empresas	24
Figura 3 - Fases da pesquisa	27
Figura 4 – Produtos provenientes da tese	28
Figura 5 - Categorias iniciais de análise	33
Figura 6 – Dendograma gerado pelo MCHD	50
Figura 7 – AFC da amostra	51
Figura 8 – Sugestões de melhorias feitas ao longo da discussão	57
Figura 9 – Dúvidas levantadas ao longo da discussão.....	58
Figura 10 – Pedido de modelo de utilidade	59
Figura 11 – Fases do modelo gerencial para estruturação de projetos de Big Data	60

SUMÁRIO

LISTA DE TABELAS.....	IX
LISTA DE FIGURAS.....	X
1 INTRODUÇÃO	13
1.1 PROBLEMA DE PESQUISA	14
1.2 OBJETIVOS	15
1.3 JUSTIFICATIVA	16
2 FUNDAMENTAÇÃO TEÓRICA.....	18
2.1 PROJETOS DE BIG DATA.....	18
2.2 ESTRUTURA DE DECISÃO EM PROJETOS DE BIG DATA	22
3 PROCEDIMENTOS METODOLÓGICOS	26
3.1 FASES DA PESQUISA	26
3.2 ENTREGAS DA TESE.....	28
3.3 PROCEDIMENTO METODOLÓGICO DO ESTUDO 1 – ESTRUTURA DE DECISÃO COM BIG DATA.....	29
3.4 PROCEDIMENTO METODOLÓGICO DO ESTUDO 2 – INVESTIGAÇÃO SOBRE PROJETOS DE BIG DATA.....	31
3.5 PROCEDIMENTO METODOLÓGICO DO ESTUDO 3 – PATENTES DE PROJETOS DE BIG DATA.....	35
3.6 PROCEDIMENTO METODOLÓGICO DO ESTUDO 4 – VALIDAÇÃO DO MODELO	38
4 RESULTADOS DOS ESTUDOS	40
4.1 RESULTADOS DO ESTUDO 1 – ESTRUTURA DE DECISÃO COM BIG DATA.....	40
4.2 RESULTADOS DO ESTUDO 2 – INVESTIGAÇÃO SOBRE PROJETOS DE BIG DATA.....	45

4.3	RESULTADOS DO ESTUDO 3 – PATENTES DE PROJETOS DE BIG DATA..	49
4.4	RESULTADOS DO ESTUDO 4 – VALIDAÇÃO DO MODELO	56
5	PRODUTO TECNOLÓGICO	59
6	CONCLUSÃO.....	62
	REFERÊNCIAS	66
	APÊNDICE A	79
	APÊNDICE B.....	97
	APÊNDICE C	111

1 INTRODUÇÃO

O Big Data expressa um aparente paradoxo em torno de seu conceito, pois ao mesmo tempo em que pode ser utilizado como uma ferramenta poderosa para lidar com vários males da sociedade, oferecendo o potencial de novos *insights* em áreas como de pesquisas médicas, combate ao terrorismo e mudanças climáticas, ele também pode permitir invasões de privacidade, diminuição das liberdades civis e aumento do controle estatal e corporativo (Boyd & Crawford, 2012). Paralelo a isso, existe um desafio em gerenciar Big Data devido a um volume cada vez maior e barato de armazenamento de dados (Demirkan & Delen, 2013).

Nesse contexto, pode-se afirmar que a maioria dos conjuntos de dados dos quais cientistas e pesquisadores conseguiram extrair um significado real ainda é muito pequeno em comparação à proporção de dados que podem ser capturados (Dobre & Xhafa, 2014). A análise de Big Data, por sua vez, compreende as fases de geração, aquisição, armazenamento e análise de dados, podendo fornecer valores úteis em cada fase por meio de julgamentos, sugestões, suporte ou decisões (Chen, Mao, & Liu, 2014). O Big Data, ainda, impulsiona uma nova geração de tecnologias e arquiteturas projetadas para extrair valor econômico (Gantz & Reinsel, 2011).

Neste contexto de obtenção de valor por meio de Big Data, surge o que Manyika *et al.* (2011) descrevem como *data-driven mind-set*, que também é descrito por Brynjolfsson, Hitt e Kim (2011) como *data driven decision making*. Decisões orientadas por dados, portanto, com base em evidências, têm se sobressaído quando comparadas com decisões com base em intuição (McAfee & Brynjolfsson, 2012). Qualquer atividade envolvendo clientes poderia se beneficiar de análise de Big Data (Russom, 2011), e avanços nas técnicas analíticas, especialmente no aprendizado de máquina, proporcionam um grande catalisador para lidar com análises de grandes conjuntos de dados (Murdoch & Detsky, 2013).

Não obstante à relevância dos avanços supracitados sobre capacidade de armazenamento e análise, ainda se faz necessário melhorar a capacidade de gerenciar projetos de Big Data. Dutta e Bose (2015) apontam que a cultura da tomada de decisão orientada a dados é essencial para o sucesso de um projeto de Big Data, porém muitos projetos de Big Data não geram o resultado esperado ou sequer são concluídos, ocasionando altos custos e prejuízos para as empresas (Barham & Daim, 2018). Portanto, apesar do progresso significativo na área de Big Data realizado nos últimos anos, praticantes destacam importantes oportunidades de

pesquisa, principalmente relacionadas a guias, modelos ou até metodologias (Lara, De Sojo, Aljawarneh, Schumaker, & Al-Shargabi, 2020).

Essa necessidade de um modelo para auxiliar os departamentos de TI não deve servir apenas para a implementação de um projeto de Big Data, mas também na melhor utilização possível dos *outputs* desse tipo de projeto para atender aos objetivos de negócios (Mousanif, Sabah, Douiji, & Sayad, 2016). Cabe reforçar que embora avanços tenham sido feitos nos algoritmos e tecnologias usados para executar as análises, pouco foi feito para determinar como a equipe deve trabalhar em conjunto para realizar um projeto de Big Data (Saltz, Shamshurin, & Connors, 2017).

1.1 PROBLEMA DE PESQUISA

Dados são ativos estratégicos, mas não possuem valor se não forem utilizados de forma construtiva e adequados para a obtenção de resultados valiosos para as organizações (Marr, 2015). Da mesma forma que muito esforço vem sendo direcionado ao desenvolvimento de projetos de Big Data, há uma alta taxa de problemas e não atingimento de objetivos estabelecidos e desejados nesses projetos, bem como estudos divulgando taxas alarmantes de fracasso em projetos de Big Data (Berman, 2018).

Os recursos de Big Data são complexos, difíceis de organizar em uma estrutura viável, que é facilmente desestruturada e, quando desestruturada, pode não ser corrigida facilmente (Berman, 2018). As organizações, por sua vez, possuem apetite por Big Data e há evidências crescentes de que os investimentos em soluções de Big Data nem sempre levam à derivação do valor pretendido (Surbakti, Wang, Indulska, & Sadiq, 2019). Apesar da oportunidade de crescimento de projetos de Big Data, há uma desconexão entre o potencial de agregação de valor em iniciativas digitais e a implementação bem-sucedida desses tipos de projeto (Mielli & Bulanda, 2019).

O motivo pelo qual esses projetos falham em produzir os resultados desejados também está na forma como foram desenvolvidos, uma vez que as organizações, geralmente, concentram-se nos próprios dados e *Data Analytics*, sem foco na tomada de decisão, que é o uso real do Big Data (Chiheb, Boumahdi, & Bouarfa, 2019). Cada vez mais as organizações precisam realinhar continuamente suas práticas de trabalho, modelos organizacionais e interesses das partes envolvidas para colher os benefícios de um projeto de Big Data (Günther,

Mehrizi, Huysman, & Feldberg, 2017). Desse modo, o uso eficaz de Big Data tem a capacidade de transformar dados em *insights* açãoáveis (Surbakti, 2021), o que torna explícito que a complexidade dos dados, ferramentas, técnicas e recursos envolvidos demandem um nível adequado de gerenciamento de projetos para que as organizações aproveitem totalmente o Big Data (Becker, 2017).

Essa problemática de insucesso condiciona um aumento da necessidade de modelos ou roteiros que auxiliem pessoas, não apenas na implementação de um projeto de Big Data, mas também, no aproveitamento máximo para atender aos objetivos de negócios (Mousanif, Sabah, Douiji, & Sayad, 2014). Nesse sentido, Saltz e Shamshurin (2016) inferem que não existe um padrão para a execução desse tipo de projeto, afirmando, ainda, que modelos, roteiros ou até uma metodologia de processo aprimorada seria útil. Esta tese, portanto, apropria-se da problemática de alta taxa de insucesso em projetos de Big Data e da ausência de um guia para estruturação desse tipo de projeto.

Os fatores negativos de projetos de Big Data, portanto podem resultar em sistemas de computação de análise de dados dispendiosos que não manipulam adequadamente os dados, e que não são operacionalizados de maneira adequada ou benéfica. Por isso, esforços com o intuito de promover soluções e alternativas que visem melhores resultados em projetos de Big Data devem ser reconhecidos.

Vale destacar que, no contexto de projetos de Big Data, as organizações muitas vezes priorizaram fatores como tempo, esforço e recursos na aquisição de soluções tecnológicas de ponta, mas a padronização, proteção e preparação dos metadados pode ter sido negligenciada ou ser insuficiente. Pesquisas sobre Big Data, normalmente, concentram-se mais na melhoria de modelos e algoritmos de dados, e pouco no uso de uma abordagem padrão para executar projetos (Ahangama & Poo, 2015). Assim, com base no cenário exposto, esta tese por estudos adota a seguinte questão de pesquisa: Como projetos de Big Data devem ser estruturados?

1.2 OBJETIVOS

Objetivo geral da tese:

- Propor um modelo gerencial para estruturação de projetos de Big Data.

Objetivos específicos da tese:

- Analisar como os Sistemas de Suporte à Decisão gerenciam Big Data para obter valor.
- Realizar uma análise exploratória de um banco de dados de patentes para coletar informações sobre as características das patentes registradas relacionadas a projetos de Big Data.
- Desenvolver um modelo gerencial para estruturação de projetos de Big Data.

1.3 JUSTIFICATIVA

Apesar do aumento em investimentos na área de Big Data, ainda há um longo caminho de amadurecimento para se chegar a um nível de excelência de entregas em projetos de Big Data. Em linha com este raciocínio, constatou-se que 87% dos projetos de ciência de dados nunca chegam à produção, pois é muito comum as equipes envolvidas tentarem “ferver o oceano”, referindo-se à intenção de quererem resolver um problema muito abrangente (Venturebeat, 2019). Adicionalmente, Gartner (2019) aponta que 80% dos *insights* analíticos não entregarão resultados de negócios até 2022 e 80% dos projetos de Inteligência Artificial permanecerão na alquimia administrados por magos, referindo-se a profissionais cujas habilidades não podem ser escaladas nas organizações.

Embora a maioria das organizações tenham acesso a Big Data, elas ainda não têm a capacidade de usá-lo de forma eficaz (Günther *et al.*, 2017). O Big Data é um recurso-chave para a tomada de decisão eficaz, mas seu crescimento exponencial sem precedentes (particularmente de dados não padronizados) cria uma série de desafios e oportunidades (Borthick & Pennington, 2017). Suas fontes multicanais e a variabilidade na qualidade dificultam a obtenção do uso econômico (Wilkin, Ferreira, Rotaru, & Gaerlan, 2020).

Empresas mais produtivas adotam o Big Data com mais rapidez que seus pares do setor que não adotam o Big Data, e investimentos de Big Data em mão de obra e arquitetura de TI ocorrem de forma complementar aos ganhos de produtividade (Bughin, 2016). Barham e Daim (2020) destacam que estudos recentes apontam que os projetos de Big Data estão enfrentando um alto percentual de falhas e não há uma razão predominante para isso, mas sim vários fatores complexos, incluindo fatores humanos relacionados a habilidades e percepção, fatores técnicos relacionados às tecnologias usadas para implementar Big Data e a natureza dos próprios dados, fatores legais relacionados a quem é o proprietário dos dados e como usá-los, fatores

organizacionais relacionados às estratégias de dados, além de alinhamento com disponibilidade e os objetivos organizacionais.

Os aspectos característicos de projetos de Big Data geralmente têm entradas incertas no projeto (por exemplo, quais dados podem ser relevantes) e também resultados incertos (ou seja, haverá informações derivadas da análise dos dados) (Saltz *et al.*, 2017). No campo da propriedade intelectual é possível perceber uma convergência de esforços para a resolução de problemas típicos de projetos de Big Data. Além de muito útil para identificar tendências tecnológicas em campos específicos (Abbas, Zhang, & Khan, 2014), as análises de patentes podem indicar quais tipos de problemas são mais recorrentes em uma área específica, na medida em que os registros de patentes se propõem a resolver algum problema prático.

Conforme argumentado pelo inventor Wang (2019), as falhas dos projetos de Big Data têm razões não técnicas, como por exemplo, quando metas do sistema de Big Data e cenários de negócios não são claros e são mais orientados para a tecnologia que para os negócios, e também razões técnicas, como por exemplo, a qualidade dos dados normalmente é ruim e a padronização não é suficiente. Os inventores Dietrich e Reiner (2017) complementam afirmando que as soluções convencionais de análise de dados estão se tornando cada vez mais limitadas devido aos tamanhos crescentes e estruturas variáveis dos conjuntos de dados com os quais essas soluções são aplicadas. Tais limitações incluem a falta de capacidade de estimar adequadamente o custo da solução de análise de dados, a inflexibilidade e a falta de otimização da solução, e a dificuldade de colocar a solução em operação.

As organizações, portanto, devem compreender o papel do Big Data associado à tomada de decisão, pois o Big Data pode guiar decisões auxiliando na identificação de oportunidades, sempre com o uso de estratégias adequadas aliadas à experiência humana (Poleto, Carvalho, & Costa, 2017). A aplicação do Big Data nos negócios modernos permite que as empresas tomem decisões mais rápidas e inteligentes, alcançando uma vantagem competitiva real (Chiheb *et al.*, 2019).

Esta pesquisa se justificou, portanto, pela contribuição teórica e prática para o campo de Big Data, mais especificamente para projetos de Big Data. Dentre as contribuições teóricas, destaca-se o panorama apresentado sobre técnicas e tecnologias utilizadas por meio de Sistemas de Suporte à Decisão, e também a identificação das grandes áreas de aplicações de patentes relacionadas a projetos de Big Data. Ademais, o modelo gerencial desenvolvido ao longo da tese, caso seja aprovada a reivindicação de propriedade intelectual, torna-se a principal contribuição prática.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção de fundamentação teórica aborda algumas definições de Big Data, assim como uma caracterização de um projeto de Big Data e, consequentemente, a estrutura de decisão em projetos de Big Data.

2.1 PROJETOS DE BIG DATA

Big Data pode ser definido como um grande conjunto de dados que é difícil de ser armazenado, processado, analisado e entendido por meio de ferramentas tradicionais de processamento de banco de dados (Huang & Chaovalltwongse, 2015). O Big Data surgiu como uma mudança paradigmática sobre como as organizações tomam decisões (Mortenson, Doherty, & Robinson, 2015). Wamba, Ngai, Riggins e Akter (2017) destacam que o Big Data pode ser adotado nos níveis estratégico, tático e operacional, podendo, assim, contribuir com a melhora da tomada de decisão.

Ao definir Big Data, Russom (2011) cita a caracterização reconhecida como os ‘3 Vs’ do Big Data (Volume, Variedade e Velocidade). O ‘Volume’ está relacionado à grande quantidade de dados que consomem grande capacidade de armazenamento ou implicam em um grande número de registros. A ‘Variedade’ é representada pelo fato de que os dados são gerados a partir de uma grande variedade de fontes e formatos e contêm campos de dados multidimensionais, incluindo dados estruturados e não estruturados. A ‘Velocidade’ está relacionada à frequência ou velocidade de geração de dados e/ou frequência de entrega de dados.

De forma incremental à definição proposta por Russom (2011), outros autores acrescentaram outros ‘V’s, como forma de complementarem a definição de Big Data. Demchenko, Grossi, de Laat e Membrey (2013) incluíram os ‘Vs’ de ‘Valor’ e ‘Veracidade’ chamando ao que se conhece como os ‘5 Vs’ do Big Data. Para os autores, o Valor é uma característica importante dos dados que é definida pelo valor agregado que os dados coletados podem trazer para o processo pretendido, atividade ou análise. A Veracidade inclui a consistência e a confiabilidade dos dados, isso permite que os dados usados sejam confiáveis, autênticos e protegidos contra acesso não autorizado e modificação.

Khan *et al.* (2019) ao tentarem trazer uma definição mais atual de Big Data propuseram uma caracterização de Big Data com 51 Vs. Segundo os autores, a disciplina de Big Data está evoluindo em torno dos Vs, e com recursos de computação em rápido crescimento, informações sendo geradas, armazenadas e acessadas por uma vasta gama de aplicativos, a definição muda por influência do ambiente, que está em constante mudança.

Um projeto de Big Data pode ser definido como um projeto intensivo de dados que apresenta problemas de grande escala com restrições de volume, variedade e velocidade dos dados (Becker, 2017). Os projetos de Big Data são caracterizados pelo grande porte, alta complexidade e tecnologia inovadora, o que agrava quaisquer deficiências nas práticas de gestão, pessoal ou processo (Kappelman, McKeeman, & Zhang, 2007). As diferenças entre projetos de Big Data e, por exemplo, projetos de desenvolvimento de *software* estão relacionadas, principalmente, com processos organizacionais desejados (Saltz, 2015).

Ainda há o desafio de determinar se os dados a serem analisados não têm problemas de incompatibilidade de tempo ou qual é o nível aceitável de qualidade de dados (Kaisler, Armour, Espinosa, & Money, 2013). Essas características não possuem paralelo em um projeto de desenvolvimento de *software* (Saltz *et al.*, 2017). Há uma recorrente discussão acerca de um projeto de Big Data bem-sucedido, contudo a complexidade desse tipo de projeto faz com que as soluções apresentadas para tal objetivo, sejam das mais variadas.

Por meio de uma abordagem de engenharia integrada em um projeto de Big Data, Jin, Wah, Cheng e Wang (2015) sugerem uma sequência abrangente de etapas, desde a coleta de requisitos, seguido do estabelecimento de uma estrutura adequada para processamento de dados, então adotando um modelo de gestão de cima para baixo e, por fim, resolvendo o problema inteiro por meio de uma solução integrada, em vez de buscar sucessos isolados.

Um estudo que explorou dificuldades e desafios enfrentados por alunos de engenharia, ao realizarem projetos de Big Data como parte de seu currículo, foi desenvolvido por Lara *et al.* (2020). Os autores apresentaram uma série de lições aprendidas como forma de auxiliar o desenvolvimento de futuros projetos de Big Data, e entre as lições aprendidas, destaca-se a evidência de maior criticidade para etapas de armazenamento e análise de dados, alto risco de baixa qualidade dos dados coletados e uma alta curva de aprendizagem na fase inicial do projeto.

Outra linha de discussão sobre projetos de Big Data, diz respeito a fatores de sucesso. Dois trabalhos os quais buscaram investigar fatores que impactam no sucesso de projetos de

Big Data são apresentados (Tabela 1) e é possível verificar certa convergência com as duas estruturas apresentadas.

Em uma pesquisa sobre como as equipes trabalham juntas para executar projetos de Big Data, Saltz e Shamshurin (2016) inferem que não há um padrão acordado para a execução desses projetos, mas há um foco crescente de pesquisas nesta área. Os autores identificaram 33 fatores críticos para a execução de esforços de Big Data, que são agrupados por seis características identificadas de uma organização de Big Data madura que são: Dados, Governança, Processos, Objetivos, Equipe e Ferramentas. Também em uma investigação de fatores críticos em projetos de Big Data, Gao, Koronios e Selle (2015) apresentaram 27 fatores divididos em seis fases, sendo elas: Negócio, Dados, Análise, Implementação, Medição e Gerais.

Há também resultados indicando a importância da escolha da abordagem na gestão de projetos de Big Data, que é condicionada não apenas pelo tamanho e criticidade do projeto, mas também pela dinâmica do ambiente (Franková, Drahošová, & Balco, 2016). Para Jin *et al.* (2015) as condições necessárias para que um projeto de Big Data seja bem-sucedido estão relacionadas a ter requisitos técnicos, sociais e econômicos muito claros, ter estruturas de dados pequenas o suficiente para caracterizar o comportamento a ser analisado e deve-se buscar modelo de gestão de cima para baixo para promover resoluções abrangentes por meio de uma abordagem integrada. Aliás, na medida em que o Big Data atinge muitos domínios da vida diária e tem impactos profundos no mundo social, a reflexão em torno da prática do Big Data torna-se cada vez mais relevante (Chen & Quan-Haase, 2018).

Dada a importância de estruturas de auxílio à tomada de decisão em projetos de Big Data, estudos propõem modelos para sanar esse *gap*. Mousanif *et al.* (2014) e Mousanif *et al.* (2016) apresentam uma abordagem holística para as organizações criarem projetos de Big Data e obterem o máximo de recompensa de seus dados. Os autores apontam para a necessidade de englobar fases de planejamento, implementação e pós-implementação do projeto. Dutta e Bose (2015) propõem outra estrutura de implantação de projetos de Big Data baseada nas fases de base estratégica, análise de dados e implementação. A estrutura foi aplicada em uma indústria de manufatura da Índia, com destaque para a relevância da compreensão clara do problema de negócios, assim como etapas bem específicas para a execução do projeto.

Tabela 1 - Fatores de sucesso de projetos de Big Data

Características	Fatores	Gao <i>et al.</i> (2015)	Saltz e Shamshurin (2016)
Dados	Gerenciamento / propriedade de dados e qualidade de dados	x	x
	Integração e segurança de dados	x	x
	Dados não estruturados / estruturados (combinação)	x	x
	Coleta de documentos / Identificação e acesso a fontes	x	x
	Visualização e Virtualização	x	
	Interpretação dos resultados analíticos	x	
	Adaptar princípios arquitetônicos	x	
Governança	Representatividade dos dados		x
	Prioridade de gestão / patrocínio / suporte	x	x
	Alinhamento da estratégia de Big Data (com a visão da organização)	x	x
	Processo de gerenciamento de projetos definido		x
	Unidade de negócio independente	x	
	Estrutura organizacional bem definida		x
	Estrutura de TI flexível	x	
	Gerenciamento de desempenho		x
	Proteção de dados e privacidade por design		x
	Cultura orientada por dados		x
Processo	Big Data como instrumento estratégico	x	
	Estratégia de informação para Big Data	x	
	Estreita colaboração entre TI e negócios		x
	Comunicação sobre os dados e iniciativas		x
	Flexibilidade e agilidade, com liberdade para experimentação (iteração)	x	x
	Foco no gerenciamento de mudanças	x	x
	Dificuldade do projeto explorada e comunicada		x
Objetivos	Clareza das entregas do projeto (claras ou ambíguas)	x	x
	Prazos apropriados	x	
	Integração de novas soluções	x	
	Terceirização	x	
	Concentração em pequenos projetos e perguntas conhecidas		x
Equipe	Caso de negócios especificado	x	x
	Estudo de viabilidade		x
	Análise de <i>gap</i> de habilidade		x
	Escopo bem definido - o que é entendido pela equipe	x	x
	Resultado mensurável do projeto	x	x
	Desenvolvimento de habilidades / treinamento	x	x
	Habilidades e capacidade de se organizar quando necessário		x
Ferramentas	Ciência de dados, tecnologia, negócios e habilidades de gerenciamento	x	x
	Equipe multidisciplinar (ou seja, em diferentes departamentos)	x	x
	Coordenação das partes interessadas / entendimento compartilhado		x
	Investimento em infraestrutura, tecnologia e ferramentas de TI		x
	Investimento em fontes de dados e armazenamento de dados	x	x
	Tecnologia de geração de relatórios e visualização		x
	Tecnologia de descoberta		x
Ferramentas inovadoras de análise		x	
<i>Hardware</i> adequado		x	

Fonte: Elaborado a partir de Gao *et al.* (2015) e Saltz e Shamshurin (2016).

Ainda há o desafio de determinar se os dados a serem analisados não têm problemas de incompatibilidade de tempo ou qual é o nível aceitável de qualidade de dados (Kaisler *et al.*, 2013). Essas características não possuem paralelo em um projeto de desenvolvimento de *software* (Saltz *et al.*, 2017).

Dada a complexidade, o fenômeno da tendência ao fracasso de projetos de Big Data remetem a alguns fatores, incluindo fatores humanos relacionados a habilidades e percepção, fatores técnicos relacionados às tecnologias usadas para implementar Big Data, a natureza dos próprios dados, fatores legais relacionados a quem possui os dados e como usá-los, fatores organizacionais relacionados às estratégias de dados, alinhamento com os objetivos organizacionais e disponibilidade de fundos (LaValle *et al.*, 2011).

Para obter resultados positivos em projetos de Big Data, profissionais e pesquisadores não devem se concentrar apenas no aspecto de dados e análise de dados ao estudar e desenvolver um projeto de Big Data, mas também ir além disso, e estudar o aspecto de decisão desse projeto (Chiheb *et al.*, 2019). Por isso o próximo tópico aborda alguns aspectos da decisão em projetos de Big Data.

2.2 ESTRUTURA DE DECISÃO EM PROJETOS DE BIG DATA

Para explorar Big Data, existem desafios nos níveis de dados, modelo e sistema, como destacam Wu, Zhu, Wu e Ding (2014). No nível de dados, as fontes de informação autônomas e a variedade de ambientes de coleta de dados, muitas vezes, resultam em dados com condições complicadas, como valores ausentes ou incertos. No nível do modelo, o principal desafio é gerar modelos globais combinando padrões descobertos localmente para formar uma visão unificadora. No nível do sistema, o desafio essencial é que uma estrutura de mineração de Big Data precisa considerar relacionamentos complexos entre amostras, modelos e fontes de dados, junto com suas mudanças em evolução com o tempo e outros fatores possíveis.

A cultura da tomada de decisão orientada a dados é essencial para o sucesso de um projeto de Big Data (Dutta & Bose, 2015). Soluções específicas têm sido apresentadas, como por exemplo, o desenvolvimento de uma abordagem de tomada de decisão baseada na metodologia difusa AHP-TOPSIS que foi desenvolvida para selecionar uma solução em nuvem apropriada para gerenciar projetos de Big Data (Boutkhoun, Hanine, Agouti, & Tikniouine, 2017). Também foi proposta uma plataforma de suporte ao gerenciamento e execução de

projetos com ênfase na parte analítica e de relatórios pelo uso das tecnologias de Business Intelligence (BI) e Big Data, essa plataforma auxilia a tomada de decisão bem sucedida em projetos (Pondel & Pondel, 2016).

Um importante processo de tomada de decisão foi desenhado por Simon (1960), processo esse que envolve três fases: inteligência, design e escolha (o modelo IDC), de modo que a fase I refere-se à inteligência que busca no ambiente condições, problemas ou oportunidades, chamando para decisão. Representado pela letra D, fase que se refere ao desenvolvimento e análise de soluções alternativas para o problema ou oportunidade e, por fim, a fase representada pela letra C refere-se à seleção de uma ou mais alternativas disponíveis. O modelo IDC não apenas influenciou outros modelos relevantes de tomada de decisão, como o apresentado por Mintzberg, Raisinghani e Theoret (1976) com as fases de identificação, desenvolvimento e seleção, como também influenciou modelos de tomada de decisão aplicados ao Big Data.

Elgendi e Elragal (2016) propuseram a estrutura B-DAD que integra ferramentas, arquitetura e análise de Big Data ao modelo IDC. No estudo, os autores identificaram produtos promocionais que deveriam ser oferecidos à venda e a influência das mídias sociais. Poleto *et al.* (2017) apresentaram um modelo integrado combinando elementos necessários para aplicar o conceito de Big Data durante as fases do processo de tomada de decisão. O modelo é baseado na integração de Big Data, BI e o *Decision Support System* (DSS) com o modelo IDC. Por fim, Chiheb *et al.* (2019) apresentam o modelo BD-Da, um modelo conceitual que descreve os níveis que devem ser considerados para desenvolver um projeto de Big Data com o objetivo de resolver um problema o qual exige uma decisão.

Ao construírem um caso para utilizar a análise de Big Data como uma base para a tomada de decisão orientada por dados em redes de cadeia de suprimentos que suportem economia circular, os autores Gupta, Chen, Hazen, Kaur e Gonzalez (2018) evidenciaram que as relações de colaboração com todas as partes interessadas ajudam a obter o acesso desejado aos dados relevantes e garantem uma análise de Big Data eficaz. Dada a importância de estruturas de auxílio à tomada de decisão, também existem propostas de roteiros específicos para a implantação de projetos de Big Data. Uma abordagem holística para as organizações criarem projetos de Big Data e obterem o máximo de recompensa de seus dados foi apresentada de modo a englobar fases de planejamento, implementação e pós-implementação do projeto (Mousanif *et al.*, 2014; Mousanif *et al.*, 2016) (Figura 1).

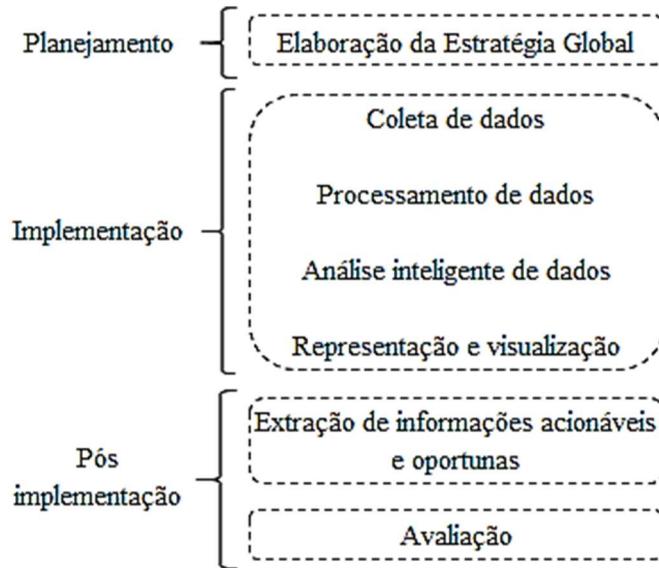


Figura 1 - Fluxo de trabalho do projeto de Big Data

Fonte: Mousanif *et al.* (2016).

Outra estrutura de implantação de projetos de Big Data foi proposta baseada nas fases de base estratégica, análise de dados e implementação (Dutta & Bose, 2015) (Figura 2). A estrutura foi aplicada em uma indústria de manufatura da Índia, com destaque para a relevância da compreensão clara do problema de negócios, assim como etapas bem específicas para a execução do projeto.



Figura 2 - Estrutura para a implementação de projetos de Big Data em empresas

Fonte: Dutta & Bose (2015).

É latente a convergência na percepção do que Chiheb *et al.* (2019) chamaram de aspecto da decisão do projeto, que vai muito além do foco exclusivo nos dados. O aspecto da decisão no projeto contempla áreas que abrangem um espectro mais amplo que a visão analítica, mas também uma importante fase estratégica de planejamento e definição de premissas, e uma fase de extração de informações valorosas que dão suporte no entendimento de fenômenos. Na próxima seção, detalha-se o procedimento metodológico da pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

A natureza da presente pesquisa se caracteriza como uma pesquisa aplicada dirigida à solução de um problema específico. Uma atitude consciente e crítica com relação aos procedimentos e paradigmas de pesquisa podem gerar avanços importantes às áreas de conhecimento destinadas (Silva, Russo, & Oliveira, 2018).

Nesta seção são descritas as seis fases que sustentaram a pesquisa e como estas delinearam o objetivo geral da tese. Para os quatro estudos contidos na pesquisa, são apresentados os procedimentos metodológicos, o método de pesquisa, a técnica de coleta de dados, a técnica de análise de dados, e os respectivos objetivos. Os quatro estudos resultaram em três artigos que compõem parte das entregas da tese, o *status* das submissões dos três artigos também foi apresentado.

3.1 FASES DA PESQUISA

A pesquisa foi estruturada em seis fases (Figura 3). Cabe destacar que as fases da pesquisa são interdependentes, pois as atividades aplicadas em uma fase servem de insumo para a fase seguinte. As fases aqui apresentadas possuem, em um primeiro momento, o propósito exploratório e, conforme a pesquisa foi avançando, atividades de validação foram sendo necessárias.

A primeira fase foi composta por uma Revisão Sistemática de Literatura (RSL) motivada pela necessidade de realizar um estudo exploratório sob o contexto da estrutura de decisão em Big Data na perspectiva das pesquisas publicadas em periódicos. A segunda fase foi composta por entrevistas com profissionais com experiência em projetos de Big Data motivada pela necessidade de compreender a percepção destes profissionais sobre o tema, assim como, identificar oportunidades de pesquisa.

A terceira fase foi composta por uma pesquisa de patentes motivada pela necessidade de explorar os registros de patentes de projetos de Big Data, visto que um dos *outputs* da fase 2 da pesquisa foi justamente desenvolver um modelo que orientasse a estruturação de um projeto de Big Data. Neste sentido, a pesquisa demandou a investigação de registros de propriedade intelectual, com o propósito de compreender o estado atual da produção técnica sobre o assunto.

Fase da pesquisa	Estudo	Motivação da fase	Artigo resultante	Método de pesquisa	Técnica de coleta dos dados	Técnica de análise dos dados	Objetivo dos estudos	Objetivo geral
1 ^a	1	Entender a discussão acadêmica sobre a estrutura de decisão com Big Data	Artigo 1 (Apêndice A)	Revisão Sistemática de Literatura	Bibliográfica	Processo recursivo com base na leitura crítica reflexiva e classificação do conteúdo de acordo com a aderência aos temas da pesquisa	Analizar como os Sistemas de Suporte à Decisão gerenciam Big Data para obter valor	Propor um modelo gerencial para estruturação de projetos de Big Data
2 ^a	2	Compreender a percepção de projetos de Big Data sob a ótica de praticantes	Artigo 3 (Apêndice C)	Entrevistas individuais	Entrevistas com questionário semi-estruturado	Processo recursivo orientado por três ciclos de codificação: aberta, axial e seletiva	Explorar a estrutura e o contexto de um projeto de Big Data	
3 ^a	3	Explorar os registros de patentes de projetos de Big Data	Artigo 2 (Apêndice B)	Tecnometria	Pesquisa em base de patentes	Análise léxica textual	Realizar uma análise exploratória de um banco de dados de patentes para coletar informações sobre as características das patentes registradas relacionadas a projetos de Big Data	
4 ^a	-	Desenvolver o modelo inicial com base nos resultados dos três primeiros estudos	-	-	-	-	-	
5 ^a	4	Validar o modelo com profissionais que atuam com projetos de Big Data	Artigo 3 (Apêndice C)	Entrevistas com grupos focais	Entrevista com questionário semi-estruturado	Processo de análise de consolidação das falas a fim de encontrar padrões de respostas	Validar o modelo para estruturação de um projeto de Big Data	
6 ^a	-	Reivindicar o registro de propriedade intelectual do modelo para estruturação de projetos de Big Data	-	-	-	-	-	

Figura 3 - Fases da pesquisa

A quarta fase foi o desenvolvimento de um modelo conceitual influenciado pelas três fases anteriores e motivado pela necessidade de desenvolver um modelo que orientasse a estruturação de um projeto de Big Data para praticantes. A quinta fase da pesquisa compreendeu a validação do modelo com grupos focais compostos por profissionais que atuam com projetos de Big Data. Por fim, a sexta fase da pesquisa foi o depósito do modelo de utilidade solicitando o direito de propriedade intelectual do modelo de estruturação de um projeto de Big Data.

As fases supracitadas, além de contribuirem para a construção do modelo conceitual e modelo de utilidade depositado, serviram de insumo para a produção acadêmica originada dessa pesquisa. A produção acadêmica é descrita em mais detalhes na próxima seção.

3.2 ENTREGAS DA TESE

A presente tese foi composta por três artigos e tem sua estrutura apresentada na Figura 4. Os estudos foram conduzidos de modo a terem distinção entre si, justificando o esforço em conduzir estudos com diferentes objetivos, e que sejam interdependentes, de modo a se complementarem para a obtenção de um objetivo final comum.

Produto	Título	Tipo de pesquisa	Status	Apêndice
Artigo 1	Big Data como um gerador de valor em Sistemas de Suporte à Decisão: uma revisão de literatura	Teórico	Artigo publicado na REGE - Revista de Gestão	A
Artigo 2	Uma análise de patentes sobre projetos de Big Data	Teórico	Artigo submetido em fase final do processo editorial	B
Artigo 3	Construindo um modelo para estruturar um projeto de Big Data	Empírico	Artigo não submetido	C
Patente de Modelo de Utilidade	Modelo gerencial para estruturação de projetos de Big Data	Não se aplica	Reivindicação solicitado ao INPI	Não se aplica

Figura 4 – Produtos provenientes da tese

O primeiro estudo resultou no primeiro artigo (APÊNDICE A), o terceiro estudo resultou no segundo artigo da tese (APÊNDICE B), e o segundo e o quarto estudos compuseram o terceiro artigo (APÊNDICE C). A última entrega da tese foi a reinvindicação de propriedade intelectual do modelo de utilidade. Na próxima seção são descritos os procedimentos metodológicos dos quatro estudos.

3.3 PROCEDIMENTO METODOLÓGICO DO ESTUDO 1 – ESTRUTURA DE DECISÃO COM BIG DATA

A necessidade de entender a discussão acadêmica sobre a estrutura de decisão com Big Data foi a motivação para a realização do primeiro estudo, e uma RSL foi apropriada, pois possibilitou ter uma percepção de um conjunto de artigos acadêmicos relacionados pelos temas tratados. O objetivo do primeiro estudo foi analisar como os Sistemas de Suporte à Decisão gerenciam Big Data para obter valor. Este estudo adotou as seguintes questões de pesquisa: Sistemas de Suporte à Decisão gerenciam Big Data para obter valor por meio de quais técnicas e tecnologias? Sistemas de Suporte à Decisão têm sido aplicados para resolver quais tipos de problemas? Para atingir o objetivo e responder as questões de pesquisa, foi conduzida uma RSL com buscas nas duas principais bases de dados acadêmicos, Scopus e Web of Science, em agosto de 2019.

A primeira etapa da RSL consistiu em uma busca no dia 21/08/2019 nas bases de dados citadas com a string: “Big Data” AND (“*Decision Theory*” or “*decision support system**” or “*decision-support system**”). A utilização do símbolo “*” permite encontrar plurais, gerúndios ou substantivos. A segunda etapa da RSL teve como critério o filtro por áreas. A base Scopus ficou restrita às seguintes áreas: *Business, Management and Accounting; Decision Sciences; Social Sciences*. Já a base Web of Science ficou restrita às seguintes áreas: *Business; Management; Social Science Interdisciplinary*.

A terceira etapa da RSL consistiu no filtro nas bases de dados por artigos, excluindo dessa forma livros, revisões, etc. A quarta etapa da RSL consistiu na verificação da disponibilidade dos artigos para *download*. A quinta etapa da RSL se deu por meio da leitura dos artigos e verificação do alinhamento com o objetivo de pesquisa. Artigos que não tratavam de DSS e Big Data foram excluídos da base de análise. A sexta etapa da RSL consistiu na verificação da duplicidade de artigos e exclusão. Por fim, a sétima etapa da RSL consistiu na

consolidação das duas bases para a análise em profundidade dos artigos. O fluxo de exclusão dos artigos é apresentado na Tabela 2.

Tabela 2 - Fluxo do filtro da RSL

Etapa	Critérios do filtro	Scopus	Web of Science
1	Pesquisa com <i>string</i> de busca	1.087 documentos	340 documentos
2	Filtro de áreas	257 documentos	25 documentos
3	Filtro por artigos	98 artigos	15 artigos
4	Exclusão de artigos indisponíveis	88 artigos	13 artigos
5	Exclusão de artigos impróprios	71 artigos	10 artigos
6	Artigos que constaram nas duas bases		9 artigos
7	Total de artigos analisados na revisão		72 artigos

A busca inicial totalizou 1.427 documentos contidos nas duas bases de dados e, ao final das sete etapas, 5% da amostra, 72 artigos, foram considerados adequados para leitura e análise na íntegra. Os artigos foram lidos na íntegra e classificados em planilha de acordo com as relações identificadas ao longo da análise e que atendessem às questões de pesquisa. O primeiro registro refere-se ao ano de 2012, e até 2018 há um crescimento no número de publicações. O ano de 2019 teve oito publicações, mas devido à data da busca nas bases de dados, não foi possível contabilizar o número total de publicações no ano (Tabela 3).

Tabela 3 - Quantidade de artigos de acordo com os anos

Ano	2012	2013	2014	2015	2016	2017	2018	2019
Quantidade de artigos	1	1	3	8	14	14	23	8

No que se refere ao processo de análise, o conteúdo foi categorizado em uma perspectiva qualitativa (Petticrew & Roberts, 2008). Foi aplicado um processo recursivo com base na leitura crítica reflexiva e classificação do conteúdo de acordo com sua aderência aos temas propostos nesta pesquisa. O processo adotado na RSL permite sua replicabilidade, conforme apontado por Tranfield, Denyer e Smart (2003).

Por meio desse estudo foi possível entender a discussão acadêmica sobre a estrutura de decisão com Big Data, e como um próximo passo, foi necessário avançar na pesquisa mas com uma perspectiva empírica, buscando um entendimento de praticantes e podendo então confrontar com os achados dos artigos da RSL.

3.4 PROCEDIMENTO METODOLÓGICO DO ESTUDO 2 – INVESTIGAÇÃO SOBRE PROJETOS DE BIG DATA

A necessidade de compreender a percepção de projetos de Big Data sob a ótica de praticantes foi a motivação para o segundo estudo, e para tanto, profissionais que atuam com projetos de Big Data foram entrevistados de modo a ser possível um entendimento real de problemas vivenciados por esses profissionais.

O objetivo do segundo estudo foi explorar a estrutura e o contexto de um projeto de Big Data. Foi um estudo empírico com a seguinte questão de pesquisa: Qual a estrutura e o contexto de um projeto de Big Data? Para atingir o objetivo proposto e responder a questão de pesquisa foram realizadas entrevistas em profundidade com 12 profissionais com experiência em projetos de Big Data. Ao todo, foram 9,52 horas de entrevistas transcritas e analisadas pela técnica de análise *Grounded Theory* e técnica de comparação constante com uma perspectiva *Data-driven* e *Theory Driven* (Charmaz, 2006; Silva, Godoi, & Melo, 2010). O processo de análise foi recursivo orientado por três ciclos de codificação: aberta, axial e seletiva.

Foi adotado o método qualitativo para compreender e explorar a temática envolvida em projetos de Big Data. Os procedimentos metodológicos conduzidos neste estudo por meio da descrição da amostra, coleta dos dados, análise dos dados e categorias de análise, proporcionaram a transparência e possibilidade de replicação dos processos utilizados, o que contribui para legitimar os achados desta pesquisa (Creswell, 2017).

A amostra foi composta por 12 profissionais com média de 7,3 anos de experiência com projetos de Big Data e que foram convidados para participar da entrevista e compartilhar conhecimentos na área adquiridos ao longo da carreira. Para a análise da duração das entrevistas considerou-se a contagem do tempo a partir do início da primeira pergunta e finalizando ao término da última resposta, com isso eliminou-se o tempo de apresentação inicial e agradecimentos ao término das entrevistas. As 12 entrevistas totalizaram 9,52 horas (em média 47,5 minutos por entrevista) que foram transcritas individualmente sem auxílio de *software* de transcrição, e que totalizaram 58 horas de transcrição em documento de texto (em média 290 minutos para transcrever cada entrevista) (Tabela 4).

A amostra contemplou profissionais com formação nas áreas de Administração, Ciência da Computação, Economia, Engenharia, Matemática e Tecnologia da Informação, sendo quatro doutores, quatro mestres e quatro especialistas. As entrevistas foram conduzidas por conversas

dirigidas, como explicitado a seguir, a fim de esgotar as discussões a respeito dos itens aqui levantados.

Para a coleta dos dados as entrevistas foram agendadas conforme a disponibilidade dos entrevistados. As estrevistas ocorreram entre o período de 16/07/2020 a 15/09/2020, datas da primeira e última entrevista respectivamente. As entrevistas foram conduzidas com o auxílio de ferramentas para comunicação digital, pois aconteceram durante o período de pandemia da COVID-19 e, por isso, não aconteceram de forma presencial. No início de cada entrevista foi solicitada a permissão para a entrevista ser gravada e com a autorização iniciava-se o protocolo com a apresentação do entrevistado, objetivo da pesquisa e informando que o entrevistado poderia pedir para que a entrevista fosse interrompida em qualquer momento, assim como, após a entrevista o entrevistado poderia ter acesso à gravação para revisão.

Tabela 4 - Descrição dos entrevistados

ID	Área de formação	Maior titulação	Duração da entrevista (min)	Perfil
E1	Tecnologia da Informação	Doutor	60	Técnico
E2	Administração	Doutor	85	Técnico
E3	Tecnologia da Informação	Especialista	35	Técnico
E4	Ciências da Computação	Especialista	47	Gestão
E5	Engenharia/Administração	Doutor	25	Gestão
E6	Matemática	Doutor	36	Técnico
E7	Tecnologia da Informação	Especialista	47	Técnico
E8	Engenharia	Mestre	40	Gestão
E9	Economia	Mestre	35	Gestão
E10	Ciências da Computação	Especialista	41	Gestão
E11	Administração	Mestre	68	Gestão
E12	Administração	Mestre	52	Técnico

O questionário elaborado previamente à realização das entrevistas se baseou em literatura sobre o tema, objetivando buscar entendimento sobre as variáveis relacionadas ao foco deste estudo. A primeira questão, que não estava vinculada a uma categoria de análise, buscava explorar a experiência profissional do entrevistado na área de Big Data, portanto, pedia-se, inicialmente, para o entrevistado comentar sua experiência na área. O processo de entrevista seguiu as prescrições de uma entrevista em profundidade (Boyce & Neale, 2006; Dilley, 2000), esse tipo de entrevista se baseia em uma conversa dirigida orientada por um conjunto de itens (Figura 5).

Com o objetivo de validar e calibrar o instrumento de coleta dos dados, para as três primeiras entrevistas foi identificada a aderência das respostas com a intenção de pesquisa, assim como, foram observadas oportunidades de novas questões serem inseridas no protocolo da pesquisa. Com isso, foram construídas perguntas orientadoras a partir dos itens constituídos conforme apresentado na Figura 5, sendo que novas perguntas emergiram durante a conversa para que se compreendesse, com mais profundidade, o fenômeno pesquisado a partir da perspectiva dos entrevistados.

Categorias	Descrição	Questão	Fonte
Definição de um projeto de Big Data	Buscou-se a definição de projeto de Big Data de acordo com os entrevistados.	Como o(a) senhor(a) define um projeto de Big Data?	Saltz, 2015; Saltz, Shamshurin, e Connors (2017); Kappelman, McKeeman, e Zhang (2007)
Fases de um projeto de Big Data	Propôs-se a identificar as fases que contemplam os projetos de Big Data de acordo com a percepção dos entrevistados.	De acordo com sua experiência, quais fases (ou etapas) contemplam um projeto de Big Data?	Wu, Zhu, Wu, e Ding (2014); Mousanif, Sabah, Douiji e Sayad (2014), e Mousanif, Sabah, Douiji e Sayad (2016).
Fatores críticos em um projeto de Big Data	Propôs-se a identificar os fatores críticos em um projeto de Big Data.	De acordo com sua experiência, quais seriam os fatores críticos em um projeto de Big Data?	LaValle, Lesser, Shockley, Hopkins, e Kruschwitz (2011); Gao, Koronios e Selle (2015); Saltz e Shamshurin (2016).
Problemas de tomada de decisão em um projeto de Big Data	Propôs-se a identificar os problemas relacionados à tomada de decisão em um projeto de Big Data.	Durante o ciclo de vida de um projeto de Big Data, quais problemas relacionados à tomada de decisão o(a) senhor(a) destacaria?	Dutta e Bose (2015); Chiheb, Boumahdi, e Bouarfa (2019).
Metodologia/Guia/Roteiro de um projeto de Big Data	Propôs-se a verificar se há algum padrão de execução ou boas práticas utilizadas para a execução de um projeto de Big Data.	Há alguma metodologia/guia/roteiro que o(a) senhor(a) utiliza nos projetos de Big Data que participa? Se não, tem conhecimento de algum(a)?	Lara, De Sojo, Aljawarneh, Schumaker, e Al-Shargabi (2020); Franková, Drahošová e Balco (2016).

Figura 5 - Categorias iniciais de análise

As questões incluídas foram: (i) Como nasce um projeto de Big Data? e (ii) Sobre habilidades conhecidas como *Hard skills* (técnicas) e *Soft skills* (comportamentais), há alguma que se sobressai em projetos de Big Data? Nota-se que as entrevistas em profundidade são compreendidas como uma conversa dirigida em que o entrevistador assume o papel de facilitador das explicações sobre o que se quer compreender (Boyce & Neale, 2006). A

entrevista teve como foco fazer o entrevistado falar o máximo possível sobre os tópicos de pesquisa, para que isso ocorresse foram utilizadas perguntas de aprofundamento: “como...”, “por quê?...”, “qual sua opinião?...”, “fale mais sobre...”.

Com base nessas informações sobre o processo de entrevistas, justifica-se a escolha pela pesquisa qualitativa, pois ela possibilitou identificar no entrevistado questões com maior nível de profundidade. A forma de condução exigiu que o entrevistador estivesse mais preparado, ou seja, munido de informações para conduzir de forma aberta a entrevista, criando um ambiente promissor e, assim, obter a maior quantidade possível de informações (Turner, 2010). Dilley (2000) descreve que na pesquisa em profundidade, deve haver alguns elementos-chave para que a entrevista seja robusta, são eles: *background* de informações, análise da entrevista e criação de um protocolo. Assim, o pesquisador deverá se munir de informações sobre o contexto cultural do entrevistado e outras informações que são relevantes para o início da entrevista.

A busca de padrões de comportamento que permitissem generalizações analíticas, ou mesmo a construção de teorias, foi feita por meio de um processo indutivo a partir do desenvolvimento de categorias ou temas apresentados pelos entrevistados (Creswell, 2017). O processo de categorização das falas dos entrevistados também seguiu um processo abdutivo, assim, seguindo a recomendação de Charmaz (2006) de buscar significados nas evidências. Assim, as categorias foram constituídas a partir das entrevistas realizadas, e construídas pela perspectiva dos dados (*Data driven*) e na perspectiva da teoria (*Theory Driven*).

Cabe destacar que a fase de análise e discussão dos dados seguiu, principalmente, uma perspectiva *data-driven*, sendo que este tipo de perspectiva permite partir dos dados das entrevistas para a construção de categorias de análise - códigos (Charmaz, 2006; Saldaña, 2012). Apesar disso, a construção do protocolo seguiu orientado por levantando prévio de literatura, entre outras pesquisas. Assim, o processo de análise foi realizado com base na codificação dos dados, que Creswell (2017) descreveu como um processo de organização do material e atribuição de rótulos das categorias destacadas com um termo (código), com base na linguagem real do participante. Para suporte na análise dos dados foi utilizado o *software* ATLAS.ti (versão 7.5.4) e seguiu-se com o processo de análise conforme orientações de Charmaz (2006), processo conhecido como técnica de *Grounded Theory* seguindo ciclos de codificação cujo objetivo é de obter um significado a partir de um corpus de análise.

Primeiramente, foi realizada a codificação aberta, que visa reconhecer incidentes, que são representados por passagens, palavras, imagens, entre outros, e que auxiliam na resposta ao problema de pesquisa. Esses códigos de primeira ordem foram agrupados para estabelecer relações por meio da codificação axial. Com base na estrutura geral construída a partir dos

incidentes e relacionando-os hierarquicamente com outros códigos de segunda ordem, categorias mais abstratas foram obtidas seguindo um terceiro ciclo de análise denominado de seletivo (Charmaz, 2006; Saldaña, 2012).

A análise também foi realizada por meio da técnica de comparação constante, na qual as atividades de codificação e análise são realizadas de forma concomitante e circular, o que permite idas e vindas ao corpus de pesquisa para validação das categorias de análise (Charmaz, 2006; Saldaña, 2012). Esse processo segue um fluxo em direção ao alcance dos objetivos da pesquisa e não está vinculado ao processo de forma linear. Assim, à medida que a amostragem teórica é aumentada (número de incidentes encontrados), são construídos os principais elementos para embasar categorias mais abstratas e suas respectivas propriedades (Silva *et al.* 2010). Estudos qualitativos com alto grau de excelência vão além da descrição e identificação e chegam a conexões temáticas complexas (Creswell, 2017). Vale ressaltar que as categorias iniciais de análise partem de uma perspectiva teórica. Essas categorias representaram os estilos de gestão e foram construídas a partir da busca e análise do referencial teórico.

A etapa de tratamento dos dados coletados permitiu ganhar confiança na análise e expandir sua percepção quanto ao dito e não dito nas entrevistas, enquanto que o protocolo traça o caminho que foi percorrido para inquirir o entrevistado. Também foi contemplada uma perspectiva de dados orientada para a compreensão de novos elementos do fenômeno pesquisado (Charmaz, 2006).

Ao final de todas as entrevistas, pediu-se, ainda, aos entrevistados para citarem alguma oportunidade de pesquisa ou o quê poderia ser desenvolvido para auxiliar a gestão de um projeto de Big Data. Por meio deste estudo foi possível compreender a percepção de projetos de Big Data sob a ótica de praticantes e as oportunidades de pesquisa identificadas condicionaram ao entendimento de que seria apropriado uma investigação de registros de propriedade intelectual, visto que um dos propósitos desta tese é de gerar um modelo de utilidade e, como forma de se assegurar o ineditismo da pesquisa, uma pesquisa em banco de dados de patentes foi adequado.

3.5 PROCEDIMENTO METODOLÓGICO DO ESTUDO 3 – PATENTES DE PROJETOS DE BIG DATA

A necessidade de explorar os registros de patentes relacionadas aos projetos de Big Data e, com isso, assegurar o ineditismo do modelo proposto foi a motivação do terceiro estudo. O

estudo ainda condicionou mais uma fonte de referência para a construção do modelo, visto que as patentes possuem informações valiosas quando se propõe esse tipo de desenvolvimento.

O objetivo do terceiro estudo foi realizar uma análise exploratória de um banco de dados de patentes para coletar informações sobre as características das patentes registradas relacionadas a projetos de Big Data. Foi um estudo teórico que possibilitou responder a seguinte questão de pesquisa: Quais tipos de problemas foram resolvidos por patentes relacionadas a projetos de Big Data? Para atingir o objetivo e responder a questão de pesquisa foi realizada uma pesquisa na base de dados Espacenet no dia 10 de janeiro de 2021 com a seguinte *string*: “*Big data project**”.

De acordo com Clarke (2018), existem diferentes tipos de busca de patentes dependendo das perguntas que estão sendo feitas: estado da arte, arte anterior/novidade/patenteabilidade, liberdade de operar, oposição e buscas de (in)validade. Com isso, a pesquisa foi classificada como estado da arte, pois foram coletados documentos relevantes publicados em uma determinada área técnica e buscado estabelecer conhecimentos e informações descritivas.

Como resultado, a busca resultou em 109 registros e, após uma análise estruturada dos documentos, seis patentes foram eliminadas da amostra por não possuírem escopo para o projeto de Big Data, mas acabaram sendo incluídas no resultado por citarem o termo da pesquisa em uma forma descontextualizada, além disso, foram identificados dois casos com duplicidade (Tabela 5).

Para a análise descritiva, foram eliminadas as seis patentes consideradas fora do escopo da pesquisa, e para a análise léxico-textual, além de descontar as seis patentes fora do escopo, foi considerada apenas uma das patentes repetidas em ambos os casos para não influenciar a qualidade dos resultados. Para a análise dos dados foi utilizado o *software* livre e de código aberto Iramuteq versão 0.7 que possibilita a análise textual de um corpus. O corpus consistiu nos resumos das patentes disponíveis nos documentos e, após o processamento do corpus, foram identificados 230 segmentos de texto e 7890 ocorrências.

Foram considerados apenas adjetivos, verbos e nomes comuns como chaves de análise. A Tabela 6 apresenta o resumo da análise do corpus textual, em que o corpus textual de 101 das patentes analisadas gerou 230 segmentos de texto e 7.890 ocorrências. Foram identificados 567 hápix, que são palavras identificadas apenas uma vez em todo o corpus textual.

Tabela 5 - Patentes eliminadas e repetidas da amostra

	Número da patente	Nome da patente
Patentes eliminadas	CN108446206A	<i>AR technology-based application availability display system</i>
	US2020324135A1	<i>Device and method for treating eight cancers with ultraviolet radiation</i>
	CN107177625A	<i>Artificial carrier system with site-specific mutagenesis and site-specific mutagenesis method</i>
	CN107671020A	<i>Automatic cleaning brush of seabed observation instrument equipment and work principle of automatic cleaning brush</i>
	CN104484582A	<i>Method and system for automatically analyzing bio information items through modular selection</i>
Patentes repetidas	WO2019100012A1	<i>Data collection and analytics based on detection of biological cells or biological substances</i>
	US9384264B1	<i>Analytic systems, methods, and computer-readable media for structured, semi-structured, and unstructured documents</i>
	EP3179387A1	
	WO2020014399A1	<i>Decentralized Cybersecure Privacy Network For Cloud Communication, Computing And Global e-Commerce</i>
	US2019386969A1	

Por meio do Método de Classificação Hierárquica Descendente (MCHD) obteve-se classes de segmentos de texto que, ao mesmo tempo, apresentaram vocabulário semelhante entre si, e vocabulário diferente dos segmentos de texto das demais classes (Camargo, 2005).

Tabela 6 - Resumo da análise do corpus

Resumo da análise do corpus	Total
Número de textos	101
Número de segmentos de texto	230
Ocorrências	7.890
Média de ocorrência de texto	78.90
Número de formas	1.269
Número de hápax	567*

*44.68% das formas, 7.19% das ocorrências.

Também foi realizada a Análise Fatorial de Correspondência (AFC), recurso do Iramuteq que apresenta uma representação gráfica no plano cartesiano e que ajuda a visualizar a proximidade e distâncias das palavras a partir do cruzamento do vocabulário do corpus textual. Os procedimentos realizados nesta análise incluem o cálculo de frequências, valores de correlação do qui-quadrado para cada palavra do corpus e a execução do AFC em uma tabela de contingência que cruza as formas ativas e as variáveis (Salviati, 2017).

Por meio deste estudo foi possível explorar os registros de patentes relacionadas aos projetos de Big Data, garantindo o ineditismo do modelo construído. Com a realização dos três

estudos, o modelo foi desenvolvido baseado no consequente avanço alcançado. Após o desenvolvimento do modelo, a validação deste foi necessária e, por isso, um quarto estudo foi necessário.

3.6 PROCEDIMENTO METODOLÓGICO DO ESTUDO 4 – VALIDAÇÃO DO MODELO

A necessidade de validar o modelo com profissionais que atuam com projetos de Big Data foi a motivação do quarto estudo. Para tanto, entrevistas com grupos focais foram realizadas de modo a buscar a legitimação do modelo com profissionais com conhecimento empírico em projetos de Big Data e obter a credibilidade necessária ao modelo.

O objetivo do quarto estudo foi validar o modelo para a estruturação de um projeto de Big Data. Foi um estudo empírico para responder a seguinte questão de pesquisa: Como um modelo gerencial pode auxiliar na estruturação de projetos de Big Data? Para atingir o objetivo proposto e responder a questão de pesquisa, foram conduzidas três entrevistas de grupos focais com profissionais que atuam com projetos de Big Data.

O grupo focal é um tipo de entrevista em profundidade realizada em grupo cujas reuniões têm características quanto à proposta, ao tamanho, à composição e aos procedimentos de condução (Oliveira & Freitas, 1998). A utilização deste procedimento se deu pela necessidade de buscar discussões entre praticantes para legitimar o *framework* elaborado, assim como, retrabalhar etapas em caso de necessidade.

A pesquisa empírica foi conduzida em formato *online*, o que possibilitou a realização de entrevistas com profissionais especializados que seriam dificilmente reunidos fisicamente em um único local. Além disso, esta estratégia foi adotada por conta do período de pandemia. Pesquisas indicam que os participantes preferem a conveniência de grupos focais online devido à flexibilidade na programação e a capacidade de participar de casa ou do escritório (Zwaanswijk & van Dulmen, 2014), e a distinção entre o mundo “real” e o mundo “virtual” está se tornando cada vez mais difícil (Stewart & Shamdasani, 2016).

A amostra foi composta por nove profissionais com experiência em projetos de Big Data (Tabela 7), um entrevistador e um facilitador. A seção 01 foi composta por dois professores de programas *strictu sensu* e um aluno de mestrado como entrevistados. A seção 02 foi composta por uma equipe de quatro profissionais que fazem parte do departamento de Ciência de Dados

de uma empresa privada do ramo de aluguel de carros. A seção 03 foi composta por dois profissionais de um banco privado.

Cada seção de grupo focal foi iniciada com uma breve apresentação do entrevistador e do facilitador e, então, o modelo era apresentado integralmente. Como procedimento de coleta dos dados foi utilizado um questionário não estruturado para dar liberdade aos entrevistados a comentarem e discutirem os pontos mais relevantes do modelo e, com isso, os entrevistados puderam apresentar seus respectivos pontos de vista. Todas as sessões foram gravadas com consentimento dos entrevistados.

Tabela 7 – Entrevistas com grupos focais

Seção	Data	Área de formação do entrevistado	Maior titulação do entrevistado	Duração da seção (min)
01	27/05/2021	Tecnologia da Informação	Doutor	
		Administração	Doutor	81
		Tecnologia da Informação	Especialista	
02	08/06/2021	Física	Mestrado	
		Tecnologia da Informação	Especialista	
		Tecnologia da Informação	Mestrado	65
03	12/07/2021	Ciência da Computação	Especialista	
		Sistemas da Informação	Especialista	
		Matemática	Especialista	57

Como procedimento de análise dos dados, após cada grupo focal a gravação era assistida com o propósito de revisar as sugestões e questionamentos realizados pelos entrevistados. Todos os apontamentos foram listados e discutidos entre o entrevistador e o facilitador, de modo a chegar em um consenso sobre a aceitação da sugestão ou não. Então, as sugestões acatadas eram incorporadas ao modelo e o próximo grupo focal era realizado com o modelo revisado.

Por meio deste estudo, foi possível validar o modelo com profissionais que atuam com projetos de Big Data e com a conclusão do quarto estudo foi possível reivindicar o registro de propriedade intelectual do modelo para estruturação de projetos de Big Data.

4 RESULTADOS DOS ESTUDOS

Nesta seção os resultados dos quatro estudos realizados são apresentados. O primeiro uma revisão sistemática, o segundo, entrevistas individuais, o terceiro, uma pesquisa em uma base de dados de patentes e o quarto, entrevistas com grupos focais.

4.1 RESULTADOS DO ESTUDO 1 – ESTRUTURA DE DECISÃO COM BIG DATA

No primeiro estudo (APÊNDICE A), cujo objetivo foi analisar como os Sistemas de Suporte à Decisão gerenciam Big Data para obter valor, foi realizada uma Revisão Sistemática de Literatura com buscas nas duas principais bases de dados acadêmicos, *Scopus* e *Web of Science*, em agosto de 2019. Ao todo foram 72 artigos selecionados, destes 63 eram artigos empíricos e nove artigos teóricos. Foram identificadas 8 técnicas e 4 tecnologias que são utilizadas por meio de Sistemas de Suporte à Decisão para obter valor de Big Data (Tabela 8).

Tabela 8 - Sistemas de Suporte à Decisão gerenciam Big Data para obter valor por meio de quais técnicas e tecnologias?

	Técnicas	Artigos
Aprendizado Supervisionado	Classificação	(Hayashi, 2016) (Rasiulis, Ustinovichius, Vilutienė, & Popov, 2016) (Rönnqvist, Svenson, Flisberg, & Jönsson, 2017) (Li, Yang, Jin, & Guo, 2017) (Ghaddar & Naoum-Sawaya, 2018) (Nimmagadda, Reiners, & Wood, 2018) (Sathiaraj, Punkasem, Wang, & Seedah, 2018) (Sivamani, Choi, & Cho, 2018) (Tashkandi, Wiese, & Wiese, 2018) (Tian, Zhang, & Zhang, 2018) (Gillingham, 2019a) (Suh, 2019)
Aprendizado Supervisionado	Rregressão	(Ilie-Zudor, Ekárt, Kemeny, Buckingham, Welch, & Monostori, 2015) (Ludwig, Feuerriegel, & Neumann, 2015) (Feuerriegel, 2016) (Chen, Achtari, Majkut, & Sheu, 2017) (Hsu, Lim, & Yang, 2017) (Nguyen <i>et al.</i> , 2017) (Jamshidi <i>et al.</i> , 2018) (Khan <i>et al.</i> , 2018) (Stein, Meller, & Flath, 2018)
Aprendizado não supervisionado	Cluster	(Hurter, Conversy, Gianazza, & Telea, 2014) (Miah, Vu, Gammack, & McGrath, 2017) (Gao, Zhang, Lu, Wu, & Du, 2018) (Papagiannidis, See-To, Assimakopoulos, & Yang, 2018) (Giglio, Bertacchini, Bilotta, & Pantano, 2019)
Aprendizado não supervisionado	Associação	(Capobianco & Liò, 2015) (Chan & Chong, 2017)
Aprendizado semi supervisionado	Classificação	(Li Li, & Zhu, 2016)
Aprendizado por reforço	Controle	(Hao, Yang, & Shi, 2018) (Wang, Geng, & Gao, 2018)

Mineração de dados	(Burattin, Cimitile, Maggi, & Sperduti, 2015) (Sasson, Ravid, & Pliskin, 2015) (Drosio & Stanek, 2016) (Goh, Tao, Zhang, & Yong, 2016) (Al Chahadah, El Refae, & Qasim, 2018) (Ahmed, Aziz, Tezel, & Riaz, 2018) (Lee, Aydin, Choi, Lekhavat, & Irani, 2018)
<i>Big Data Analytics</i>	(Ashrafi, Kuilboer, Joshi, Ran, & Pande, 2014) (Groves, Collins, Gini, & Ketter, 2014) (Vera-Baquero, Colomo-Palacios, Molloy, & Elbattah, 2015) (Giannakis & Louis, 2016) (Osuszek, Stanek, & Twardowski, 2016) (Power, 2016) (Shrestha, Cater-Steel, & Toleman, 2016) (Long, 2017) (Moore, 2017) (Brinch, Stentoft, Jensen, & Rajkumar, 2018) (Fredriksson, 2018) (Wang, Kung, & Byrd, 2018) (Ghasemaghaei & Calic, 2019) (Mello & Martins, 2019) (Power, Cyphert, & Roth, 2019) (Schneider & Seelmeyer, 2019)

Tecnologias	Artigos
<i>Cloud computing</i>	(Khansa, Forcade, Nambari, Parasuraman, & Cox, 2012) (Demirkan & Delen, 2013) (Neaga, Liu, Xu, Chen, & Hao, 2015) (Boutkhoum, Hanine, Agouti, & Tikniouine, 2016) (Boutkhoum <i>et al.</i> , 2017) (Schnase <i>et al.</i> , 2017)
<i>Business Intelligence</i>	(Pape, 2016) (Jin & Kim, 2018)
<i>Data Warehouse</i>	(Almeida, Bernardino, & Furtado, 2015) (Jukic, Jukic, Sharma, Nestorov, & Arnold, 2017) (Janković, Mladenović, Mladenović, Vesović, & Glavić, 2018)
Sistemas de informação	(Lan, Zhang, Zhong, & Huang, 2016) (Semanjski, Bellens, Gautama, & Witlox, 2016) (Baechle, Agarwal, & Zhu, 2017) (Deal, Pan, Pallathucheril, & Fulton, 2017) (Pettit <i>et al.</i> , 2018) (Aversa, Cabantous, & Haefliger, 2018) (Gillingham, 2019b)

O maior número de trabalhos foi de técnicas relacionadas à utilização de *Big Data Analytics* (BDA), 16 artigos representando 22% da amostra, seguido de aprendizado supervisionado por classificação, 12 artigos representando 17% da amostra. Em relação às tecnologias, sistemas de informação tiveram a maior frequência, com sete artigos representando 10% da amostra, seguido de *Cloud computing* com seis artigos, que representaram 8% da amostra.

Foram identificados 11 campos de aplicação de Sistemas de Suporte à Decisão para obtenção de valor de Big Data (Tabela 9). Ressalta-se que os artigos direcionados à área de logística e tráfego apresentaram a maior taxa, representando 19%, seguidos pelas áreas de saúde, organização e mercado, com 17% cada.

Cabe destacar também que entre os anos de 2012 e 2013 apenas tecnologias foram identificadas nos artigos e, a partir de 2014, houve registros de artigos utilizando técnicas de análise de Big Data. Foi observado ainda, que o crescimento do uso da tecnologia ocorreu até 2016, com queda nos anos subsequentes. Quanto às técnicas, desde o registro do primeiro artigo

(2014) seu uso tem crescido de forma constante, caracterizando assim a crescente aplicabilidade das técnicas para esse fim.

Tabela 9 - Sistemas de Suporte à Decisão têm sido aplicados para resolver quais tipos de problemas?

Campo	Escopo	Artigos
Logística & Tráfego	Otimização de rotas para economia de custos, desempenho e melhorias.	(Hurter <i>et al.</i> , 2014) (Ilie-Zudor <i>et al.</i> , 2015) (Neaga <i>et al.</i> , 2015) (Lan <i>et al.</i> , 2016) (Chen <i>et al.</i> , 2017) (Hsu <i>et al.</i> , 2017) (Rönnqvist <i>et al.</i> , 2017) (Aversa <i>et al.</i> , 2018) (Hao <i>et al.</i> , 2018) (Jamshidi <i>et al.</i> , 2018) (Jin & Kim, 2018) (Lee <i>et al.</i> , 2018) (Sathiaraj <i>et al.</i> , 2018) (Yongheng Wang <i>et al.</i> , 2018)
Saúde	Previsão de saúde da população, assistência à detecção de doenças e mapeamento de informações para melhorar o atendimento.	(Khansa <i>et al.</i> , 2012) (Ashrafi <i>et al.</i> , 2014) (Burattin <i>et al.</i> , 2015) (Capobianco & Liò, 2015) (Feuerriegel, 2016) (Goh <i>et al.</i> , 2016) (Baechle <i>et al.</i> , 2017) (Nguyen <i>et al.</i> , 2017) (Ghaddar & Naoum-Sawaya, 2018) (Sivamani <i>et al.</i> , 2018) (Tashkandi <i>et al.</i> , 2018) (Yichuan Wang <i>et al.</i> , 2018)
Organizacional & Mercado	Assistência na tomada de decisões de negócios e melhoria do relacionamento social.	(Demirkan & Delen, 2013) (Boutkhoum <i>et al.</i> , 2016) (Osuszek <i>et al.</i> , 2016) (Pape, 2016) (Boutkhoum <i>et al.</i> , 2017) (Moore, 2017) (Ghaddar & Naoum-Sawaya, 2018) (Janković <i>et al.</i> , 2018) (Tian <i>et al.</i> , 2018) (Ghasemaghaei & Calic, 2019) (Mello & Martins, 2019) (Power <i>et al.</i> , 2019)
Indústria & T.I.	Eficiência de processos, custos operacionais e otimização de atividades.	(Almeida <i>et al.</i> , 2015) (Sasson <i>et al.</i> , 2015) (Rasiulis <i>et al.</i> , 2016) (Shrestha <i>et al.</i> , 2016) (Jukic <i>et al.</i> , 2017) (Li <i>et al.</i> , 2017) (Ahmed <i>et al.</i> , 2018) (Nimmagadda <i>et al.</i> , 2018) (Stein <i>et al.</i> , 2018)
Supply Chain	Supporte para tomada de decisão estratégica e controle de indicadores de desempenho.	(Groves <i>et al.</i> , 2014) (Vera-Baquero <i>et al.</i> , 2015) (Giannakis & Louis, 2016) (Long, 2017) (Brinch <i>et al.</i> , 2018) (Papagiannidis <i>et al.</i> , 2018)
Tempo	Análise climática para prever o fluxo de consumo de varejo, tráfego e gerenciamento de crises.	(Ludwig <i>et al.</i> , 2015) (Drosio & Stanek, 2016) (Schnase <i>et al.</i> , 2017) (Lee <i>et al.</i> , 2018) (Sathiaraj <i>et al.</i> , 2018) (Tian <i>et al.</i> , 2018)
Turismo &/ou Mídias sociais	Análise de tendências de fluxo turístico de acordo com o registro de visitas.	(Jiexun Li <i>et al.</i> , 2016) (Miah <i>et al.</i> , 2017) (Nguyen <i>et al.</i> , 2017) (Gao <i>et al.</i> , 2018) (Giglio <i>et al.</i> , 2019)
Financeiro	Supporte à decisão na análise de riscos, tendências contábeis e de índices de ações.	(Hayashi, 2016) (Chan & Chong, 2017) (Al Chahadah <i>et al.</i> , 2018) (Khan <i>et al.</i> , 2018)
Problemas sociais	Assistência na identificação de problemas sociais e mitigação de decisões enganosas com base em banco de dados.	(Gillingham, 2019a) (Gillingham, 2019b) (Suh, 2019) (Schneider & Seelmeyer, 2019)
Smart city	Assistência para sistemas de planejamento e crescimento urbano sustentável.	(Semanjski <i>et al.</i> , 2016) (Deal <i>et al.</i> , 2017) (Pettit <i>et al.</i> , 2018)
Governamental	A centralização da coleta de Big Data pode fornecer condições de controle sobre a população.	(Fredriksson, 2018) (Power, 2016)

Os artigos relacionados ao tema aumentaram no número de publicações ao longo dos anos. A facilidade de acesso aos componentes que permitam a coleta e análise de dados explica a melhor portabilidade, transferência remota e acesso dos dados digitalizados de um contexto para outro, assim como, interconectividade, possibilidade de sintetizar Big Data e encontrar conexões a partir da interação entre inteligência humana e algorítmica (Günther *et al.*, 2017), contribuiu com a melhora na obtenção de valor de Big Data.

A aplicação de Sistemas de Suporte à Decisão para obtenção de valor de Big Data, além do conhecimento em tecnologias de coleta, armazenamento e processamento de dados e técnicas de análise de dados, exige que os profissionais sejam capazes de identificar problemas específicos e complexos, o que requer um conhecimento profundo desses profissionais nessas áreas de aplicação do problema.

Nota-se uma relação inversa na complexidade da forma de análise dos dados quanto à necessidade de tomada de decisão humana baseada na intuição. Por exemplo, Vera-Baquero *et al.* (2015) usando análise descritiva utilizou mecanismos de correlação de eventos para monitorar instâncias e processos para auxiliar os analistas em sua tomada de decisão. Miah *et al.* (2017), por meio de análise algorítmica preditiva, identificou áreas de alta atividade fotográfica para encontrar grupos em locais turísticos populares. Pape (2016) apresentou uma estrutura de análise prescritiva para priorizar itens na análise de negócios e aplicá-los em um ambiente organizacional.

Neste estudo, os artigos tiveram uma predominância na análise preditiva, sugerindo uma posição intermediária da intensidade da complexidade na análise dos dados e da necessidade de tomada de decisão humana. Vale destacar uma tendência de aumento do trabalho com análises prescritivas, dado o crescente interesse, por exemplo, em estudos com carros autônomos e otimização de rotas.

Em geral, os tipos de análise são definidos de acordo com o entendimento do problema, de forma que, para análise de dados de eventos passados, técnicas de Big Data *Analytics* são utilizadas para análises descritivas, e para problemas de inferência e predição, na qual busca identificar possíveis resultados e probabilidades de um evento, o aprendizado de máquina foi usado.

Embora tenham sido apresentados separadamente, as relações entre os tipos de problemas, técnicas e tecnologias são utilizadas como uma forma de resolver problemas complexos, dada a limitação humana para identificar relacionamentos em grandes conjuntos de

dados. Isso requer o auxílio de tecnologias e técnicas, que fornecem mecanismos para a obtenção de *insights*, que auxiliam na tomada de decisões. Pode-se concluir que o momento do *insight*, ao dar condições para que os tomadores de decisão identifiquem a melhor decisão a ser tomada para resolver um problema complexo, é quando se obtém o valor do Big Data, e o Sistema de Suporte à Decisão é um meio para isso.

Análises da literatura de Big Data foram realizadas em algumas áreas, Seles *et al.* (2018), por exemplo, argumentaram que a gestão de Big Data significa uma grande oportunidade em sustentabilidade corporativa, devido ao seu potencial para gerar um melhor entendimento das oportunidades e desafios das mudanças climáticas. Li, Xu, Tang, Wang e Li (2018) argumentaram sobre algumas desvantagens do uso de Big Data no turismo relacionado à qualidade dos dados, custo dos dados e questões de privacidade. Bhardwaj, Wodajo, Spano, Neal e Coustasse (2018) ao investigarem a aplicação de análises de Big Data na área da saúde, identificaram benefícios em várias fases do gerenciamento de doenças crônicas e poderiam ajudar a reduzir a carga de doenças crônicas nos pacientes.

Por se tratar de uma área ainda nova, a pesquisa em Big Data é muito recente e, ao mesmo tempo, ainda existem muitas oportunidades de pesquisas que buscam investigar esse ambiente, seus desafios e avanços obtidos nos últimos anos. Com isso, propôs-se a realizar esta pesquisa abrangente para que fosse possível registrar as áreas em que a aplicação do Big Data tem sido utilizada para solucionar problemas reais e também, os tipos de técnicas e tecnologias utilizadas de forma mais recorrente.

Com isso, algumas oportunidades de pesquisa emergem dos resultados, como, por exemplo, investigações sobre o perfil comportamental dos profissionais envolvidos com o desenvolvimento de Sistemas de Suporte à Decisão, quais capacidades se destacam para determinadas áreas, e quais impactos negativos podem resultar de decisões enviesadas tomadas a partir desses mecanismos. Outro ponto pouco explorado e que merece maior atenção está relacionada às questões éticas envolvidas nesse tipo de contexto, portanto seria oportuno investigar como a ética está sendo discutida no desenvolvimento de Sistemas de Suporte à Decisão envolvendo, por exemplo, controle civil com restrições de liberdade, incentivo ao consumo e automação de julgamentos de condenação baseados em padrões.

4.2 RESULTADOS DO ESTUDO 2 – INVESTIGAÇÃO SOBRE PROJETOS DE BIG DATA

No segundo estudo (APÊNDICE C), cujo objetivo foi explorar a estrutura e o contexto de um projeto de Big Data, foram realizadas entrevistas em profundidade com 12 profissionais com experiência em projetos de Big Data. Ao todo, foram 9,52 horas de entrevistas transcritas e analisadas pela técnica de análise *Grounded Theory* e técnica de comparação constante com uma perspectiva *Data-driven* e *Theory Driven* (Charmaz, 2006; Silva *et al.*, 2010). O processo de análise foi recursivo orientado por três ciclos de codificação: aberta, axial e seletiva.

Um projeto de Big Data, como definido por Becker (2017), é um projeto intensivo de dados com problemas de volume, variedade e velocidade, mas, na análise, destaca-se a ênfase dada na necessidade de haver um problema de negócio a ser resolvido. Com isso, entende-se que uma fase predecessora ao projeto em si, mas que não deve ser negligenciada é a percepção de quais problemas precisam ser solucionados, e então questionar, como sugerido por Mousanif *et al.* (2014), se esses problemas podem ser resolvidos com *softwares* e técnicas para análises convencionais de dados. Frente a isso, é possível inferir que um projeto de Big Data pode ser definido como uma forma de resolução de um problema complexo, por meio de tecnologias de computação distribuída e técnicas de análises de dados, que resulta em uma informação capaz de auxiliar a tomada de decisão.

Para obter o máximo de benefícios do Big Data, profissionais e pesquisadores não devem se concentrar apenas no aspecto de dados e análise de dados ao estudar e desenvolver um projeto de Big Data para apoiar a tomada de decisão, mas também ir além disso, e estudar o aspecto de decisão deste projeto (Chiheb *et al.*, 2019). Foi percebido uma maturidade acerca dos entrevistados por não negligenciarem o aspecto da tomada de decisão, contudo, foi notado que fatores externos, fora da alçada dos gestores, podem impactar a qualidade das decisões, como por exemplo, o viés do nível executivo em desejar algum tipo de resultado, ou ainda influenciando a aquisição de uma tecnologia que não é a mais adequada.

Quando analisadas as fases de um projeto de Big Data (Tabela 10), foi constatado que não há um consenso de etapas ou fases que contemplam o projeto. Para Mousanif *et al.* (2016) o fluxo de um projeto de Big Data é linear iniciando com a elaboração da estratégia global como forma de planejamento, seguindo por etapas de implantação e, por fim, com a etapa de pós-implementação. Ao apresentarem um mapa holístico para implantação de projetos de Big Data, Dutta e Bose (2015) apresentaram um *framework* com as fases de base estratégica, análise de

dados e implementação. Em consonância com esta discussão, Dietrich e Reiner (2017) produziram uma invenção que é um programa de computador para provisionar um sistema de computação de acordo com um ciclo de vida analítico de dados. De acordo com os inventores, o processo atende às necessidades emergentes relacionadas a projetos de Big Data sendo efetiva para obter benefícios operacionais de fontes de Big Data.

Tabela 10 - Fluxo de projetos de Big Data

Mousanif <i>et al.</i> (2016)	Dutta e Bose (2015)	Dietrich e Reiner (2017)
1) Elaboração de estratégia global	1) Problema de negócio	1) Explorar o contexto de negócios e enquadrar a hipótese
2) Coleta dos dados	2) Pesquisa	2) Descobrir e ingerir fontes de dados
3) Processamento dos dados	3) Formação de equipe multifuncional	3) Integrar e preparar dados
4) Análise inteligente dos dados	4) Roteiro do projeto	4) Planejar modelo, métodos e variáveis-chave
5) Representação e visualização	5) Coleta e examinação dos dados	5) Desenvolver teste, conjunto de dados e modelo de construção
6) Extração de <i>insights</i> açãoáveis e oportunos	6) Modelagem e análise dos dados	6) Aplicar o modelo aos cenários de análise e de decisão
7) Avaliação	7) Visualização dos dados 8) Geração de <i>insights</i>	7) Envolver e operacionalizar o modelo
	9) Integração com sistema de TI 10) Treinando pessoas	
Linear	Cíclico	Cíclico

A partir da análise deste estudo, pode-se concluir que um projeto de Big Data tem uma fase inicial relacionada à estratégia, onde se delimita o problema complexo a ser resolvido, seus requisitos e paradigmas de análise. Então, tem-se a fase tecnológica que é onde se estrutura toda a arquitetura de coleta, tratamento e processamento de dados. Em seguida tem-se a fase analítica que é o momento da realização das análises dos dados provenientes da fase anterior e, a partir disso, deve-se avaliar se foi gerado algum tipo de *insight*. Por fim, tem-se uma nova fase estratégica, onde ocorre a aferição do resultado ter atingido o objetivo inicial.

Em uma pesquisa sobre como as equipes trabalham juntas para executar projetos de Big Data, Saltz e Shamshurin (2016) inferem que não há um padrão acordado para a execução desses projetos, mas há um foco crescente de pesquisa nesta área e que uma metodologia de processo aprimorada seria útil. Os autores ainda identificaram 33 fatores críticos para a execução de esforços de Big Data, que são agrupados por seis características identificadas de uma organização de Big Data madura que são: Dados, Governança, Processos, Objetivos, Equipe e Ferramentas. Também em uma investigação de fatores críticos em projetos de Big

Data, Gao *et al.* (2015) apresentaram 27 fatores divididos em seis fases, sendo elas: Negócio, Dados, Análise, Implementação, Medição e Gerais.

Como resultado deste estudo, foram identificados 17 fatores críticos, com maior incidência para *Soft skills*, Tecnologia, *Hard skills* e Planejamento. Destaca-se a percepção que os entrevistados dão às habilidades das pessoas envolvidas com os projetos, também com a aquisição e utilização de tecnologias adequadas e, ainda, com o planejamento adequado para a execução de projetos de Big Data. Esses fatores críticos convergem com problemas de tomada de decisão identificados nas respostas desta pesquisa, e conforme afirmaram Chiheb *et al.* (2019) muitos projetos de Big Data fornecem resultados decepcionantes que não atendem às necessidades dos tomadores de decisão por vários motivos, como por exemplo, negligenciar o estudo do aspecto de tomada de decisão dos projetos.

Ao investigarem qual abordagem deve ser aplicada na gestão de projetos de Big Data, Franková *et al.* (2016) constataram que a abordagem ágil é preferível devido à frequente mudança de requisitos provenientes de novos questionamentos que vão surgindo ao longo do projeto. Neste estudo, percebeu-se a aderência na utilização de métodos ágeis na medida em que os gestores procuram entregar resultados fragmentados com pequenos casos de uso, aceitando falhas e promovendo sucessivas revisões, de modo a seguir uma orientação iterativa.

Como também evidenciado por Franková *et al.* (2016), constatou-se que a escolha da abordagem na gestão de projetos de Big Data é condicionada não apenas pelo tamanho e criticidade do projeto, mas também pela dinâmica do ambiente, pelas habilidades dos responsáveis pelo trabalho do projeto e pela cultura organizacional. Para Jin *et al.* (2015) as condições necessárias para que um projeto de Big Data seja bem-sucedido estão relacionadas a ter requisitos técnicos, sociais e econômicos muito claros, ter estruturas de dados pequenas o suficiente para caracterizar o comportamento a ser analisado e deve-se buscar modelo de gestão de cima para baixo para promover resoluções abrangentes, por meio de uma abordagem integrada. Identificou-se relações em vários aspectos, principalmente, na relação do ambiente em que os projetos de Big Data estão inseridos com a *expertise* da equipe e cultura organizacional.

Com a última pergunta do questionário (O que poderia ser desenvolvido para auxiliar a gestão de um projeto de Big Data?) buscou-se identificar oportunidades de pesquisa pela perspectiva dos entrevistados, de modo a ouvir sugestões do que poderia ser desenvolvido para auxiliar a gestão em projetos de Big Data. Como resultado, teve-se sugestões que serviram como uma agenda de pesquisas futuras, tornando-se algo complementar a este estudo.

Dentro do contexto do problema do *hype* envolvido em um projeto de Big Data, E2 sugere o desenvolvimento de uma ferramenta de planejamento que oriente decisões baseadas em critérios técnicos. A percepção de que um *framework* teria valor foi compartilhada por E3, E5, E7, E8 e E11. E8 destaca, ainda, que o Big Data pode ser aplicado em qualquer empresa ou qualquer negócio, a dificuldade é ter profissionais qualificados que consigam estruturar adequadamente o ciclo de vida para a obtenção de resultados. Ao argumentar que a academia precisa passar por uma transformação empírica, E11 sugere a elaboração de um *framework* ético na área de Big Data. Segundo o entrevistado, há pouca discussão sobre o aspecto ético de projetos de Big Data, quando comparado com a discussão sobre qualidade ou segurança, e poderia até ser desenvolvida uma fase sobre ética como um processo dentro de um *Design Thinking*.

No entendimento de E4, falta uma “metodologia não muito engessada e também não muito rápida”, ou seja, algo pouco burocrático, mas que proporcione o controle dos processos e que seria uma intersecção de *waterfall* e ágil, e que chamou de “uma metodologia ágil documentável”. E12 sugere a elaboração de uma linha mestre de condução de projetos de Big Data, considerando as especificidades de cada projeto, e E8 propôs o desenvolvimento de métricas de gestão associadas à metodologia e, com isso, mensurar eficiência e desempenho ao longo da gestão.

Por fim, E6 sugere a elaboração de um material para profissionais que vão desempenhar a função de gestor de um projeto de Big Data, mas que não são da área de gestão de projetos. Com o interesse cada vez maior na área de Big Data, profissionais acabam sendo inseridos nesse ambiente, mas nem sempre dominam técnicas e conceitos de gestão, e isso acaba sendo uma barreira na condução desses tipos de projetos, dessa forma, o entrevistado entende que o desenvolvimento de algo que desse suporte nesse tipo de situação seria de grande valor.

Apesar do progresso significativo na área de Big Data realizado nos últimos anos, praticantes destacam importantes oportunidades de pesquisa, principalmente, relacionadas a guias, modelos ou até metodologias (Lara *et al.*, 2020). Dessa forma, entende-se que uma oportunidade de contribuir para a comunidade de profissionais que atuam com projetos de Big Data, poderia ser algo relacionado ao desenvolvimento e validação empírica de um modelo que oriente a estruturação de um projeto de Big Data.

4.3 RESULTADOS DO ESTUDO 3 – PATENTES DE PROJETOS DE BIG DATA

No terceiro estudo (APÊNDICE B), cujo objetivo foi realizar uma análise exploratória de um banco de dados de patentes para coletar informações sobre as características das patentes registradas relacionadas a projetos de Big Data, realizou-se uma pesquisa na base de dados Espacenet no dia 10 de janeiro de 2021 com patentes que continham a seguinte *string* em alguma parte do documento: “*Big data project**”.

Conforme demonstrado na Tabela 11, a China apresentou um número expressivo de registros com 81 patentes, correspondendo a 78,64% da amostra. A China registrou uma patente em 2015 e apresentou crescimento ano a ano até 2020. Os Estados Unidos da América (EUA) são o segundo país com mais registros, dez no total, e a Organização Mundial de Propriedade Intelectual (WIPO) é o terceiro, com seis registros. Além disso, a República da Coreia (ROK) possui três registros, o European Patent Office (EPO) dois e o Japão um. O grande número de registros e o aumento constante da produtividade ao longo dos anos destacam o interesse da China pelo tema.

Tabela 11 - Mapa de intensidade do número de registros em países/regiões relacionados a projetos de Big Data

País/Região	2015	2016	2017	2018	2019	2020	Total
China	1	2	6	13	17	42	81
USA	2	1	1		3	3	10
WIPO		1		1	1	3	6
ROK				1		2	3
EPO				1		1	2
Japão						1	1
Total	3	4	8	15	22	51	103

Ao estudar patentes de Big Data na China, Yang *et al.* (2017) mostraram que de 1980 a 2016 o aumento nos registros de patentes de Big Data na China ocorreu após 2005, e os pedidos de patentes dos reclamantes eram principalmente de universidades e empresas. Teixeira e Queirós (2016) destacam o “Serviço de Propriedade Intelectual Intermediária”, que, desde 2008, tem sido uma estratégia de desenvolvimento tecnológico para fazer da China um país com alto nível de criação de propriedade intelectual. De fato, o mapa de intensidade ilustra a predominância de registros na China, confirmando o sucesso dessa estratégia.

Na geração do dendrograma (Figura 6), que é um diagrama em árvore, mostram-se os grupos formados pelo agrupamento em seus níveis de similaridade, três classes foram identificadas e analisadas de acordo com as relações entre as palavras. A classe 1 representa 32,8% do corpus e possui palavras como “system”, “method”, “model” e “management”, que se relacionam ao corpus como procedimentos ou formas de organização. A classe 1 também contém palavras como “cloud”, “compute”, “platform” e “service”, que podem indicar uma direção para o uso de serviços de computação em nuvem. A classe 2 representa 32% do corpus e é composta por palavras como “efficiency”, “improve”, “process”, “reduce”, “speed” e “solution”, sugerindo que o corpus é direcionado para a otimização de soluções. Além disso, a Classe 2 também contém palavras como “apparatus”, “device”, “equipment” e “wireless”, referindo-se a objetos que facilitam a otimização da solução. Por fim, a Classe 3 representa 35,2% do corpus e é caracterizada por palavras como “store”, “database”, “block”, “record”, “structure”, “module” e “share”, que são indicativos de armazenamento de dados e compartilhamento de estruturas.

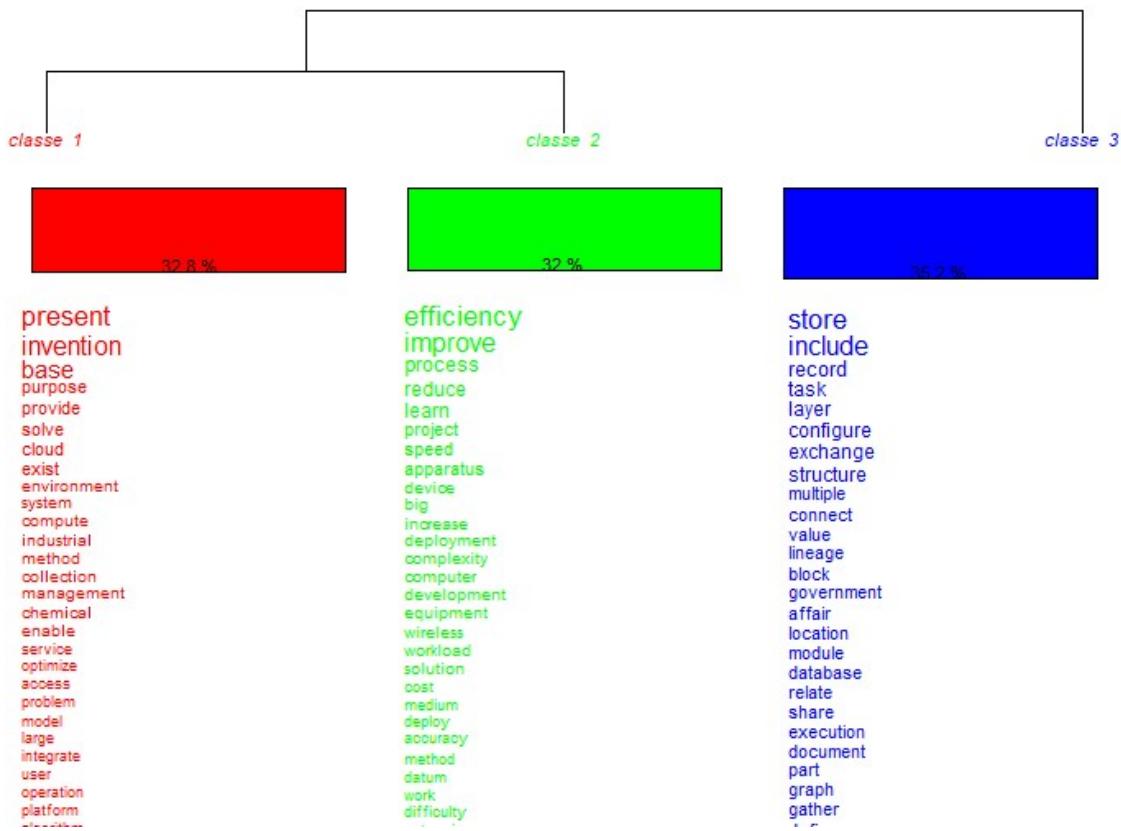


Figura 6 – Dendrograma gerado pelo MCHD

Deve-se ressaltar que uma vez que as Classes 1 e 2 possuem equivalência em termos de meios de palavras em relação à Classe 3; portanto, eles poderiam ter sido combinados em uma classe.

A AFC possibilitou a visualização do corpus textual distribuído no plano cartesiano. Conforme mostrado na Figura 7, as três classes distintas não se sobrepõem. O corpus textual foi formado pelas palavras do resumo da patente, e cada palavra teve que ser repetida no mínimo dez vezes para ser considerada na AFC. A interpretação do posicionamento das palavras é subjetiva, por isso foi necessário interpretar por que as palavras têm essa distribuição. Assim, após o *software* gerar a AFC, foi realizado uma análise profunda direcionada a todas as patentes por suas respectivas palavras para identificar as relações entre elas e formar um corpus comum que justifique cada grupo de palavras.

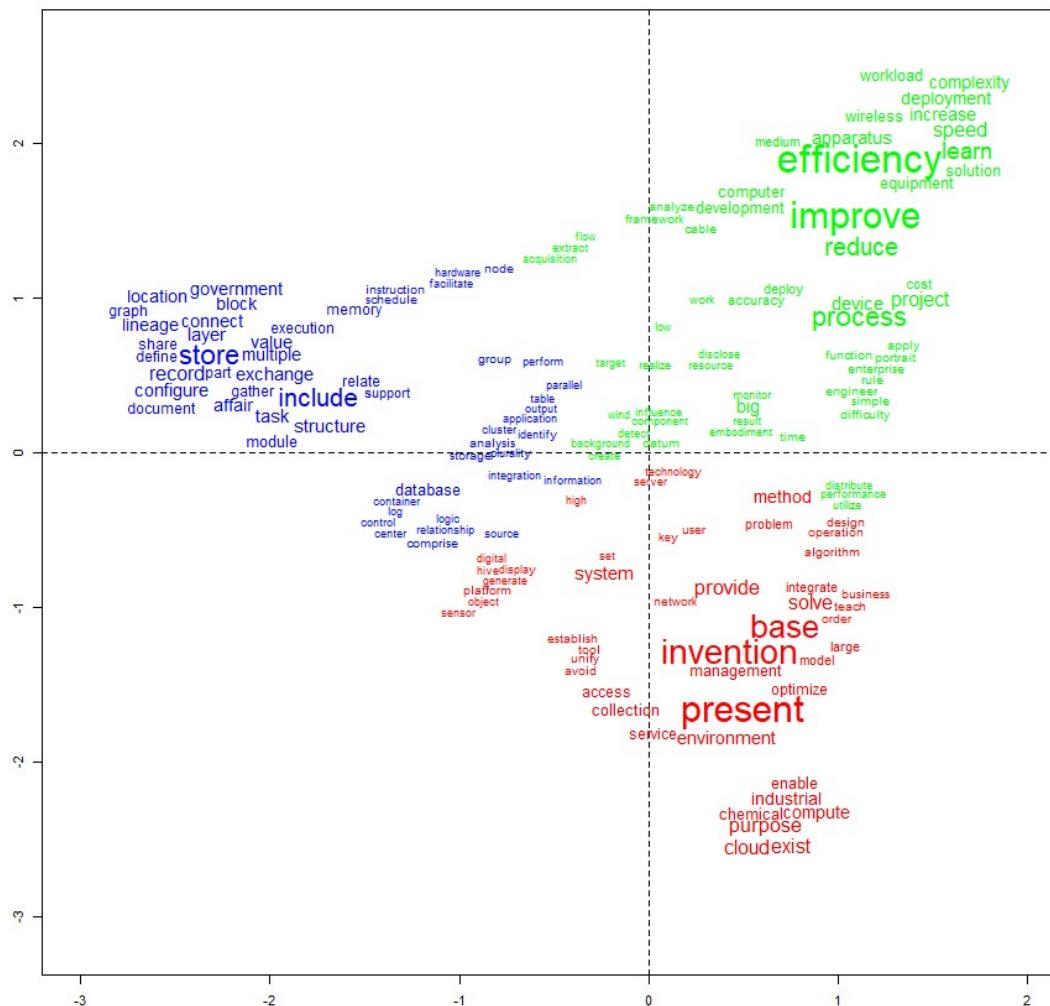


Figura 7 – AFC da amostra

Ao interpretar as distâncias entre as palavras e as distribuições ao longo do plano, o eixo *x* parece referir-se às tecnologias de coleta e armazenamento de dados (à esquerda) e às técnicas de processamento e desempenho (à direita). Em outras palavras, patentes relacionadas a aplicativos de design de Big Data. Por outro lado, o eixo *y* indica uma relação de tangibilidade tal que as partes inferior e superior do plano representam soluções intangíveis e tangíveis, respectivamente.

Devido ao desenvolvimento contínuo, a tecnologia IoT e Big Data se tornaram ferramentas amplamente aplicadas em muitos campos técnicos (Wang, 2020). A rápida evolução dos sistemas de suporte à decisão fez os especialistas perceberem que colocar Big Data na nuvem é um desafio real que as empresas devem considerar. Nesse sentido, a seleção de soluções em nuvem é reconhecida como uma questão de pesquisa essencial para projetos de Big Data (Boutkhoun *et al.*, 2016).

Um desafio significativo enfrentado por esse setor é que o Big Data gerado a partir de vários sensores IoT contém informações espaciotemporais ricas (Yang, Huang, Li, Liu, & Hu, 2016). Foi identificado um projeto de Big Data IoT na nuvem que fornece um método para analisar o grau de segurança contra poluição medindo a luz refletida de uma parede de cortina de vidro (Chen, Gu, & Hua, 2020). Segundo os autores, a invenção ajuda a monitorar continuamente o fenômeno da poluição luminosa na parede de cortina de vidro. Além disso, utilizando Big Data, a patente analisa e visualiza de forma objetiva a segurança da iluminação do prédio, protegendo a visão e melhorando a produtividade dos funcionários.

Foi relatado que o avanço da IoT deve fornecer acesso público fácil a resultados valiosos de Big Data e computação em nuvem (Yang *et al.*, 2016). No entanto, o acesso às melhores informações precisa ser fornecido ao público para uma melhor tomada de decisão (Abbas *et al.*, 2015). A invenção de Pan e Rui (2020) se refere ao campo técnico de fábricas inteligentes e, mais especificamente, a um sistema de processamento de modelagem química baseado em Big Data industrial e IoT. Para os inventores, a invenção visa integrar e implementar tecnologia de IA e Big Data de forma independente e controlável nas linhas de produção tradicionais, ajudando as empresas de produção a obter dados sinérgicos, produção e fluxos de controle, melhorar a eficiência da produção e reduzir os custos de produção. Em outras palavras, essa tecnologia pode contribuir para o desenvolvimento de uma manufatura inteligente, independente e controlável.

Prevê-se que a tecnologia Blockchain superará, em breve, desafios significativos enfrentados nesta área de pesquisa e desenvolvimento. Nesse contexto, foi desenvolvido um

método de extração e gerenciamento de resultados de projetos gerados por especialistas, executados por um servidor e armazenados em uma Blockchain (Kim, 2020). Além disso, também foi identificado um algoritmo e plataforma de consenso Blockchain, tecnologia de criptografia Blockchain e rede evolutiva desenvolvida para reduzir os custos e o tempo de transação de criptomoeda (Ji, Jean, Tae, Cheol, & Dam, 2020). Alkhamisi e Alboraei (2019) afirmam que, com o surgimento da tecnologia Blockchain, pesquisas recentes se concentraram na adoção do Blockchain para resolver problemas de segurança associados à IoT. Além disso, Ferrag *et al.* (2018) classificaram os modelos de ameaças considerados pelos protocolos Blockchain em redes IoT em cinco categorias principais: ataques baseados em identidade, ataques baseados em manipulação, ataques criptanalíticos, ataques baseados em reputação e ataques baseados em serviço.

Na amostra, também foram detectadas patentes relacionadas a um sistema de controle de orientação remota para serviços corporativos (Feng, 2018). Esta invenção torna a consulta de informações de demanda de serviço de uma empresa mais eficiente e conveniente, combinando, de forma eficiente, o cliente, os dados de nuvem da plataforma e agências de serviço terceirizadas. Apesar das inúmeras vantagens, é um desafio encontrar sistemas de recomendação com uma fórmula específica que forneça diretamente parâmetros otimizados. Esse contratempo é atribuído, principalmente, ao fato de que mais e mais parâmetros são gerados conforme os modelos de aprendizado de máquina se tornam mais e mais complexos. Notavelmente, foi proposto que um algoritmo de descida gradiente mínimo baseado na dimensão *Gene*, *Behavior* e *Phenotypic* poderia mitigar esse problema (Liu, Wang, & Wu, 2020). Os sistemas de recomendação coletam classificações históricas do usuário, interações, retratos, redes sociais e contexto e analisam interesses históricos para recomendar informações relevantes para os usuários atenderem às necessidades e algoritmos de recomendação, que são o núcleo dos sistemas de recomendação.

A mineração de dados desempenha um papel crucial na análise de grandes conjuntos de dados. A invenção de Bender, Celi, Greenlee e Jaramillo (2019) refere-se a um método implementado por um computador para analisar documentos técnicos relativos a um corpus. De acordo com os inventores, vários algoritmos podem realizar pesquisas de dados; no entanto, a recuperação de informações costuma ser um processo impreciso com o surgimento de projetos de Big Data. Portanto, melhorias na recuperação da informação são necessárias. Também encontrou-se uma série de patentes relacionadas à otimização de desempenho. Por exemplo, Dai (2014) descreve um método e dispositivo para otimizar o desempenho de um banco de

dados distribuído. Chen, Lin e Luo (2016) fornecem um método para atualizar e monitorar Big Data, que resolve os problemas associados aos componentes durante a coleta de dados distribuída tradicional.

Quando um processo ETL é implementado como um fluxo de trabalho, em que as tarefas processam dados e são conectadas por fluxos de dados, ferramentas ETL devem ser desenvolvidas para executar esses fluxos de trabalho (Ali & Wrembel, 2017). A invenção de Chen, Liao, Lin e Yang (2017) é uma ferramenta ETL que aborda um sistema de agendamento Big Data ETL, apoiando a visualização e o fluxo de processo, eliminando operações complicadas em segundo plano, melhorando a velocidade e eficiência de desenvolvimento ETL e reduzindo os custos de implementação. Foi relatado que a estratégia de arquitetura ETL processa os dados de maneira uniforme (Jianmin *et al.*, 2020). Além disso, Zhao (2017) divulga um método de extensão de aplicativo Apache NiFi para importar dados de origem para um banco de dados. De acordo com o método, diferentes ferramentas de acesso não são necessárias quando várias fontes de dados têm acesso ao banco de dados. Consequentemente, o processo de acesso aos dados é unificado. Um *plug-in* de acesso também pode ser estendido automaticamente.

A cultura de tomada de decisão baseada em dados promoveu mudanças estruturais em processos, pessoas e tecnologia (Elgendi & Elragal, 2016). O valor gerado a partir do uso de Big Data no processo de tomada de decisão está na habilidade dos tomadores de decisão em alocar recursos e fazer ajustes no processo de forma mais assertiva. Na verdade, algumas patentes identificadas na amostra melhoram potencialmente a tomada de decisão. No entanto, como Boutkhoum *et al.* (2017) apontaram, selecionar a solução mais adequada para projetos de Big Data é complexo e requer um processo de avaliação abrangente.

A invenção de Wang e Xu (2016) apresentou um modelo analítico que avalia Big Data para conservação de água. Este método aplica um processo de hierarquia analítica e uma teoria matemática difusa para Big Data, considerando de forma abrangente vários esquemas de serviço de água. Uma vez que os esquemas são classificados, o melhor esquema de tomada de decisão é selecionado. Outra patente (Wang, 2020) se refere a um método e sistema para coletar dados de custos de engenharia com base em Big Data, equipamento de computador e um meio de armazenamento legível por computador. De acordo com os inventores, o método tem a vantagem de que diferentes tomadores de decisão podem dominar os dados de custo para cada estágio em cada campo da engenharia de maneira conveniente. O custo total de engenharia é obtido como um todo e, por fim, uma decisão razoável é tomada.

Destaca-se também a patente de Du *et al.* (2020), que pertence ao campo de visão mecânica. Esta invenção divulga um método de gestão digital para o plantio de cânhamo na China. As empresas de plantio de cânhamo da China não monitoram os principais pontos de controle, e a operação manual, geralmente, gera algumas informações de histerese. Consequentemente, o tempo de tomada de decisão e ações de correção de desvio são afetados negativamente. Esta invenção usa um banco de dados de biblioteca de vídeo de imagens de crescimento de plantas e monitora as imagens para construir uma biblioteca de modelos e uma base de conhecimento. O sistema pode realizar processamento e reconhecimento de imagens em lote, configurando parâmetros para auxiliar na tomada de decisões e melhorar a produtividade.

O Big Data nas empresas modernas permite que as empresas tomem decisões mais rápidas e sábias e obtenham vantagens competitivas. No entanto, muitos projetos de Big Data fornecem resultados decepcionantes que não atendem às necessidades do tomador de decisão. Esses resultados ruins ocorrem porque eles negligenciam a decisão desses projetos, que é um aspecto do processo de tomada de decisão (Chiheb *et al.*, 2019). Afirma-se que com a crescente demanda pelo uso do Big Data para aproveitar suas oportunidades, as organizações buscam soluções e diretrizes claras e simples para a gestão do Big Data (Elgendi & Elragal, 2016). Conforme destacado por Poleto *et al.* (2017), o uso do Big Data só terá sucesso se houver uma estratégia de integração de dados para geração de informações relevantes e gestão do conhecimento.

Por fim, destaca-se requerentes com mais de um registro na amostra. Ao todo, nove candidatos se enquadram nesse requisito, incluindo Empresas Limitadas com 11 registros, seguidas por universidades com dois, um instituto de pesquisa com um e um indivíduo com um (Tabela 12).

Os dados de patentes apresentam percepções valiosas para inventores, engenheiros, empresas e tomadores de decisão (Akers, 2003). É também considerado um indicador de inovação e reflete mudanças tecnológicas e científicas (Archibugi, 1992).

Tabela 12 - Requerentes com mais de uma patente registrada

Requerentes	Número da patente	Nome da patente
<i>Chongqing Socialcredits Big Data Tech Co Ltd</i>	CN109408712A	<i>A method for constructing a multi-dimensional information portrait of a travel agency user</i>
	CN109345348A	<i>Recommendation method of multi-dimensional information portrait based on travel agency users</i>
<i>Electric Power Res Inst</i>	CN109190381A	<i>A detection method for Hadoop security vulnerability</i>
	CN109961376A	<i>Distributed energy storage equipment management and control system and method</i>
<i>H3C Big Data Tech Co Ltd</i>	CN108062399A	<i>Data processing method and device</i>
	CN109344163A	<i>Data verification method, apparatus, and computer-readable medium</i>
<i>Lenovo Beijing Co Ltd</i>	CN108804233A	<i>Memory space recycling method, device, electronic equipment and storage medium</i>
	CN106293929A	<i>Data processing method and first electronic equipment</i>
<i>Shenzhen OneConnect Intelligent Technology Co Ltd</i>	CN111984685A	<i>Data tilt detection method, device, computer equipment and readable storage medium</i>
	CN111984630A	<i>Log association method, device and computer equipment</i>
	CN111158642A	<i>Data construction method and device, computer equipment and storage medium</i>
<i>South China University of Technology</i>	CN111444236A	<i>Mobile terminal user portrait construction method and system based on Big Data</i>
	CN107817787A	<i>Intelligent production line manipulator fault diagnosis method based on machine learning</i>
<i>Xiamen University</i>	CN207423929U	<i>Novel air quality intelligence prediction system</i>
	CN111310868A	<i>Water-writing handwritten character recognition method based on convolutional neural network</i>
<i>Wang Yujue</i>	CN109650186A	<i>Communication cable taking-up and paying-off device for communication Big Data project</i>
	CN109650187A	<i>Communication cable taking-up and paying-off device for communication Big Data project</i>
<i>Xinhua Sanda Data Tech Co Ltd</i>	CN107548151A	<i>Method and device for marking wireless access point (AP)</i>
	CN107547519A	<i>Wireless Big Data deployment method and device</i>

Nesse sentido, o mapeamento de invenções e, consequentemente, os requerentes podem revelar fortes indicadores de tendências tecnológicas e, em alguns casos, domínio de uma técnica ou tecnologia. Portanto, tanto a academia quanto a indústria devem reconhecer os padrões emergentes de tecnologias de Big Data, que são de importância primordial no crescimento das empresas orientadas a dados (Saheb & Saheb, 2020).

4.4 RESULTADOS DO ESTUDO 4 – VALIDAÇÃO DO MODELO

No quarto estudo, cujo objetivo foi validar o modelo para a estruturação de um projeto de Big Data, foram conduzidas três entrevistas com grupos focais com profissionais que atuam com projetos de Big Data. O processo de validação do modelo gerencial ocorreu com a

apresentação integral do modelo e, posteriormente, os entrevistados faziam comentários indicando pontos de atenção, realizando questionamentos e sugerindo melhorias. Toda sugestão de melhoria era avaliada após a seção. Cabe destacar que nem todas as sugestões foram contempladas no modelo, pois foram avaliadas quanto à viabilidade e foco da proposta. Também era comum a discussão entre os próprios entrevistados de modo a legitimar alguma colocação feita com um exemplo passado no dia a dia.

Desse modo, após a aplicação das sessões de grupo focal, evidenciou-se mais sugestões de melhorias para a Fase estratégica I, sendo seis ao todo (Figura 8). Ficou evidente a preocupação com a primeira fase do modelo de modo a direcionar as fases subsequentes. Destaca-se a sugestão para a classificação dos problemas em ‘puxado’ e ‘empurrado’, onde se tem a preocupação em definir se o problema a ser resolvido parte de uma demanda clara ou se é um problema orientado a um viés ou desejo. Nesta mesma linha, as sugestões para a estruturação de uma fase de gestão de expectativas, em que é realizado um mapeamento de requisitos, onde se identificam riscos, faz-se um alinhamento sobre a possibilidade de insucesso do projeto, ocorre uma análise de viabilidade com especialistas e também se realiza o *input* de lições aprendidas de projetos anteriores.

Seção	Fase Estratégica I	Fase Tecnológica	Fase Analítica	Fase Estratégica II	TOTAL
01	3		2	1	6
02	2	1	2		5
03	1	1		1	3
TOTAL	6	2	4	2	

Figura 8 – Sugestões de melhorias feitas ao longo da discussão

Na fase tecnológica, com duas sugestões, houve um incremento no processo de adequação do projeto conforme legislação ou regras onde também foi considerada a definição de políticas de acesso aos dados. Na fase analítica, com quatro sugestões, foi incorporado um processo de legitimação por especialistas, em que deve ser composta uma equipe multidisciplinar com capacidade de percepção do contexto no qual o problema está inserido. Esses especialistas podem mitigar o risco de alguma variável não ter sido considerada, ou ainda, compartilhar conhecimento tácito, que muitas vezes é difícil de ser incorporado nos processos.

Por fim, na Fase estratégica II com duas sugestões, houve a inclusão de um ponto de decisão onde se verifica se, após toda a análise realizada e *insight* gerado, ainda há *small data* a ser analisado, proveniente de algum fracionamento do banco de dados. Com isso tem-se a possibilidade de realizar novas análises, que podem indicar tendências a partir desse conjunto de dados menor.

Um ponto de destaque foi que a cada seção, menos sugestões foram dadas de modo a indicar uma aceitação cada vez maior do modelo, visto que, com menos sugestões notou-se a legitimização e concordância dos processos pelos entrevistados. Com relação às dúvidas levantadas ao longo da discussão, foram registradas oito dúvidas que ocorreram ao longo das apresentações (Figura 9).

Seção	Fase Estratégica	Fase Tecnológica	Fase Analítica	Fase Estratégica	TOTAL
01	1				1
02	2	1			3
03	2	2			4
TOTAL	5	3	0	0	

Figura 9 – Dúvidas levantadas ao longo da discussão

As dúvidas surgiram durante a apresentação do modelo e sempre eram sanadas no momento que o questionamento era realizado. As dúvidas ocorriam por uma demonstração de curiosidade de algum dos entrevistados que participaram das sessões e, na maioria das vezes, etapas subsequentes ao processo que estava sendo explicado respondia a dúvida levantada pelo entrevistado.

Ao final das três sessões, como forma de buscar a aceitação do modelo e considerá-lo oportuno para a utilização, o seguinte questionamento era feito aos grupos focais: Vocês entendem que o modelo pode ser útil quando utilizado como um guia para estruturação de projetos de Big Data? Nas três sessões as respostas foram positivas, e em uma das sessões, um dos entrevistados inclusive retomou uma situação vivida e, conforme sua fala, se ele tivesse esse modelo na época, teria evitado um problema que ele acabou passando.

Com as validações evidenciadas em cada seção, o modelo foi considerado validado por especialistas que atuam profissionalmente com projeto de Big Data, com isso foi realizado o processo de solicitação de propriedade intelectual do modelo.

5 PRODUTO TECNOLÓGICO

Foi reivindicada a autoria de um modelo de utilidade de título “Modelo gerencial para estruturação de projetos de Big Data” como produto tecnológico desta tese. Modelos de utilidade são concebidos para proteger melhorias funcionais em registros de uso prático, e o modelo de utilidade proveniente desta tese é um modelo gerencial para estruturação de projetos de Big Data, baseado nas fases estratégicas, tecnológica e analítica, podendo ser aplicado para projetos de diferentes áreas e setores.

Na Figura 10, consta parte do comprovante da solicitação do registro de modelo de utilidade, cujo depósito foi realizado no dia 02/08/2021 no Instituto Nacional de Propriedade Industrial (INPI).



Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

Número do Processo: BR 20 2021 015271 6

Dados do Depositante (71)

Depositante 1 de 3

Nome ou Razão Social: GUSTAVO GRANDER

Tipo de Pessoa: Pessoa Física

Figura 10 – Pedido de modelo de utilidade

O produto tecnológico aqui apresentado é uma produção de alto teor inovativo, devido à combinação de conhecimentos pré-estabelecidos e a criação a partir da pesquisa de novos procedimentos. Possui aplicabilidade potencialmente elevada, visto que pode ser utilizado por outras pessoas que enfrentam os mesmos problemas em outros ambientes profissionais, corporativos ou acadêmicos. O produto tecnológico também possui uma alta complexidade, visto que as variáveis relacionadas ao estudo são diversas e interconectadas. Além disso,

destaca-se uma alta possibilidade de impacto, pelo potencial para promover considerável mudança no ambiente social inserido, visto o potencial para resolução de problemas.

O modelo é caracterizado por um conjunto de fases que representam o ciclo de um projeto de Big Data (Figura 11). A primeira fase do modelo, Fase Estratégica I, contempla um conjunto de etapas que vão sustentar o planejamento do projeto de Big Data, com o início dado após um gatilho, que pode ser um evento ou um fenômeno externo e que estimula a reflexão do problema a ser resolvido, ser puxado ou empurrado. Após essa definição, tem-se as etapas de gestão de expectativas, a definição da hipótese ou objetivo gerado, então é definido o tipo de análise que deve ser considerada, de acordo com a maneira que se pretende olhar e, com isso, encerra-se a primeira fase.

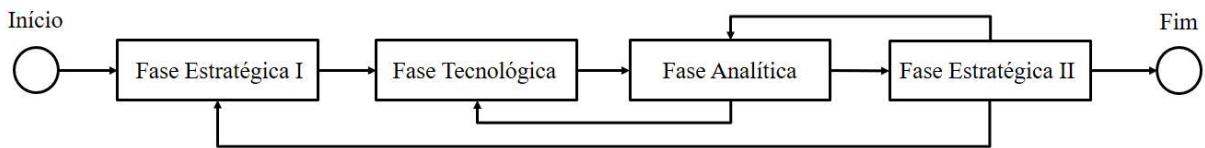


Figura 11 – Fases do modelo gerencial para estruturação de projetos de Big Data

A segunda fase, a Fase Tecnológica, inicia com o questionamento de haver algum impacto ético ou legal no que se deseja investigar, e com a posterior adequação em casos afirmativos, parte-se então para a modelagem da arquitetura, coleta dos dados e armazenamento, quando aplicável. A terceira fase, a Fase Analítica, é a fase mais extensa do modelo com o início no pré-processamento de dados, então com base no tipo de análise definido na primeira fase, define-se a técnica de análise dos dados que será utilizada, podendo resultar em algoritmos de aprendizado supervisionado, semi-supervisionado, ou não supervisionado para análises preditivas ou prescritivas, e técnicas de *Analytics*, em casos de análises descritivas.

Por fim, a quarta e última fase é a Fase Estratégica II, que se inicia com o suporte para tomada de decisão consequente do *output* da fase analítica. Esse *output* condiciona a incorporação de um novo processo no ambiente do projeto de Big Data, em seguida gera-se a documentação de lições aprendidas do projeto, e então deve-se questionar se o *insight* gerado resultou em um novo gatilho. Em caso afirmativo, o processo deve ser retomado para o início de todo o fluxo do projeto de Big Data com o novo questionamento e definição do tipo de problema. Em caso negativo, deve-se questionar se há *small data* não analisado, proveniente de alguma filtragem ou limpeza na base de dados coletada. *Small datas* são conjuntos menores de dados, em relação ao Big Data, sendo conjuntos de menor complexibilidade de análise e, muitas

vezes, representam um comportamento mais específico de um grupo. Em caso afirmativo, o processo deve ser reiniciado a partir da fase analítica e, em caso negativo, encerra-se o ciclo de vida do projeto de Big Data.

O principal benefício do modelo é proporcionar a integração das fases pertinentes ao projeto de Big Data e a relação de dependência entre elas, característica muitas vezes negligenciada em projetos que não seguem um roteiro pré-definido, gerando fragilidades na estruturação do projeto. O modelo ainda mitiga o problema de adoção de um projeto de Big Data pelo modismo, pois cada fase e processo acaba sendo um importante filtro de refino do projeto e, na medida que as etapas forem negligenciadas, a fragilidade do projeto fica latente.

O modelo possui alta aderência ao Programa de Pós Graduação em Administração - Gerenciamento de Projetos (PPGP) da Universidade Nove de Julho, por conta do apelo de integração entre o ambiente acadêmico e profissional, assim como, por se tratar de um instrumento com propósito claro de resolução de problemas práticos em contextos profissionais e sociais ligados à gestão de projetos. Cabe salientar que os critérios de avaliação utilizados foram baseados no guia de Produção Técnica (GT 06 – Qualis Técnico/Tecnológico), instituído pela CAPES.

Esta pesquisa está inserida na linha de pesquisa de Gerenciamento em Projetos e eixo temático “Sociedade 4.0 e Economia Circular” pertencentes ao PPGP. Este eixo temático visa desenvolver pesquisas que possibilitem compreender e integrar o mundo digital e o mundo físico a fim de fornecer soluções e explicações que auxiliam em lidar com mudanças nas relações sociais, comerciais e produtivas.

6 CONCLUSÃO

O trabalho desenvolvido nesta tese traz contribuições tanto conceituais quanto práticas para a área de projetos de Big Data. No primeiro estudo foi apresentado um panorama das técnicas e tecnologias que vêm sendo discutidas na literatura para solucionar problemas em determinadas áreas, a classificação descritiva deste estudo serve como um guia para novos pesquisadores que buscam entender as pesquisas envolvendo sistemas de suporte à decisão e Big Data para agregar valor na sociedade.

Alinhado com os resultados do primeiro estudo, o artigo resultante classificou técnicas como aprendizado supervisionado, semi-supervisionado e não supervisionado, bem como técnicas de *Big Data Analytics*, e também tecnologias como *Cloud Computing*, *Business Intelligence* e *Data Warehouse*. Além das classificações citadas, também se destacam os tipos de tratamentos de dados possíveis, de acordo com o objetivo de análise, podendo ser análises descritivas, preditivas ou prescritivas. Cabe salientar, ainda, o número expressivo de análises preditivas provenientes dos estudos. Estas foram as principais contribuições teóricas do estudo. Por meio do primeiro estudo, o primeiro objetivo específico da tese foi atingido, que foi analisar como sistemas de suporte à decisão gerenciam Big Data para obtenção de valor.

No segundo estudo, em que foi explorada a estrutura e o contexto de um projeto de Big Data, foi possível sintetizar a definição, as fases, fatores críticos e problemas de tomada de decisão em projetos de Big Data, pela percepção de profissionais que atuam na área. Por meio desse estudo, foi possível identificar como oportunidade de pesquisa o desenvolvimento do modelo proposto nesta tese, como forma de auxiliar a estruturação de projetos de Big Data, visto que os profissionais que participaram deste estudo destacaram que seria algo valoroso.

A partir dos resultados obtidos nas entrevistas foi possível estruturar um esboço do modelo gerencial com base na realidade observada. Ressalta-se, aqui, a importância das entrevistas em profundidade, para depreender as especificidades de um projeto de Big Data. Neste sentido, os entrevistados apontaram os desafios e as possíveis soluções para gerenciar um projeto de Big Data, sendo esta uma contribuição relevante para a construção do modelo e avanço conceitual da tese. Essa fase da pesquisa suportou o estudo de patentes de forma estruturada para alcançar um modelo que, de fato, traga contribuições aos praticantes e à academia.

No terceiro estudo, em que se buscou compreender as características de patentes relacionadas a projetos de Big Data, por meio de uma análise textual, foram detectadas três classes de palavras que direcionam as patentes para (i) uma direção de computação em nuvem, (ii) otimização de soluções, e (iii) estruturas de armazenamento e compartilhamento de dados. Foram constatadas tecnologias emergentes como Blockchain e IoT, que são utilizadas em soluções de projetos de Big Data e cabe destacar a China como país predominante em número de registros de propriedade intelectual, representando mais de 80% total de patentes analisadas.

O terceiro estudo possibilitou concluir o segundo objetivo específico da tese, que foi realizar uma análise exploratória de um banco de dados de patentes para coletar informações sobre as características das patentes registradas relacionadas à projetos de Big Data. Os estudos um, dois e três possibilitaram desenvolver um modelo gerencial para estruturação de projetos de Big Data.

No quarto estudo, e com influência do segundo, foi desenvolvido um modelo gerencial para estruturação de projetos de Big Data. Este levou em consideração as fases estratégicas, tecnológica e analítica, sendo esta uma contribuição prática que atende o terceiro objetivo específico da tese. O modelo em questão pode ser aplicado para projetos de diferentes áreas e setores. Assim, os quatro estudos, permitiram atingir o objetivo geral da tese, que foi propor um modelo gerencial para estruturação de projetos de Big Data.

Parte-se da premissa de que o modelo, pelo aspecto gerencial, pode ser aplicado para projetos de diferentes áreas e setores, característica assimilada pelos entrevistados do quarto estudo. Contudo, um aspecto que gerou reflexão foi a expectativa do modelo ter um maior nível de detalhamento, principalmente na fase tecnológica. É entendido que, quanto maior o nível de detalhamento do modelo, menor a sua capacidade de aplicação em um escopo mais amplo de projetos de Big Data, pois ele passa a ser mais restrito a determinadas áreas ou setores. O mesmo pode acontecer com as prescrições técnicas, que limitam o campo de aplicação do modelo.

Portanto, o modelo, em um nível mais abrangente, mantém o aspecto gerencial possibilitando um nível de gestão de alto nível nas organizações. Recomenda-se que, durante a estruturação do projeto de Big Data, os processos devam ser estruturados com maior nível de detalhamento, conforme os praticantes entenderem necessário. O modelo serve como uma referência e deve ser utilizado por praticantes de acordo com a situação e a própria experiência e recursos disponíveis da equipe do projeto.

O modelo possui potencial para mitigar a alta taxa de falhas em projetos de Big Data (Barham & Daim, 2020), pois as diferentes fases do modelo deixam evidente o perfil multidisciplinar que a equipe deve possuir para atuar com aspectos estratégicos, tecnológico e analítico. Há ainda, o aspecto do adequado alinhamento do projeto, promovendo um maior controle e reduzindo as entradas e resultados incertos dos projetos (Saltz *et al.*, 2017).

O modelo visa portanto, e de uma forma mais abrangente, otimizar a obtenção de resultados desejados pela forma de estruturar esses projetos, dando condições para tomadas de decisões mais assertivas (Chiheb *et al.*, 2019). E, como já mencionado anteriormente, o uso eficaz de Big Data se dá pela capacidade de transformar dados em *insights* acionáveis (Surbakti, 2021), e o modelo de utilidade proposto proporciona uma maior clareza ao longo dos processos, deixando essa percepção dos *insights* como algo mais perceptível.

O fato do modelo não ser técnico como alguns dos entrevistados demonstraram interesse, faz com que este aspecto seja um limitador para o modelo. Neste sentido, em uma das sessões das entrevistas com grupos focais, um dos entrevistados expressou sua vontade de ter um detalhamento mais técnico, principalmente para a fase tecnológica. Esta situação poderá ser evidenciada por praticantes quando o modelo se tornar público – com o aceite e publicação do modelo de utilidade pelo INPI. Embora essa seja uma limitação, cabe destacar que este não era o escopo desta pesquisa ou do modelo proposto. Caso aprovada a reivindicação de propriedade intelectual do modelo, esta se torna a principal contribuição prática gerada nesta tese.

Essa restrição, por outro lado, pode ser uma importante oportunidade de pesquisa futura, pois novos estudos podem se aprofundar no modelo com níveis técnicos complementares, ou estudos direcionados a um maior nível de detalhamento em alguma fase específica do modelo. Também, destaca-se a possibilidade de aplicação do modelo em um projeto de Big Data, visto que a validação realizada no estudo 4 ocorreu conforme a percepção dos entrevistados.

Cabe registrar também, importantes lições aprendidas provenientes desta pesquisa. A sequência dos estudos foi um fator relevante para que fosse possível ter um modelo de utilidade como *output*. Realizar uma revisão sistemática da literatura como primeiro estudo, sob um prisma mais abrangente, possibilitou uma melhor compreensão do fenômeno de pesquisa. Buscar uma oportunidade de pesquisa com profissionais que atuam na área trouxe robustez e credibilidade ao objetivo de pesquisa, visto que esforços foram direcionados para desenvolver algo que sanasse uma carência latente demonstrada pelos profissionais entrevistados no segundo estudo.

A pesquisa em uma base de patentes também foi um importante passo da tese, pois a investigação em patentes possibilitou identificar se já haviam reivindicações de modelos semelhantes com o que foi desenvolvido, isso seria uma barreira quanto ao ineditismo da reivindicação. Outro ponto importante da pesquisa de patentes é que o próprio INPI realiza uma investigação de ineditismo das solicitações, então a pesquisa realizada com antecedência deu sustentação para se seguir com a reivindicação da propriedade inelectual do modelo.

A discussão tratada no início desta tese, sobre a problemática de insucesso em projetos de Big Data (Mousanif *et al.*, 2014), pode, portanto, ser mitigada por meio do modelo proveniente desta pesquisa. O modelo, por meio de sucessivos processos em que se exige tomadas de decisões estratégicas, tecnológicas e analíticas, condiciona a um *insight* açãoável e perceptível – este que deve ser o objetivo fim de um projeto de Big Data (Surbakti, 2021). Assim, esta é a principal lição aprendida pela lente da redução de altas taxas de insucesso em projetos de Big Data, a condução desses projetos devem proporcionar tomadas de decisões assertivas para atender objetivos de negócios.

O modelo não foi apresentado integralmente neste documento por uma opção em manter o ineditismo no documento depositado ao INPI. Por mais que exista o período de graça que garante o direito de divulgação em até um ano antes do depósito, sem problemas com a incidência do art. 12, da Lei de Propriedade Industrial, o modelo só poderá ser acessado quando publicado.

REFERÊNCIAS

- Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3–13. <https://doi.org/10.1016/j.wpi.2013.12.006>
- Ahangama, S. & Poo, D. (2015). Improving health analytic process through project, communication and knowledge management. In: *The International Conference on Information Systems*.
- Ahmed, V., Aziz, Z., Tezel, A., & Riaz, Z. (2018). Challenges and drivers for data mining in the AEC sector. *Engineering, Construction and Architectural Management*, 25(11), 1436–1453.
- Akers, L. (2003). The future of patent information—a user with a view. *World Patent Information*, 25(4), 303–312. [https://doi.org/10.1016/s0172-2190\(03\)00086-3](https://doi.org/10.1016/s0172-2190(03)00086-3)
- Al Chahadah, A., El Refae, G. A., & Qasim, A. (2018). The use of data mining techniques in accounting and finance as a corporate strategic tool: an empirical investigation on banks operating in emerging economies. *International Journal of Economics and Business Research*, 15(4), 442–452.
- Ali, S. M. F., & Wrembel, R. (2017). From conceptual design to performance optimization of ETL workflows: current state of research and open problems. *The VLDB Journal*, 26(6), 777–801. <https://doi.org/10.1007/s00778-017-0477-2>
- Alkhamisi, A.O., & Alboreai, F. (2019). Privacy-aware Decentralized and Scalable Access Control Management for IoT Environment. *JKAU*, 8, 71–84. <https://doi.org/10.4197/comp.8-1.7>.
- Almeida, R., Bernardino, J., & Furtado, P. (2015). Testing SQL and NoSQL approaches for big data warehouse systems. *International Journal of Business Process Integration and Management*, 7(4).
- Archibugi, D. (1992). Patenting as an indicator of technological innovation: A review. *Science and Public Policy*, 19(6), 357–368. <https://doi.org/10.1093/spp/19.6.357>
- Ashrafi, N., Kuilboer, J., Joshi, C., Ran, I. & Pande, P. (2014). Health Informatics in the Classroom: An Empirical Study to Investigate Higher Education's Response to Healthcare Transformation. *Journal of Information Systems Education*, 25(4), 305–315.
- Aversa, P., Cabantous, L., & Haefliger, S. (2018). When decision support systems fail: Insights for strategic information systems from Formula 1. *The Journal of Strategic Information Systems*.
- Baechle, C., Agarwal, A., & Zhu, X. (2017). Big data driven co-occurring evidence discovery in chronic obstructive pulmonary disease patients. *Journal of Big Data*, 4(1).
- Barham, H., & Daim, T. (2020). The Use of Readiness Assessment for Big Data Projects.

- Sustainable Cities and Society*, 102233. <https://doi.org/10.1016/j.scs.2020.102233>
- Barham, H., & Daim, T. (2018). Identifying Critical Issues in Smart City Big Data Project Implementation. *In Proceedings of the 1st ACM/EIGSCC Symposium on Smart Cities and Communities*, 1 – 9.
- Becker, D. K. (2017). Predicting Outcomes for Big Data Projects: Big Data Project Dynamics (BDPD). *In: 2017 IEEE International Conference on Big Data (BIGDATA)*, 2320 – 2330.
- Bender, M., Celi, J., Greenlee, G. G., Jaramillo, D. (2019). Analyzing technical documents against known art (U.S. Patent No. US10282603B2).
<https://worldwide.espacenet.com/patent/search/family/062783206/publication/US10282603B2?q=US10282603B2>
- Berman, J. J. (2018). Big Data Failures and How to Avoid (Some of) Them. *Principles and Practice of Big Data*, 321–349. <https://doi.org/10.1016/b978-0-12-815609-4.00015-7>
- Bhardwaj, N., Wodajo, B., Spano, A., Neal, S., & Coustasse, A. (2018). The Impact of Big Data on Chronic Disease Management. *Health Care Manager*, 37(1), 90–98.
<https://doi.org/10.1097/HCM.0000000000000194>
- Boutkhoum, O., Hanine, M., Agouti, T., & Tikniouine, A. (2017). A decision-making approach based on fuzzy AHP-TOPSIS methodology for selecting the appropriate cloud solution to manage big data projects. *International Journal of System Assurance Engineering and Management*, 8(S2).
- Boutkhoum, O., Hanine, M., Agouti, T., & Tikniouine, A. (2016). Selection problem of cloud solution for big data accessing: fuzzy AHPPROMETHEE as a proposed methodology. *Journal of Digital Information Management*, 14(6), 368 – 382.
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication, & Society* 15(5), 662-679.
- Borthick, A. F. & Pennington, R. R. (2017). When data become ubiquitous, what becomes of accounting and assurance? *Journal of Information Systems*, 31(3), 1–4.
<https://doi.org/10.2308/isys-10554>
- Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Pathfinder International Tool Series.
- Brinch, M., Stentoft, J., Jensen, J. K., & Rajkumar, C. (2018). Practitioners understanding of big data and its applications in supply chain management. *The International Journal of Logistics Management*, 29(2), 555–574.
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?*
- Bughin, J. (2016). Big data, Big bang? *Journal of Big Data*, 3(1).

<https://doi.org/10.1186/s40537-015-0014-3>

- Burattin, A., Cimitile, M., Maggi, F. M., & Sperduti, A. (2015). Online Discovery of Declarative Process Models from Event Streams. *IEEE Transactions on Services Computing*, 8(6), 833–846.
- Camargo, B. V. (2005). *Alceste: Um programa informático de análise quantitativa de dados textuais*. João Pessoa: E. Universitária.
- Capobianco, E., & Liò, P. (2015). Comorbidity networks: beyond disease correlations. *Journal of Complex Networks*, 3(3), 319–332.
- Chan, S. W. K., & Chong, M. W. C. (2017). Sentiment analysis in financial texts. *Decision Support Systems*, 94, 53–64.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative research*. London: Sage.
- Chen, B., Lin, K. & Luo, Y. (2016). *Big-data distributed data acquisition device deploying, upgrading and monitoring solution* (CN Patent No. CN106130801A).
<https://worldwide.espacenet.com/patent/search/family/057273073/publication/CN106130801A?q=CN106130801A>
- Chen, C., Achtari, G., Majkut, K., & Sheu, J.-B. (2017). Balancing equity and cost in rural transportation management with multi-objective utility analysis and data envelopment analysis: A case of Quinte West. *Transportation Research Part A: Policy and Practice*, 95, 148–165.
- Chen, C., Gu, W. & Hua, Y. (2020). Glass curtain wall light reflection pollution safety degree analysis method (CN Patent No. CN112001023A).
<https://worldwide.espacenet.com/patent/search/family/073464775/publication/CN112001023A?q=CN112001023A>
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Netw Appl*, 19, 171–209.
- Chen, S., Liao, Y., Lin, Z. & Yang, Z. (2017). *Big data ETL dispatching system, capable of supporting visualization and process-oriented implementation* (CN Patent No. CN107145576A).
<https://worldwide.espacenet.com/patent/search/family/059776923/publication/CN107145576A?q=CN107145576A>
- Chen, W., & Quan-Haase, A. (2018). Big Data Ethics and Politics: Toward New Understandings. *Social Science Computer Review*, 1–7.
<https://doi.org/10.1177/0894439318810734>
- Chiheb, F., Boumahdi, F., & Bouarfa, H. (2019). A Conceptual Model for Describing the Integration of Decision Aspect into Big Data. *International Journal of Information System Modeling and Design*, 10(4), p. 1 – 23.
- Clarke, N. S. 2018. The basics of patent searching. *World Patent Information*, 54, S4–S10.

<https://doi.org/10.1016/j.wpi.2017.02.006>

Creswell, J. W. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dai, B. (2014). *Method and device for performance optimization of distributed database* (CN Patent No. CN103823881A).

<https://worldwide.espacenet.com/patent/search/family/050758945/publication/CN103823881A?q=CN103823881A>

Demirkan, H, & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55, 412-421.

Deal, B., Pan, H., Pallathucheril, V., & Fulton, G. (2017). Urban Resilience and Planning Support Systems: The Need for Sentience. *Journal of Urban Technology*, 24(1), 29–45.

Demchenko, Y., Grosso, P., de Laat, C., & Membrey, P. (2013). Addressing big data issues in Scientific Data Infrastructure. *2013 International Conference on Collaboration Technologies and Systems (CTS)*. <https://doi.org/10.1109/cts.2013.6567203>

Dietrich, D. I., Reiner, D. S (2017). *Holistic methodology for big data analytics* (U.S. Patent No. US9798788B1).

<https://worldwide.espacenet.com/patent/search/family/060082219/publication/US9798788B1?q=US9798788B1>

Dilley, P. (2000). *Conducting Successful Interviews: Tips for Intrepid Research*. Theory Into Practice, 39(3), 131–137.

Dobre, C., & Xhafa, F. (2014). Intelligent services for Big Data Science. *Future Generation Computer Systems*, 37, 267 – 281.

Drosio, S., & Stanek, S. (2016). The Big Data concept as a contributor of added value to crisis decision support systems. *Journal of Decision Systems*, 25, 228–239.

Du, Y., Liu, Q., Liu, T., Tian, X., Wang, J., Wang, T., Yang, Y., Yang, Z., Zhou, L., & Zhu, M. (2020). *Digital management method for China-hemp planting* (CN Patent No. CN111652090A).

<https://worldwide.espacenet.com/patent/search/family/072352300/publication/CN111652090A?q=CN111652090A>

Dutta, D. & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165, 293 – 306.

Elgendi, N. & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, 100, 1071 – 1084.

Feng, X. (2018). *Remote assistant guidance and control system for enterprise services* (CN Patent No. CN107748764A).

<https://worldwide.espacenet.com/patent/search/family/061255161/publication/CN107748764A?q=CN107748764A>

8764A?q=CN107748764A

- Ferrag, M. A., Derdour, M., Mukherjee, M., Derhab, A., Maglaras, L., & Janicke, H. (2018). Blockchain Technologies for the Internet of Things: Research Issues and Challenges. *IEEE Internet of Things Journal*, 1–1. <http://dx.doi.org/10.1109/jiot.2018.2882794>
- Feuerriegel, S. (2016). Decision support in healthcare: determining provider influence on treatment outcomes with robust risk adjustment. *Journal of Decision Systems*, 25(4), 371–390.
- Franková, P., Drahošová, M., & Balco, P. (2016). Agile Project Management Approach and its Use in Big Data Management. *Procedia Computer Science*, 83, 576-583. <http://dx.doi.org/10.1016/j.procs.2016.04.272>
- Fredriksson, C. (2018). Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support. *Journal of Decision Systems*, 27(1), 1–18.
- Gantz, J., & Reinsel, D. (2011). *Extracting Value from Chaos*. IDC iview.
- Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in Big Data analytics projects. *2015 Americas Conference on Information Systems, AMCIS 2015*, 1–14.
- Gao, W., Zhang, Q., Lu, Z., Wu, D., & Du, X. (2018). Modelling and application of fuzzy adaptive minimum spanning tree in tourism agglomeration area division. *Knowledge-Based Systems*, 143, 317–326.
- Gartner. (2019). Our Top Data and Analytics Predicts for 2019. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/
- Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, 265(3), 993–1004.
- Giannakis, M., & Louis, M. (2016). A multi-agent based system with big data processing for enhanced supply chain agility. *Journal of Enterprise Information Management*, 29(5), 706–727.
- Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, 72, 306–312.
- Gillingham, P. (2019a). Decision Support Systems, Social Justice and Algorithmic Accountability in Social Work: A New Challenge. *Practice*, 1–14.
- Gillingham, P. (2019b). *Can Predictive Algorithms Assist Decision-Making in Social Work with Children and Families?* Child Abuse Review.
- Ghasemaghaei, M., & Calic, G. (2019). Can big data improve firm decision quality? The role

- of data quality and data diagnosticity. *Decision Support Systems*, 120, 38-49.
- Goh, W. P., Tao, X., Zhang, J., & Yong, J. (2016). Decision support systems for adoption in dental clinics: A survey. *Knowledge-Based Systems*, 104, 195–206.
- Groves, W., Collins, J., Gini, M., & Ketter, W. (2014). Agent-assisted supply chain management: Analysis and lessons learned. *Decision Support Systems*, 57, 274–284.
- Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- Gupta, S., Chen, H., Hazen, B. T., Kaur, S., & Gonzalez, E. D. R. S. (2018). Circular economy and big data analytics: A stakeholder perspective. *Technological Forecasting and Social Change*. <https://doi.org/10.1016/j.techfore.2018.06.030>
- Hao, S., Yang, L., & Shi, Y. (2018). Data-driven car-following model based on rough set theory. *IET Intelligent Transport Systems*, 12(1), 49–57.
- Hayashi, Y. (2016). Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. *Operations Research Perspectives*, 3, 32–42.
- Hsu, C.-Y., Lim, S. S., & Yang, C.-S. (2017). Data mining for enhanced driving effectiveness: an eco-driving behaviour analysis model for better driving decisions. *International Journal of Production Research*, 55(23), 7096–7109.
- Huang, S., & Chaovalltwongse, W. A. (2015). Computational optimization and statistical methods for big data analytics: Applications in neuroimaging. *INFORMS Tutorials in Operations Research*. <https://doi.org/10.1287/educ.2015.0135>.
- Hurter, C., Conversy, S., Gianazza, D., & Telea, A. C. (2014). Interactive image-based information visualization for aircraft trajectory analysis. *Transportation Research Part C: Emerging Technologies*, 47, 207–227.
- Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P., & Monostori, L. (2015). Advanced predictive-analysis-based decision support for collaborative logistics networks. *Supply Chain Management: An International Journal*, 20(4), 369–388.
- Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núñez, A., Dollevoet, R., De Schutter, B. & Li, Z. (2018). A decision support approach for condition-based maintenance of rails based on big data analysis. *Transportation Research Part C: Emerging Technologies*, 95, 185–206.
- Janković, S., Mladenović, S., Mladenović, D., Vesović, S., & Glavić, D. (2018). Schema on read modeling approach as a basis of big data analytics integration in EIS. *Enterprise Information Systems*, 1–22.
- Ji, C. Y., Jean, K. Y., Tae, K. Y., Cheol, L. S., & Dam, Y. S. (2020). *Consensus algorithm of blockchain cryptographic technique and evolution network in blockchain* (KR Patent No. KR 1020200000000).

- KR20200045089A).
<https://worldwide.espacenet.com/patent/search/family/070732495/publication/KR2020045089A?q=KR20200045089A>
- Jianmin, W., Wenbin, Z., Tongrang, F., Shilong, Y., & Hongwei, L. (2020). An improved join-free snowflake schema for ETL and OLAP of data warehouse. *Concurrency and Computation: Practice and Experience*. <https://doi.org/10.1002/cpe.5519>
- Jin, D.-H., & Kim, H.-J. (2018). Integrated Understanding of Big Data, Big Data Analysis, and Business Intelligence: A Case Study of Logistics. *Sustainability*, 10(10), 3778.
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2, 59 – 64.
- Jukic, N., Jukic, B., Sharma, A., Nestorov, S., & Arnold, B. K. (2017). Expediting analytical databases with columnar approach. *Decision Support Systems*, 95, 61–81.
- Kaisler, S., Armour, F., Espinosa, J. A., Money, W. (2013) Big data: issues and challenges moving forward. In: *System Sciences (HICSS) 44th Hawaii International Conference on System Sciences*, 995–1004.
- Kappelman, L., McKeeman, R., & Zhang, L. (2007). Early Warning Signs of it Project Failure: The Dominant Dozen. *EDPACS*, 35(1), 1–10.
<https://doi.org/10.1080/07366980701238939>
- Khan, U., Aadil, F., Ghazanfar, M. A., Khan, S., Metawa, N., Muhammad, K., Mehmood, I. &, Nam, Y. (2018). A Robust Regression-Based Stock Exchange Forecasting and Determination of Correlation Between Stock Markets. *Sustainability*, 10(10), 3702.
- Khan, N., Naim, A., Hussain, M. R., Naveed, Q. N., Ahmad, N., & Qamar, S. (2019). The 51 V's Of Big Data. *Proceedings of the International Conference on Omni-Layer Intelligent Systems - COINS '19*. <https://doi.org/10.1145/3312614.3312623>
- Khansa, L., Forcade, J., Nambari, G., Parasuraman, S., & Cox, P. (2012). Proposing an Intelligent Cloud-Based Electronic Health Record System. *International Journal of Business Data Communications and Networking*, 8(3), 57–71.
- Kim, B. E. (2020). *Method and system for managing, on basis of Big Data, project results created by expert* (WO Patent No. WO2020175753A1).
<https://worldwide.espacenet.com/patent/search/family/068542153/publication/WO2020175753A1?q=WO2020175753A1>
- Lan, S., Zhang, H., Zhong, R. Y., & Huang, G. Q. (2016). A customer satisfaction evaluation model for logistics services using fuzzy analytic hierarchy process. *Industrial Management & Data Systems*, 116(5), 1024–1042.
- Lara, J. A., De Sojo, A. A., Aljawarneh, S., Schumaker, R. P., & Al-Shargabi, B. (2020). Developing Big Data Projects in Open University Engineering Courses: Lessons Learned. *IEEE Access*, 8, 22988–23001. <https://doi.org/10.1109/access.2020.2968969>

- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21.
- Lee, H., Aydin, N., Choi, Y., Lekhavat, S., & Irani, Z. (2018). A decision support system for vessel speed decision in maritime logistics using weather archive big data. *Computers & Operations Research*, 98, 330–342.
- Li, J., Li, X., & Zhu, B. (2016). User opinion classification in social media: A global consistency maximization approach. *Information & Management*, 53(8), 987–996.
- Li, F., Yang, J., Jin, C., & Guo, C. (2017). A new effect-based roughness measure for attribute reduction in information system. *Information Sciences*, 378, 348–362.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big Data in tourism research: A literature review. *Tourism Management*, 68, 301–323.
<https://doi.org/10.1016/j.tourman.2018.03.009>
- Liu, Z., Wang, Q & Wu, J. (2020). Mini-GD method based on GBP dimension (CN Patent No. CN111581524A).
<https://worldwide.espacenet.com/patent/search/family/072127637/publication/CN111581524A?q=CN111581524A>
- Long, Q. (2017). A framework for data-driven computational experiments of inter-organizational collaborations in supply chain networks. *Information Sciences*, 399, 43–63.
- Ludwig, N., Feuerriegel, S., & Neumann, D. (2015). Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems*, 24(1), 19–36.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011) *Big Data: The Next Frontier For Innovation, Competition, And Productivity*. Mckinsey Global Institute: Washington, Dc, Usa.
- Marr, B. (2015). *Big Data: Too Many Answers, Not Enough Questions*. Forbes.
- McAfee, A., & Brynjolfsson, E. (2012). *Big data: the management revolution*. Harvard business review, 90(10), 60-68.
- Mello, R., & Martins, R. A. (2019). Can big data analytics enhance performance measurement systems? *IEEE Engineering Management Review*, 1–1.
- Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771–785.
- Mielli, F., & Bulanda, N. (2019). Digital Transformation: Why Projects Fail, Potential Best Practices and Successful Initiatives. *2019 IEEE-IAS/PCA Cement Industry Conference (IAS/PCA)*. <https://doi.org/10.1109/citcon.2019.8729105>
- Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of “unstructured”

- decision processes. *Administrative Science Quarterly*, 21(2), 246–275. <https://doi.org/10.2307/2392045>
- Moore J. (2017). Data visualization in support of executive decision making. *Interdisciplinary Journal of Information, Knowledge, and Management*, 12, 125-138.
- Mortenson, M. J., Doherty, N. F. & Robinson, S. (2015), Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3), 583–595.
- Mousanif, H., Sabah, H., Douiji, Y., & Sayad, Y. O. (2014). From Big Data to Big Projects: a Step-by-step Roadmap. In: *2014 International Conference on Future Internet of Things and Cloud*. <https://doi.org/10.1109/FiCloud.2014.66>
- Mousannif, H., Sabah, H., Douiji, Y., & Sayad, Y. O. (2016). Big data projects: just jump right in! *International Journal of Pervasive Computing and Communications*, 12(2), 260–288. <https://doi.org/10.1108/ijpcc-04-2016-0023>
- Murdoch, T. B., & Detsky, A. S. (2013). The Inevitable Application of Big Data to Health Care. *JAMA*, 309(13), 1351-1352.
- Neaga, I., Liu, S., Xu, L., Chen, H., & Hao, Y. (2015). Cloud Enabled Big Data Business Platform for Logistics Services: A Research and Development Agenda. *Lecture Notes in Business Information Processing*, 22–33.
- Nguyen, T., Larsen, M. E., O'Dea, B., Nguyen, D. T., Yearwood, J., Phung, D., Venkatesh, S. & Christensen, H. (2017). Kernel-based features for predicting population health indices from geocoded social media data. *Decision Support Systems*, 102, 22–31.
- Nimmagadda, S. L., Reiners, T., & Wood, L. C. (2018). On big data-guided upstream business research and its knowledge management. *Journal of Business Research*, 89, 143–158.
- Oliveira, M., Freitas, H. (1998). Focus group, pesquisa qualitativa: resgatando a teoria, instrumentalizando o seu planejamento. *Revista de Administração da Universidade de São Paulo*, 33(3), 83-91.
- Osuszek, L., Stanek, S., & Twardowski, Z. (2016). Leverage big data analytics for dynamic informed decisions with advanced case management. *Journal of Decision Systems*, 25, 436–449.
- Pan, M. & Rui, Q. (2020). *Chemical process modeling processing system based on industrial big data and industrial Internet of Things* (CN Patent No. CN111338302A). <https://worldwide.espacenet.com/patent/search/family/071183939/publication/CN111338302A?q=CN111338302A>
- Papagiannidis, S., See-To, E. W. K., Assimakopoulos, D. G., & Yang, Y. (2018). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age? *Computers & Operations Research*, 98, 355–366.

- Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research*, 252(2), 687–698.
- Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
- Pettit, C., Bakelman, A., Lieske, S. N., Glackin, S., Thomson, G., Shearer, H., ... & Newman, P. (2018). Planning support systems for smart cities. *City, culture and society*, 12, 13-24.
- Poleto, T., Carvalho, V. D. H. de, & Costa, A. P. C. S. (2017). The Full Knowledge of Big Data in the Integration of Inter-Organizational Information: An Approach Focused on Decision Making. *International Journal of Decision Support System Technology*, 9(1), p. 16 – 31.
- Pondel, J., & Pondel, M. (2016). The Concept of Project Management Platform using BI and Big Data Technology. In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016)*, 1, 166-173.
- Power, D. J., Cyphert, D., & Roth, R. M. (2019). Analytics, bias, and evidence: the quest for rational decision making. Special issue on cognitive bias. *Journal of Decision Systems*, 1–18.
- Power, D. J. (2016). “Big Brother” can watch us. *Journal of Decision Systems*, 25, 578 – 588.
- Rasiulis, R., Ustinovichius, L., Vilutienė, T., & Popov, V. (2016). Decision model for selection of modernization measures: public building case. *Journal of Civil Engineering and Management*, 22(1), 124–133.
- Rönnqvist, M., Svenson, G., Flisberg, P., & Jönsson, L.-E. (2017). Calibrated Route Finder: Improving the Safety, Environmental Consciousness, and Cost Effectiveness of Truck Routing in Sweden. *Interfaces*, 47(5), 372–395.
- Russom, P. (2011). *Big data analytics*. Tdwi Best Practices Report, Fourth Quarter.
- Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00287-9>
- Saldaña, J. (2012). *The Coding Manual for Qualitative Researchers*. 2nd ed. SAGE Publications.
- Saltz, J. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In: *3rd IEEE International Conference on Big Data*.
- Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project’s success. *2016 IEEE International Conference on Big Data (Big Data)*. <https://doi.org/10.1109/bigdata.2016.7840936>

- Saltz, J., Shamshurin, I., & Connors, C. (2017). A Framework for Describing Big Data Projects. *In: International Conference on Business Information Systems*, 183 – 195.
- Salviati, M. E. 2017. Manual do Aplicativo Iramuteq (versão 0.7 Alpha 2 e R Versão 3.2.3). Retrieved from <http://iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>
- Sathiaraj, D., Punkasem, T., Wang, F., & Seedah, D. P. K. (2018). *Data-driven analysis on the effects of extreme weather elements on traffic volume in Atlanta, GA, USA*. Computers, Environment and Urban Systems.
- Sasson, E., Ravid, G., & Pliskin, N. (2015). Improving similarity measures of relatedness proximity: Toward augmented concept maps. *Journal of Informetrics*, 9(3), 618–628.
- Schnase, J. L., Duffy, D. Q., Tamkin, G. S., Nadeau, D., Thompson, J. H., Grieg, C. M., McInerney, M. A. & Webster, W. P. (2014). MERRA Analytic Services: Meeting the Big Data challenges of climate science through cloud-enabled Climate Analytics-as-a-Service. *Computers, Environment and Urban Systems*, 61, 198–211.
- Schneider, D., & Seelmeyer, U. (2019). Challenges in Using Big Data to Develop Decision Support Systems for Social Work in Germany. *Journal of Technology in Human Services*, 1–16.
- Seles, B. M. R. P., Jabbour, A. B. L de S., Jabbour, C. J. C., Fiorini, P. de C., Mohd-Yusoff, Y., & Thomé, A. M. T. (2018). Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for Big Data management towards a low carbon society. *Journal of Cleaner Production*, 189, 763–774. <https://doi.org/10.1016/j.jclepro.2018.04.113>
- Semanjski, I., Bellens, R., Gautama, S., & Witlox, F. (2016). Integrating Big Data into a Sustainable Mobility Policy 2.0 Planning Support System. *Sustainability*, 8(11), 1142.
- Silva, L. F. da, Russo, R. de F. S. M., & Oliveira, P. S. G. de. (2018). Quantitativa ou qualitativa? Um alinhamento entre pesquisa, pesquisador e achados em pesquisas sociais. *Revista Pretexto*, 19(4), 30-45.
- Silva, A. B., Godoi C. K., & Melo, R. B. (2010). *Pesquisa Qualitativa em Estudos Organizacionais*. São Paulo: Saraiva.
- Simon, H. A. (1960). *The executive as decision maker*. The new science of management decision, 1-7.
- Sivamani, S., Choi, J., & Cho, Y. (2017). A service model for nutrition supplement prediction based on Fuzzy Bayes model using bigdata in livestock. *Annals of Operations Research*, 265(2), 257–268.
- Shrestha, A., Cater-Steel, A., & Toleman, M. (2016). Innovative decision support for IT service management. *Journal of Decision Systems*, 25, 486–499.
- Stein, N., Meller, J., & Flath, C. M. (2018). Big data on the shop-floor: sensor-based decision-

- support for manual processes. *Journal of Business Economics*, 88(5), 593–616.
- Stewart, D. W., & Shamdasani, P. (2016). Online Focus Groups. *Journal of Advertising*, 46(1), 48–60. <https://doi.org/10.1080/00913367.2016.1252288>
- Suh, J. H. (2019). SocialTERM-Extractor: Identifying and Predicting Social-Problem-Specific Key Noun Terms from a Large Number of Online News Articles Using Text Mining and Machine Learning Techniques. *Sustainability*, 11(1).
- Surbakti, F. P. S. (2021). What is Effective Use of Big Data? The Consensual Definition of Effective Use of Big Data. *IOP Conf. Series: Materials Science and Engineering*. <https://doi.org/10.1088/1757-899X/847/1/012003>
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2019). Factors Influencing Effective Use of Big Data: A Research Framework. *Information & Management*. <https://doi.org/10.1016/j.im.2019.02.001>
- Tashkandi, A., Wiese, I., & Wiese, L. (2018). Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems. *Big data research*, 13, 52–64.
- Teixeira, A. A. C., & Queirós, A. S. S. (2016). Economic growth, human capital and structural change: A dynamic panel data analysis. *Research Policy*, 45(8), 1636–1648. <https://doi.org/10.1016/j.respol.2016.04.006>
- Tian, J., Zhang, Y., & Zhang, C. (2018). Predicting consumer variety-seeking through weather data analytics. *Electronic Commerce Research and Applications*, 28, 194–207.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207-222.
- Turner, D. W. (2010). Qualitative Interview Design: A Practical Guide for Novice Investigators. *The Qualitative Report*, 15(3).
- Venturebeat (2019). *Why do 87% of data science projects never make it into production?* <https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/>
- Vera-Baquero, A., Colomo-Palacios, R., Molloy, O. &, Elbattah, M. (2015). Business process improvement by means of Big Data based Decision Support Systems: a case study on Call Centers. *International Journal of Information Systems and Project Management*, 3(1), 5-26.
- Wamba, S. F., Ngai, E.W.T., Riggins, F.J. & Akter, S. (2017). Guest editorial: transforming operations and production management using big data and business analytics: future research directions. *International Journal of Operations and Production Management*, 37(1), 2–9.
- Wang, W. (2020). *Engineering cost data collection method and system based on big data*,

- computer equipment and computer readable storage medium* (CN Patent Number CN111444446A).
<https://worldwide.espacenet.com/patent/search/family/071649085/publication/CN1114446A?q=CN111444446A>
- Wang, B. (2019). System for effectively improving output result data quality in big data processing process and control method thereof (CN Patent No. CN110347741A).
<https://worldwide.espacenet.com/patent/search/family/068178762/publication/CN110347741A?q=CN110347741A>
- Wang, P. & Xu, Y. (2016). *Water conservancy big data service analysis and evaluation model* (CN Patent No. CN106096733A).
<https://worldwide.espacenet.com/patent/search/family/057252185/publication/CN106096733A?q=CN106096733A>
- Wang, Y., Geng, S., & Gao, H. (2018). A proactive decision support method based on deep reinforcement learning and state partition. *Knowledge-Based Systems*, 143, 248–258.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.
- Wilkin, C., Ferreira, A., Rotaru, K., & Gaerlan, L. R. (2020). Big data prioritization in SCM decision-making: Its role and performance implications. *International Journal of Accounting Information Systems*, 100470. <https://doi.org/10.1016/j.accinf.2020.100470>
- Wu, X., Zhu, X., Wu, G., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97–107.
<https://doi.org/10.1109/TKDE.2013.109>
- Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2016). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13–53. <https://doi.org/10.1080/17538947.2016.1239771>
- Yang, Z., Zhang, Z., Zhang, S., Wang, J., Lin, H., & Zeng, B. (2017). Patent abstract analysis on Chinese big data. *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*.
<https://doi.org/10.1109/fskd.2017.8393098>
- Zhao, Z. (2017). Application extension method and system of Apache NiFi in aspect of importing source data into database (CN Patent No. CN107330028A).
<https://worldwide.espacenet.com/patent/search/family/060194770/publication/CN107330028A?q=CN107330028A>
- Zwaanswijk, M., & van Dulmen, S. (2014). Advantages of asynchronous online focus groups and face-to-face focus groups as perceived by child, adolescent and adult participants: a survey study. *BMC Research Notes*, 7(1), 756. <https://doi.org/10.1186/1756-0500-7-756>

APÊNDICE A

The current issue and full text archive of this journal is available on Emerald Insight at:
<https://www.emerald.com/insight/2177-8736.htm>

Big data as a value generator in decision support systems: a literature review

Gustavo Grander and Luciano Ferreira da Silva
PPGP, Universidade Nove de Julho, São Paulo, Brazil, and
 Ernesto Del Rosário Santibañez Gonzalez
Universidad de Talca, Talca, Chile

Big data as a
 value
 generator in
 DSS

Received 10 March 2020
 Revised 15 October 2020
 12 January 2021
 19 March 2021
 Accepted 14 May 2021

Abstract

Purpose – This paper aims to analyze how decision support systems manage Big data to obtain value.
Design/methodology/approach – A systematic literature review was performed with screening and analysis of 72 articles published between 2012 and 2019.

Findings – The findings reveal that techniques of big data analytics, machine learning algorithms and technologies predominantly related to computer science and cloud computing are used on decision support systems. Another finding was that the main areas that these techniques and technologies are been applied are logistic, traffic, health, business and market. This article also allows authors to understand the relationship in which descriptive, predictive and prescriptive analyses are used according to an inverse relationship of complexity in data analysis and the need for human decision-making.

Originality/value – As it is an emerging theme, this study seeks to present an overview of the techniques and technologies that are being discussed in the literature to solve problems in their respective areas, as a form of theoretical contribution. The authors also understand that there is a practical contribution to the maturity of the discussion and with reflections even presented as suggestions for future research, such as the ethical discussion. This study's descriptive classification can also serve as a guide for new researchers who seek to understand the research involving decision support systems and big data to gain value in our society.

Keywords Decision support systems, Big data, Machine learning, Analytics, Cloud computing, Algorithm
Paper type Literature review

1. Introduction

While Big Data can be used as a powerful tool to treat various social diseases, offering the potential for new insights into areas such as medical research, counterterrorism and climate change, its use also allows invasions of privacy, diminished civil liberties and increased state and corporate control (Boyd & Crawford, 2012). There is a big challenge in managing Big data due to the increasing and cheap volume of data storage (Demirkhan & Delen, 2013).

Most data sets from which scientists and researchers have been able to extract real meaning are still very small compared to the proportion of data that can be captured (Dobre & Xhafa, 2014). Big data analysis should include the phases of data generation, acquisition, storage and analysis and can provide useful values at each stage through judgments, suggestions, support or decisions (Chen, Mao & Liu, 2014a). Some authors point out that

© Gustavo Grander, Luciano Ferreira da Silva and Ernesto Del Rosário Santibañez Gonzalez. Published in *Revista de Gestão*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).



Revista de Gestão
 Emerald Publishing Limited
 e-ISSN: 2177-8736
 p-ISSN: 1809-2276
 DOI: 10.1108/REG-03-2020-0014

REGE

advances in analytical techniques, especially machine learning, have been a major facilitator for dealing with large data set analysis (Murdoch & Detsky, 2013).

Big data are defined as a large set of data that are difficult to store, process analyze and understand using traditional database processing tools (Huang & Chaovalltwongse, 2015), Big data emerge as a paradigm shift in how organizations make decisions (Mortenson, Doherty & Robinson, 2015). Therefore, through decision support systems (DSS), it is possible to process large volumes of data using output models with accessible interfaces (Constantiou & Kallinikos, 2014). Research in big data and DSS has presented technological aspects and big data design challenges as the main focus (Chen, Mao, Zhang & Leung, 2014b).

We, therefore, have the opportunity to propose an in-depth analysis of how DSS manages big data to obtain value. In this context, we performed a systematic literature review (SLR) of the use of big data in DSS. To achieve this goal, we propose to answer two questions: DSS manages big data to obtain value which techniques and technologies? And, DSS has been applied to solve what types of problems? To answer the research questions, we conducted an SLR with research in the academic databases Scopus and Web of Science (WoS), in August 2019.

As a contribution of our study, we show that DSS is used for the management of big data mainly through techniques such as big data analytics (BDA), machine learning algorithms and technologies such as cloud computing. The main areas in which these techniques and technologies are applied are logistics, traffic, health, organization and market.

This article is structured as follows: In addition to this Introduction section, we present in Section 2 the methodological procedures applied in the research, the results are presented in Section 3, discussions in Section 4 and in Section 5, we present the conclusion of our study.

2. Methodology

We performed the SLR based on the guidelines suggested by Petticrew & Roberts (2006) to understand the use of big data in DSS. We have chosen to develop an SLR because of the rigor of the research methodology that involves systematic data collection procedures, descriptive and qualitative data analysis techniques.

2.1 Data collection

First, we start with the identification phase with the application of a search string, initially based on preliminary searches. Then, we apply the string to the database search tools in order to identify other ways in which the searched terms are referenced (Petticrew & Roberts, 2006). Thus, after a few rounds of searching, new terms were identified and added to the initial string to obtain a corpus of analysis. Therefore, the first stage of the SLR consisted of searching the Scopus and WoS databases, using the following Boolean terms: "Big Data" AND ("Decision Theory" or "decision support system*" or "decision-support system*"). The use of the asterisk (*) in the search string serves to obtain variations of words in their plural form.

The second stage of the SLR had as criteria the filter by areas. The Scopus database was restricted to the following areas: business, management and accounting; decision sciences; social sciences. The WoS base was restricted to the following areas: business, management and social science interdisciplinary.

The third stage of the SLR consisted of filtering the databases keeping only articles, thereby excluding books, reviews, etc. The fourth step of SRL consisted of checking the availability of downloadable articles. The fifth stage of the SLR was reading the articles and checking their alignment with the research objective; articles without objective strongly dealing with the subject were considered inappropriate for the analysis, and therefore, they

were excluded from the sample. The sixth stage of the SLR consisted of identifying repeated articles and consequently reducing duplication. Finally, the seventh stage of SLR was the consolidation of the two bases for an in-depth analysis of the articles. The exclusion flow of articles is presented in [Table 1](#).

The initial search totaled 1,427 documents contained in both databases, and at the end of the seven steps, 5% of the sample, 72 articles, were considered adequate for full reading and analysis.

2.2 Data analysis

We performed the SLR of the 72 selected articles, of which 63 were empirical and 9 were theoretical. The articles were read in full and classified in a spreadsheet according to relationships identified throughout the analysis and that met the research questions. We note that the first record refers to the year 2012, and until 2018, there is a growth in the number of publications. The year 2019 had eight publications, but due to the date of the search in the databases, it was not possible to account for the total number of publications in the year ([Table 2](#)).

Regarding the analysis process, after selecting and collecting the database, the researchers started categorizing the contents from a qualitative perspective ([Pettricrew & Roberts, 2008](#)). A recursive process was applied based on reflective critical reading and classification of the content according to its adherence to the themes proposed in this research. The process of an SLR must allow its transparency and replicability as pointed out by [Tranfield, Denyer & Smart \(2003\)](#). The following are the results of this study and their respective analyzes.

3. Results

To answer the first research question, we classified the articles into two groups: The first group deals with techniques, and the second group with technologies. There is a wide variety of techniques and technologies for capturing, selecting, analyzing and visualizing big data, and these tools focus on three classes: batch processing, flow processing and interactive analysis tools ([Chen & Zhang, 2014](#)).

3.1 Techniques applied in big data

The items presented here are related to the application of data analysis techniques to obtain value from big data.

Stage	Criteria	Scopus	WoS
1	Survey string search	1,087 documents	340 documents
2	Filter by area	257 documents	25 documents
3	Filter by articles	98 articles	15 articles
4	Exclusion unavailable articles	88 articles	13 articles
5	Exclusion inappropriate articles	71 articles	10 articles
6	Articles in both bases	9 articles	
7	Total articles analyzed in the review	72 articles	

Table 1.
Exclusion flow of
articles

Year	2012	2013	2014	2015	2016	2017	2018	2019
Quantity of paper	1	1	3	8	14	14	23	8

Table 2.
Quantity of papers
according to years

REGE

3.1.1 Machine learning technique with supervised learning algorithm. Classification algorithms from this group were used, for example, as a management tool for the creation of a DSS that enabled credit risk assessment in the financial market (Hayashi, 2016), for building modernization assessment allowing the decision-maker to select the best alternatives in terms of energy consumption and installation cost (Rasiulis, Ustinovichius, Vilutiene & Popov, 2016), to deal with business data heterogeneity and multidimensionality (Nimmagadda, Reiners & Wood, 2018). We also found application in route optimization models through DSS to view traffic volume and weather interactions applicable to transport planners, traffic control rooms and urban infrastructure DSS (Sathiaraj, Punkasem, Wang & Seedah, 2018) and through an online route generation system with Dijkstra algorithm, which resulted in changes of route paradigms that were determined together (Rönnqvist, Svenson, Flisberg & Jö). Finally, we found an application for the purpose of identifying and predicting social issues through online news analytics (Suh, 2019).

Other findings were with vector machine for decision-making optimization through patient similarity (Tashkandi, Wiese & Wiese, 2018) and in optimizing data collection for cancer classification and online critic sentiments (Ghaddar & Naoum-Sawaya, 2018), the fuzzy rule to determine the health status of cattle to predict nutritional intake (Sivamani, Choi & Cho, 2018), Mehrabian–Russell model for forecasting consumer purchases using climate parameters (Tian, Zhang & Zhang, 2018), the use of a set of attribute reduction and data set analysis in information systems (Li, Yang, Jin & Guo, 2017) and rapid safety feedback used to identify child maltreatment (Gillingham, 2019a).

We have identified that techniques applied in the above studies were for predictive data analysis, which helps anticipate changes based on understanding patterns and anomalies within a database (Hurwitz & Kirsch, 2018).

The second group of articles in our technique analysis was linear regression that has been observed in stock market index prediction studies (Khan *et al.*, 2018), in DSS for collaborative logistics networks (Ilie-Zudor *et al.*, 2015) and in predicting fuel consumption based on driving behaviors (Hsu, Lim & Yang, 2017). The least absolute shrinkage operator regression was used to estimate risk-adjusted performance in hospitals (Feuerriegel, 2016), for forecasting electricity prices based on historical data and analysis of weather conditions (Ludwig, Feuerriegel & Neumann, 2015) and for predicting population health indices from social media data to improve predictive performance (Nguyen *et al.*, 2017).

Regression of neural networks was applied to estimate rail integrity conditions to assist in maintenance control (Jamshidi *et al.*, 2018), through regression for multipurpose utility analysis, a methodology was developed to analyze resource scarcity problems in rural transport management (Chen, Achtari, Majkut & Sheu, 2017), and a regression tree model was used to quantify and optimize leak detection on the shop floor (Stein, Meller & Flath, 2018).

In this group of articles, we note the predominance of predictive data analysis. However, one article analyzed data also with prescriptive analysis (Stein *et al.*, 2018), and in which case, the results obtained by prescriptive analysis stood out in relation to predictive results. Prescriptive analyses are required to determine and evaluate alternative decisions involving objectives and requirements characterized by high volume and complexity (Wang, Gunasekaran, Ngai & Papadopoulos, 2016).

Based on these studies and on the condition of predicting situations, the value built from the use of big data in decision-making is the ability of decision-makers to allocate resources, make adjustments in their process or even to direct assertive actions to deal with child abuse as in case of Gillingham's (2019a) study. The ability to relate several factors from the big data analysis allows improving decision-making considerably; in this group, the gains also come from saving resources that can be applied in certain activities, for example, in the measurement of risks.

3.1.2 Machine learning technique with unsupervised learning algorithm. We have identified the application of cluster learning with image-based information visualization technique for

analyzing emerging data sets from the air traffic control domain (Hurter, Conversy, Gianazza & Telea, 2014), and hierarchical clustering technique used as DSS to identify near real-time industrial clusters (Papagiannidis, See-To, Assimakopoulos & Yang, 2018).

We also identified articles developed through the diffuse adaptive minimum spanning tree method that offered opportunities for the development of intelligent tourism DSS for government policymaking (Gao, Zhang, Lu, Wu & Du, 2018), P-DBSCAN algorithm that identifies areas of high photographic activity, using the information on the number of tourists providing photos and the number of photos sent to find groups at popular tourist sites (Miah, Vu, Gammack & McGrath, 2017) and user behavior map analysis through image recognition algorithms that enable tourism scenario formulation conditions for marketing strategies (Giglio, Bertacchini, Bilotta & Pantano, 2019).

In this group, again predictive analyzes formed the majority of articles, however, Hurter *et al.* (2014) conducted a study through descriptive data analysis. Descriptive analyses help to understand the historical data set and then to understand the current reality (Hurwitz & Kirsch, 2018).

Another identified group was unsupervised association learning. We have found the use of transmissible comorbidity map for patient monitoring according to risk profiles (Capobianco & Liò, 2015) and the use of sentiment analysis mechanisms that benefit from grammar-based linguistic analysis to predict stock market index trends (Chan & Chong, 2017). Both articles appropriated the decision tree as a technique and treated the data for predictive analysis.

This group of techniques and applications allows for real-time monitoring, which brings benefits and diagnostic capacity. Based on this application, decision-makers can have more agility in raising problems and opportunities organically and evolutionarily.

3.1.3 Machine learning technique with semi-supervised learning algorithm. We have only identified one article that used semi-supervised classification learning as a technique to address a Twitter user opinion rating issue in discussions from other people's retweets, using the label propagation technique for predictive analytics (Li, Li & Zhu, 2016).

3.1.4 Machine learning techniques with reinforcement learning algorithms. Studies classified as reinforcement learning have in common the traffic optimization ratio to avoid congestion during the journey. A car tracking model based on approximate set theory data was used to consider the hidden information in a field data set (Hao, Yang & Shi, 2018), and a proactive method of controlling traffic congestion with the use of deep belief network has been applied to eliminate unwanted future states (Wang, Geng & Gao, 2018b). Both studies presented a predictive analysis of data.

3.1.5 Data mining techniques. The articles emphasized the added value resulting from the use of big data in the treatment of heterogeneous data sources accessed by crisis management frameworks (Drosio & Stanek, 2016). The perception of the use of data mining techniques as a strategic management tool in the banking sector from an accounting and financial perspective was also explored (Al Chahadah, El Refae & Qasim, 2018). A co-word analysis framework model has been proposed that encompasses and synthesizes unstructured textual data from web sources to improve decision-making processes to gain and maintain competitive advantage (Sasson, Ravid & Pliskin, 2015).

In the area of health, barriers to DSS adopted by dentists were identified due to their perceived lack of utility, difficulties arising from social and economic factors and in interpreting information from the used systems (Goh, Tao, Zhang & Yong, 2016). In the construction industry sector, it was discussed how the key challenges of data mining from the nature of fragmented data require cultural change and more structured collection systems (Ahmed, Aziz, Tezel & Riaz, 2018).

Data mining techniques using the Bresenham algorithm was used to identify the impact of weather conditions on a given travel route, focusing on fuel cost savings (Lee, Aydin, Choi,

REGE

Lekhavat & Irani, 2018). And a combination of flow mining algorithms and online discovery of declaration models, and a model of declarative processes for the transmission of event-producing data was applied in an academic hospital to cancer patients diagnosed by updated images as behavior runtime (Burattin, Cimitile, Maggi & Sperduti, 2015).

The use of data from different sources and in different types is a great challenge for decision-makers. In this research, we can perceive possibilities such as those mentioned above that allow considerable gains for public and private agents in the process of diagnosis and construction of strategies to deal with problems. Thus, when compiling the data in a single system, after collection and treatment, professionals who work with big data can provide inputs for various activities in society.

3.1.6 Big data analytics (BDA) techniques. The health area has been a fertile field for the application of big data (Ashrafi, Kuilboer, Joshi, Ran & Pande, 2014). And the healthcare industry has yet to fully understand the benefits of using BDA (Wang, Kung & Byrd, 2018a). There are still challenges in the adoption of DSS, such as data protection, prejudice and the probability of manipulation in the area of social work (Schneider & Seelmeyer, 2019).

In the governmental field, a paradoxical situation has been identified because on the one hand, we realize the creation of information and knowledge from big data analysis and its use as DSS support in organizations improving the quality and reliability of decisions (Fredriksson, 2018). On the other hand, the monopoly of government control of large amounts of data can have a dangerous influence on citizens for security reasons and/or unreasonable access (Power, 2016).

The use of big data mediated by data diagnosis did not significantly affect the quality of decisions in business studies (Ghasemaghaei & Calic, 2019). However, evidence-based decision-making must address a wide range of issues such as knowledge, belief, social factors and technical capabilities (Power, Cyphert & Roth, 2019). Therefore, we need to understand and present the data source and journey through the right visualization technologies and intuitive platforms for delivering successful messages (Moore, 2017), as the use of unstructured external data challenges companies as to the information accuracy mainly concerning performance measurements (Mello & Martins, 2019).

We highlight the limited availability of tools to provide decision support to improve service management processes (Shrestha, Cater-Steel & Toleman, 2016). However, methodologies and frameworks (Vera-Baquero, Colomo-Palacios, Molloy & Elbattah, 2015), integrated structures of data-driven computational experiments (Long, 2017) and multiagent technologies (Giannakis & Louis, 2016) are proposed to condition scalability and big data processing power. These technologies present themselves as promising alternatives to meet the needs of modern supply chain management at a level of collaboration between organizations and decentralizing operations.

Big Data have been widely applied in logistics, service, planning and manufacturing processes; however, supply chain executives still seem to have a slow adoption in relation to the real potential (Brinch, Stentoft, Jensen & Rajkumar, 2018). Note that when automating business processes, business agility should be considered especially about how the automated process will respond situations where market assumptions may be violated (Groves, Collins, Gini & Ketter, 2014). In general, articles using BDA techniques seek to provide solutions in complex, turbulent environments that often lack a centralized, standardized information collection framework, making it even more difficult for managers to manage data (Osuszek, Stanek & Twardowski, 2016).

3.2 Technologies for big data

The items presented here are related to the application of data analysis technologies to obtain value from big data.

Big data as a value generator in DSS

3.2.1 Cloud computing technology. We identified articles that were rated for using cloud computing technology. In order to benefit from the use of big data, companies can instead look for cloud computing for technical flexibility (Boutkhoum, Hanine, Agouti & Tilkiouine, 2016); economies of scale, scope and speed in service analysis (Demirkhan & Delen, 2013) and further reduce uncertainty and ambiguity in the classification of decision-makers (Boutkhoum, Hanine, Agouti & Tilkiouine, 2017). A big data logistics business platform for supply chain network management services has been proposed as a way to modernize logistics services (Neaga, Liu, Xu, Chen & Hao, 2015). In healthcare, an intelligent cloud-based electronic health record system has been developed to improve the privacy and security of electronic records, and it has enabled this application to reduce medical errors and improve patients' quality of life, as well as reducing costs and increasing productivity of health organizations (Khansa, Forcade, Nambari, Parasuraman & Cox, 2012). Finally, the term Climate Analytics-as-a-Service has been presented as a form of business where cloud computing plays an important role as it provides a mobility-driven way of working (Schnase *et al.*, 2017).

3.2.2 Technologies for business intelligence (BI). In our analysis, only two articles were identified with BI technology. The first study highlighted the importance of BI-based decision-making and how its use is critical to ensuring competitiveness and sustainable growth in enterprises (Jin & Kim, 2018). The second presented a prescriptive framework for prioritizing items in business analysis and applying them to the business environment (Pape, 2016).

3.2.3 Technologies for data warehouse. Janković, Mladenović, Mladenović, Vesković & Glavić (2018) present a flexible integration approach from heterogeneous external sources. Big data analysis has enabled integration into big data applications based on batch-oriented processing, as well as flexibility, reuse of raw data and query of various data types and repositories at once. Jukic, Jukic, Sharma, Nestorov & Arnold (2017) describe improvements in tactical decision-making quality that have also been found by providing critical information through analytical databases that combined columnar database technology with an approach based on denormalization of data tables for analysis and decision support. Almeida, Bernardino & Furtado (2015) performed the evaluation of - structured query language (SQL) and no structured query language (NoSQL) scalability platforms, where they concluded that the mechanism was considered more appropriate in transactional systems.

3.2.4 Systems technology. The last group of articles classified in this study was that with those papers that developed or used some information system as a technology application. Baechle, Agarwal & Zhu (2017) described the benefits of improved accuracy of services provided over the short term by applying large dictionaries of chronic obstructive pulmonary disease-related terms. Gillingham (2019b) presents through the principles of algorithmic responsibility, and combination with ethical codes of social work, the possibility for social workers to identify when wrong or biased recommendations are made by the deployed DSS and thereby develop defense strategies for social service users.

Pettit *et al.* (2018) present a tool called planning support system (PSS) used in a study related to smart cities, which assisted in the execution of the main tasks associated with the urban planning process, besides allowing better coordination between the planning and infrastructure agencies of the city, state and the country. Semanjski, Bellens, Gautama & Witlox (2016) demonstrate a case with the applicability of a big data integration approach to the smart city planning process, where the observed benefits were higher data punctuality, shorter cycle time and more informed and agile decision-making for both citizens and city planners. Deal, Pan, Pallathucheril & Fulton (2017) explored the potential of a new generation of conscious PSSs and the challenges of achieving big data.

We can point out from the analysis that as much as DSS are sophisticated tools used to optimize decision-making processes when used in high-pressure and turbulent environments,

REGE

they can have negative consequences (Aversa, Cabantous & Haefliger, 2018). Therefore, the use of techniques such as the diffuse analytical hierarchy process (Lan, Zhang, Zhong & Huang, 2016) ends up being a way to optimize the decision-making process in a system.

4. Discussion

Table 3 was structured to answer the first research question of this study. Thus, the application of big data and DSS is applied through techniques and technologies. We can infer that among the identified techniques, the use of BDA was the most, with 16 articles, representing 22% of the sample, followed by supervised classification learning, with 12 articles, which represented 17% of the entire sample analyzed. Regarding technologies, systems had the highest frequency, with seven articles, representing 10% of the sample, followed by cloud computing, with six articles, representing 8% of the sample.

We highlight that the articles in the years 2012 and 2013 used only technologies, and only from 2014, articles using big data analysis techniques were recorded in our sample. We also note that the growth in technology use occurred until 2016, with a decrease in subsequent years. As for the techniques, since the registration of the first article (2014), its use has grown steadily, thus characterizing the increasing applicability of techniques for this purpose. For this finding, we disregard the year 2019 since it was not a complete year for the collection of articles.

Table 3.
Techniques and
technologies used in
DSS to get value from
Big Data

Techniques	Articles
Supervised learning	Classification
	Hayashi (2016), Rasiulis <i>et al.</i> (2016), Rönnqvist <i>et al.</i> (2017), Li <i>et al.</i> (2017), Ghaddar & Naoum-Sawaya (2018), Nimmagadda <i>et al.</i> (2018), Sathiaraj <i>et al.</i> (2018), Sivamani <i>et al.</i> (2018), Tashkandi <i>et al.</i> (2018), Tian <i>et al.</i> (2018), Gillingham (2019a), Suh (2019)
Supervised learning	Regression
	Ilie-Zudor <i>et al.</i> (2015), Ludwig <i>et al.</i> (2015), Feuerriegel (2016), Chen <i>et al.</i> (2017), Hsu <i>et al.</i> (2017), Nguyen <i>et al.</i> (2017), Jamshidi <i>et al.</i> (2018), Khan <i>et al.</i> (2018), Stein <i>et al.</i> (2018)
Unsupervised learning	Cluster
Unsupervised learning	Association
Semi-supervised learning	Classification
Reinforced learning	Control
Data mining	
BDA	
Technologies	Articles
Cloud computing	Khansa <i>et al.</i> (2012), Demirkhan & Delen (2013), Neaga <i>et al.</i> (2015), Boutkhoum <i>et al.</i> (2016), Boutkhoum <i>et al.</i> (2017), Schnase <i>et al.</i> (2017)
Business intelligence	Pape (2016), Jin & Kim (2018)
Data warehouse	Almeida <i>et al.</i> (2015), Jukic, Jukic, Sharma, Nestorov & KorallusArnold (2017), Janković <i>et al.</i> (2018)
Information systems	Lan <i>et al.</i> (2016), Semanjski <i>et al.</i> (2016), Baechle <i>et al.</i> (2017), Deal <i>et al.</i> (2017), Pettit <i>et al.</i> (2018), Aversa <i>et al.</i> (2018), Gillingham (2019b)

BDA techniques have been widely used as a form of data analysis to interpret past behaviors and relationships. The amount of work using this technique shows us the importance and relevance that BDA has in environments that want to analyze large data sets, mainly for metrics management. Another technique that stood out for the high number of articles in our sample was supervised learning, specifically by classification and regression. Supervised learning techniques seek to define a function that can predict unknown labels from a previously labeled data set.

The low frequency recorded in techniques such as reinforced learning and semi-supervised learning does not necessarily indicate low interest in these applications by researchers. Reinforced learning, for example, has a very fertile field for applications where it is desired to have the least possible decision-making requirement of a person; with this, we understand that they are still new techniques but that have a great potential to be more and more used in our life.

[Table 4](#) was organized to answer the second research question of this study. We note that the issue of scalability in the use of decision systems based on big data is a major challenge. Otherwise, advances make it possible to perceive gains in areas such as healthcare, logistics and supply chain as can be seen in [Table 4](#).

When analyzing the percentage of each field, we noticed that articles directed to the logistics and traffic area had the highest rate, representing 19%; these studies had a predominant focus on the use of algorithms to optimize routes to reduce costs and improve performance. Health, the second group with the most articles (17%), had a predominant focus with studies of disease prediction through diagnoses and support in diagnoses. Business and market also representing 17% of the sample; they presented studies aimed at better decisions for business and social relations. Consideration should also be given to the predominance of studies directed to predictive analysis, that is, an analysis technique that helps to anticipate changes based on understanding patterns and anomalies within data ([Hurwitz & Kirsch, 2018](#)).

In line with the findings of [Li, Xu, Tang, Wang & Li \(2018\)](#), we reinforce the relationship between tourism and big data as users are sources of data generation mainly due to devices that allow the collection and interpretation of their movement. In our study, this was evident through the use of social media. We still evidence the impact that big data management has on the market and the business world; [Seles, Jabbour, Jabbour, Fiorini, Mohd-Yusoff & Thome \(2018\)](#) showed the positive effect from an environmental, operational and economic perspective. Convergent also to the study by [Rialti, Marzi, Ciappei, & Busso \(2019\)](#), we highlight the need for big data management in organizations increasingly focused on strategies for decision-making and knowledge management. These data-driven decisions, therefore, are evidence based and stand out when compared to intuition-based decisions ([McAfee & Brynjolfsson, 2012](#)).

The verification of the types of problems involved in these applications reflects the capabilities of the professionals involved in this context. The use of big data in DSS, in addition to knowledge in data collection, storage and processing technologies and data analysis techniques, requires professionals to be able to identify specific complex problems, and this requires a profound knowledge of these professionals about these application areas of big data.

We note an inverse relationship in the complexity of the form of data analysis concerning the need for intuition-based human decision-making. For example, [Vera-Baquero *et al.* \(2015\)](#) using descriptive analysis used event correlation mechanisms to monitor instances and processes to assist analysts in their decision-making. [Miah *et al.* \(2017\)](#) through predictive algorithmic analysis identified areas of high photographic activity to find groups at popular tourist sites. And [Pape \(2016\)](#) presented a prescriptive analysis framework for prioritizing items in business analysis and applying them in a business environment.

In this study, the articles had predominance in predictive analysis, thus suggesting an intermediate position of the intensity of complexity in data analysis and the need for human decision-making ([Figure 1](#)).

Big data as a value generator in DSS

REGE	Field	Scope	Articles
	Logistic and traffic	Route optimization for cost savings, performance and improvements	Hurter <i>et al.</i> (2014), Ilie-Zudor <i>et al.</i> (2015), Neaga <i>et al.</i> (2015), Lan <i>et al.</i> (2016), Chen <i>et al.</i> (2017), Hsu <i>et al.</i> (2017), Rönnqvist <i>et al.</i> (2017), Aversa <i>et al.</i> (2018), Hao <i>et al.</i> (2018), Jamshidi <i>et al.</i> (2018), Jin & Kim (2018), Lee <i>et al.</i> (2018), Sathiaraj <i>et al.</i> (2018), Wang <i>et al.</i> (2018a)
	Health	Population health prediction, disease detection assistance and information mapping to improve care	Khansa <i>et al.</i> (2012), Ashrafi <i>et al.</i> (2014), Burattin <i>et al.</i> (2015), Capobianco & Liò (2015), Feuerriegel (2016), Goh <i>et al.</i> (2016), Baechle <i>et al.</i> (2017), Nguyen <i>et al.</i> (2017), Ghaddar & Naoum-Sawaya (2018), Sivamani <i>et al.</i> (2018), Tashkandi <i>et al.</i> (2018), Wang <i>et al.</i> (2018a)
	Business and market	Assistance in business decision-making and social relationship improvement	Demirkan & Delen (2013), Boutkhoum <i>et al.</i> (2016), Osuszek <i>et al.</i> (2016), Pape (2016), Boutkhoum <i>et al.</i> (2017), Moore (2017), Ghaddar & Naoum-Sawaya (2018), Janković <i>et al.</i> (2018), Tian <i>et al.</i> (2018), Ghasemaghaei & Calic (2019), Mello & Martins (2019), Power <i>et al.</i> (2019)
	Industry and IT	Process efficiency, operating costs and activity optimization	Almeida <i>et al.</i> (2015), Sasson <i>et al.</i> (2015), Rasulius <i>et al.</i> (2016), Shrestha <i>et al.</i> (2016), Jukic <i>et al.</i> (2017), Li <i>et al.</i> (2017), Ahmed <i>et al.</i> (2018), Nimmagadda <i>et al.</i> (2018), Stein <i>et al.</i> (2018)
	Supply chain	Support for strategic decision-making and control of key performance indicators	Groves <i>et al.</i> (2014), Vera-Baquero <i>et al.</i> (2015), Giannakis & Louis (2016), Long (2017), Brinch <i>et al.</i> (2018), Papagiannidis <i>et al.</i> (2018)
	Weather	Climate analysis to predict retail consumption flow, traffic and crisis management	Ludwig <i>et al.</i> (2015), Drosio & Stanek (2016), Schnase <i>et al.</i> (2017), Lee <i>et al.</i> (2018), Sathiaraj <i>et al.</i> (2018), Tian <i>et al.</i> (2018)
	Tourism and/or social media	Tourist flow trend analysis according to visit log	Li <i>et al.</i> (2016), Miah <i>et al.</i> (2017), Nguyen <i>et al.</i> (2017), Gao <i>et al.</i> (2018), Giglio <i>et al.</i> (2019)
	Financial	Decision support in risk analysis, accounting and stock index trends	Hayashi (2016), Chan & Chong (2017), Al Chahadah <i>et al.</i> (2018), Khan <i>et al.</i> (2018)
	Social problems	Assistance in identifying social problems and mitigating misleading decisions based on the database	Gillingham (2019a), Gillingham (2019b), Suh (2019), Schneider & Seelmeyer (2019)
	Smart city	Assistance for planning systems and sustainable urban growth	Semanjski <i>et al.</i> (2016), Deal <i>et al.</i> (2017), Pettit <i>et al.</i> (2018)
Table 4. Types of issues solved with DSS	Government	Big data collection centralization can give conditions of control over the population	Fredriksson (2018), Power (2016)

We can infer based on the analyzed information that, in the coming years, there should be an increase in the number of works with prescriptive analyzes approaching, given the increase in applications, for example, in studies with autonomous cars and route optimization.

In general, the types of analysis are defined according to the understanding of the problem, so that for data analysis of past events, BDA techniques are used for descriptive analysis, and for inference and prediction problems, in which it seeks to identify possible results and probabilities of an event, machine learning has been used.

Although they have been presented separately, the relationship between the types of problems, techniques and technologies is used to ensure DSS is used as a way to solve

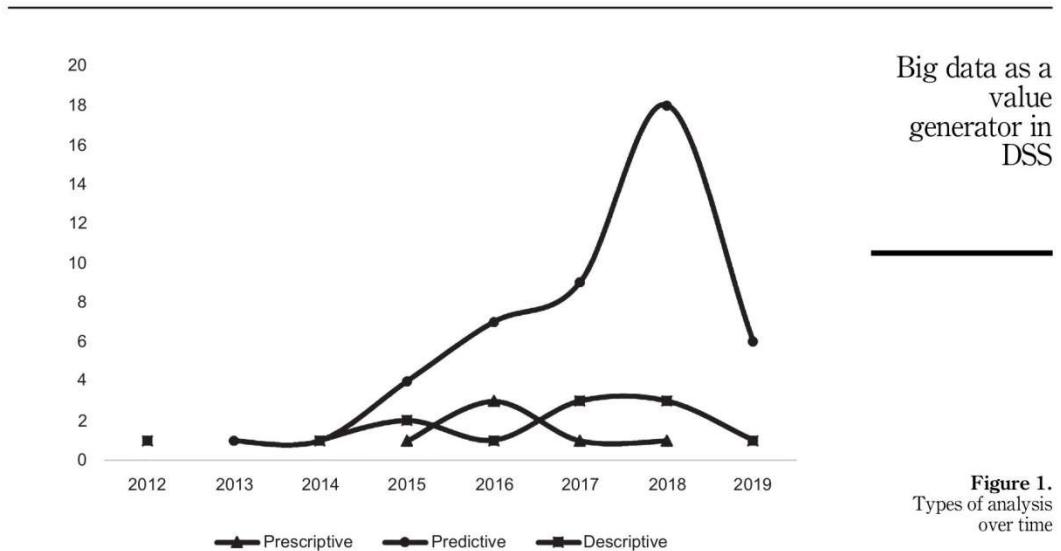


Figure 1.
Types of analysis over time

complex problems, given the human limitation to identify relationships in large data sets. This requires the help of technologies and techniques, which provide mechanisms for obtaining insights that help decision-making.

The value generated by big data in DSS, therefore, occurs when there is better decision-making. For this, techniques and technologies support applications in the most different areas as previously presented. However, the quality of decision-making depends not only on big data but also on the ability to manage the big data chain, from data collection to final insight.

4.1 Research agenda

Analyses of the big data literature have been carried out in some areas; [Seles et al. \(2018\)](#), for example, argued that big data management means a major opportunity in corporate sustainability, due to its potential to generate a better understanding of the opportunities and challenges of climate change. [Li et al. \(2018\)](#) argued about some disadvantages of using big data in tourism-related situations, due to data quality, data cost and privacy issues. And [Bhardwaj, Wodajo, Spano, Neal & Coustasse \(2018\)](#), when investigating the application of big data analyses in the health area, identified benefits in several phases of the management of chronic diseases and could help to reduce the burden of chronic diseases in patients.

As it is still a new area, research on big data is very recent and, at the same time, there are still many research opportunities to investigate this environment, its challenges and advances obtained over the last few years. With that, we proposed to carry out this comprehensive research, so that it was possible to register areas that the application of big data has been used to solve real problems and also the types of techniques and technologies used more recurrently.

With this, some research opportunities emerge from our results, such as, for example, investigations on the behavioral profile of professionals involved with the development of DSS, which capabilities stand out for certain areas. Another point underexplored and that deserves more attention is related to the ethical issues involved in this type of context, and then it could be researched how ethics is being discussed when developing DSS for decision-

REGE

making involving civil control with freedom restrictions, incentive consumption and automatic judgments based on the previous decisions.

Another point that could be explored as an extension to this study would be a research on articles that discuss big data as a source of information for decision-making, without necessarily referring to DSS. The reason for this is that DSS is a terminology that cannot be used by all those who discuss the use of Big Data as a support for decision-making.

5. Concluding remarks

As a result of our study, we note the use of techniques such as BDA, data mining and machine learning algorithms, which were classified into supervised, semi-supervised, unsupervised and reinforcement learning. We also highlight the use of technologies such as BI to report, analyze and present data; cloud computing as a distributed data system provided as a service via the network; data warehouse as a specialized and optimized reporting database, often used to store large amounts of structured data, and information systems. We believe that this is the main contribution to theory.

Through the emerging and incipient subject that the article addresses, we seek to present an overview of the techniques and technologies that are being used to solve problems in their respective areas, as a form of theoretical contribution. This understanding contributes as a way to fill a research gap involving an increasingly important issue for the whole of society.

We have also identified several application areas, such as logistics and traffic for route optimization and operational cost reduction, health for disease detection and treatment guidance, business and market-driven decision-making for competitive advantage and areas with lower application rates in our analyzes as a government in social control and smart cities, in support of urban planning. Within this context of technique or technology applied in some areas, we note the inverse relationship between the complexity of data processing and the need for human decision-making, which results in the form of analysis, which may be descriptive, predictive or prescriptive. In our opinion, this is the main practical contribution of the study.

We also understand that there is a practical contribution to the maturation of the discussion and with reflections even presented as suggestions for future research, such as, for example, the ethical discussion. Our descriptive classification can also serve as a guide for new researchers who seek to understand the research environment involving DSS and big data to obtain value; we believe that this contribution is the most important for society.

We observed that articles related to the theme have increased in number over the years. The ease of access to components that enable data collection and analysis explains the better portability, remote transfer and access of scanned data from one context to another and interconnectivity, the possibility of synthesizing big data and finding connections from the interaction between human and algorithmic intelligence (Günther, Mehrizi, Huysman & Feldberg, 2017), made it possible to improve capturing data and increase the value of big data.

We conclude that the value obtained from big data through DSS occurs with the use of techniques and technologies, which may or not be used together, and that help decision-makers. These insights would not be obtained without using the techniques and technologies in question, given the large volume of data that would make it impossible for a person to collect, store, process and analyze without the aid of DSS.

References

- Ahmed, V., Aziz, Z., Tezel, A., & Riaz, Z. (2018). Challenges and drivers for data mining in the AEC sector. *Engineering, Construction and Architectural Management*, 25(11), 1436–1453, doi: [10.1108/ECAM-01-2018-0035](https://doi.org/10.1108/ECAM-01-2018-0035).

- Al Chahadah, A.R., El Refae, G.A., & Qasim, A. (2018). The use of data mining techniques in accounting and finance as a corporate strategic tool: An empirical investigation on banks operating in emerging economies. *International Journal of Economics and Business Research*, 15(4), 442–452, doi: [10.1504/IJEBR.2018.092149](https://doi.org/10.1504/IJEBR.2018.092149).
- Almeida, R., Bernardino, J., & Furtado, P. (2015). Testing SQL and NoSQL approaches for big data warehouse systems. *International Journal of Business Process Integration and Management*, 7(4), 322–334, doi: [10.1504/IJBPM.2015.073656](https://doi.org/10.1504/IJBPM.2015.073656).
- Ashrafi, N., Kuilboer, J.P., Joshi, C., Ran, I., & Pande, P. (2014). Health informatics in the classroom: An empirical study to investigate higher education's response to healthcare transformation. *Journal of Information Systems Education*, 25(4), 305–315.
- Aversa, P., Cabantous, L., & Haefliger, S. (2018). When decision support systems fail: Insights for strategic information systems from Formula 1. *Journal of Strategic Information Systems*, 27(3), 221–236, doi: [10.1016/j.jsis.2018.03.002](https://doi.org/10.1016/j.jsis.2018.03.002).
- Baechle, C., Agarwal, A., & Zhu, X. (2017). Big Data driven co-occurring evidence discovery in chronic obstructive pulmonary disease patients. *Journal of Big Data*, 4(1), doi: [10.1186/s40537-017-0067-6](https://doi.org/10.1186/s40537-017-0067-6).
- Bhardwaj, N., Wodajo, B., Spano, A., Neal, S., & Coustasse, A. (2018). The impact of big data on chronic disease management. *Health Care Manager*, 37(1), 90–98, doi: [10.1097/HCM.0000000000000194](https://doi.org/10.1097/HCM.0000000000000194).
- Boutkhoum, O., Hanine, M., Agouti, T., & Tikniouine, A. (2016). Selection problem of cloud solution for big data accessing: Fuzzy AHP-PROMETHEE as a proposed methodology. *Journal of Digital Information Management*, 14(6), 368–382.
- Boutkhoum, O., Hanine, M., Agouti, T., & Tikniouine, A. (2017). A decision-making approach based on fuzzy AHP-TOPSIS methodology for selecting the appropriate cloud solution to manage big data projects. *International Journal of Systems Assurance Engineering and Management*, 8, 1237–1253, doi: [10.1007/s13198-017-0592-x](https://doi.org/10.1007/s13198-017-0592-x).
- Boyd, D., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, 15(5), 662–679, doi: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878).
- Brinch, M., Stentoft, J., Jensen, J.K., & Rajkumar, C. (2018). Practitioners understanding of big data and its applications in supply chain management. *International Journal of Logistics Management*, 29(2), 555–574, doi: [10.1108/JLM-05-2017-0115](https://doi.org/10.1108/JLM-05-2017-0115).
- Burattin, A., Cimitile, M., Maggi, F.M., & Sperduti, A. (2015). Online discovery of declarative process models from event streams. *IEEE Transactions on Services Computing*, 8(6), 833–846, doi: [10.1109/TSC.2015.2459703](https://doi.org/10.1109/TSC.2015.2459703).
- Capobianco, E., & Liò, P. (2015). Comorbidity networks: beyond disease correlations. *Journal of Complex Networks*, 3(3), 319–332, doi: [10.1093/comnet/cnu048](https://doi.org/10.1093/comnet/cnu048).
- Chan, S.W.K., & Chong, M.W.C. (2017). Sentiment analysis in financial texts. *Decision Support Systems*, 94, 53–64, doi: [10.1016/j.dss.2016.10.006](https://doi.org/10.1016/j.dss.2016.10.006).
- Chen, C.L.P., & Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275, 314–347, doi: [10.1016/j.ins.2014.01.015](https://doi.org/10.1016/j.ins.2014.01.015).
- Chen, M., Mao, S., & Liu, Y. (2014a). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209, doi: [10.1007/s11036-013-0489-0](https://doi.org/10.1007/s11036-013-0489-0).
- Chen, M., Mao, S., Zhang, Y., Leung, V.C. (2014). *Big data: Related technologies, challenges and future prospects*. Heidelberg: Springer.
- Chen, C., Achtari, G., Majkut, K., & Sheu, J.B. (2017). Balancing equity and cost in rural transportation management with multi-objective utility analysis and data envelopment analysis: A case of quinte west. *Transportation Research Part A: Policy and Practice*, 95, 148–165, doi: [10.1016/j.tra.2016.10.015](https://doi.org/10.1016/j.tra.2016.10.015).

REGE

- Constantiou, I.D., & Kallinikos, J. (2014). New games, new rules: Big Data and the changing context of strategy. *Journal of Information Technology*, 30(1), 44–57, doi: [10.1057/jit.2014.17](https://doi.org/10.1057/jit.2014.17).
- Deal, B., Pan, H., Pallathucheril, V., & Fulton, G. (2017). Urban resilience and planning support systems: The need for sentience. *Journal of Urban Technology*, 24(1), 29–45, doi: [10.1080/10630732.2017.1285018](https://doi.org/10.1080/10630732.2017.1285018).
- Demirkiran, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and Big Data in cloud. *Decision Support Systems*, 55(1), 412–421, doi: [10.1016/j.dss.2012.05.048](https://doi.org/10.1016/j.dss.2012.05.048).
- Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, 37, 267–281, doi: [10.1016/j.future.2013.07.014](https://doi.org/10.1016/j.future.2013.07.014).
- Drosio, S., & Stanek, S. (2016). The Big Data concept as a contributor of added value to crisis decision support systems. *Journal of Decision Systems*, 25(June), 228–239, doi: [10.1080/12460125.2016.1187404](https://doi.org/10.1080/12460125.2016.1187404).
- Feuerriegel, S. (2016). Decision support in healthcare: Determining provider influence on treatment outcomes with robust risk adjustment. *Journal of Decision Systems*, 25(4), 371–390, doi: [10.1080/12460125.2016.1187553](https://doi.org/10.1080/12460125.2016.1187553).
- Fredriksson, C. (2018). Big data creating new knowledge as support in decision-making: Practical examples of big data use and consequences of using big data as decision support. *Journal of Decision Systems*, 27(1), 1–18, doi: [10.1080/12460125.2018.1459068](https://doi.org/10.1080/12460125.2018.1459068).
- Gao, W., Zhang, Q., Lu, Z., Wu, D., & Du, X. (2018). Modelling and application of fuzzy adaptive minimum spanning tree in tourism agglomeration area division. *Knowledge-Based Systems*, 143, 317–326, doi: [10.1016/j.knosys.2017.06.007](https://doi.org/10.1016/j.knosys.2017.06.007).
- Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, 265(3), 993–1004, doi: [10.1016/j.ejor.2017.08.040](https://doi.org/10.1016/j.ejor.2017.08.040).
- Ghasemaghaei, M., & Calic, G. (2019). Can big data improve firm decision quality? The role of data quality and data diagnosticity. *Decision Support Systems*, 120, 38–49, doi: [10.1016/j.dss.2019.03.008](https://doi.org/10.1016/j.dss.2019.03.008).
- Giannakis, M., & Louis, M. (2016). Journal of enterprise information management article information: A multi-agent based system with big data processing for enhanced supply chain agility. *Journal of Enterprise Information Management*, 29(5), 706–727.
- Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, 72(January 2018), 306–312, doi: [10.1016/j.tourman.2018.12.007](https://doi.org/10.1016/j.tourman.2018.12.007).
- Gillingham, P. (2019a). Can predictive algorithms assist decision-making in social work with children and families? *Child Abuse Review*, 28(March), 114–126, doi: [10.1002/car.2547](https://doi.org/10.1002/car.2547).
- Gillingham, P. (2019b). Decision support systems, social justice and algorithmic accountability in social work: A new challenge. *Practice*, 31(4), 277–290, doi: [10.1080/09503153.2019.1575954](https://doi.org/10.1080/09503153.2019.1575954).
- Goh, W.P., Tao, X., Zhang, J., & Yong, J. (2016). Decision support systems for adoption in dental clinics: A survey. *Knowledge-Based Systems*, 104, 195–206, doi: [10.1016/j.knosys.2016.04.022](https://doi.org/10.1016/j.knosys.2016.04.022).
- Groves, W., Collins, J., Gini, M., & Ketter, W. (2014). Agent-assisted supply chain management: Analysis and lessons learned. *Decision Support Systems*, 57(1), 274–284, doi: [10.1016/j.dss.2013.09.006](https://doi.org/10.1016/j.dss.2013.09.006).
- Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191–209, doi: [10.1016/j.jsis.2017.07.003](https://doi.org/10.1016/j.jsis.2017.07.003).
- Hao, S., Yang, L., & Shi, Y. (2018). Data-driven car-following model based on rough set theory. *IET Intelligent Transport Systems*, 12(1), 49–57, doi: [10.1049/iet-its.2017.0006](https://doi.org/10.1049/iet-its.2017.0006).
- Hayashi, Y. (2016). Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. *Operations Research Perspectives*, 3, 32–42, doi: [10.1016/j.orp.2016.08.001](https://doi.org/10.1016/j.orp.2016.08.001).

- Hsu, C.Y., Lim, S.S., & Yang, C.S. (2017). Data mining for enhanced driving effectiveness: An eco-driving behaviour analysis model for better driving decisions. *International Journal of Production Research*, 55(23), 7096–7109, doi: [10.1080/00207543.2017.1349946](https://doi.org/10.1080/00207543.2017.1349946).
- Huang, S., & Chaovallitwongse, W.A. (2015). Computational optimization and statistical methods for big data analytics: Applications in neuroimaging. *The Operations Research Revolution*, 71–88, doi: [10.1287/educ.2015.0135](https://doi.org/10.1287/educ.2015.0135).
- Hurter, C., Conversy, S., Gianazza, D., & Telea, A.C. (2014). Interactive image-based information visualization for aircraft trajectory analysis. *Transportation Research Part C: Emerging Technologies*, 47(P2), 207–227, doi: [10.1016/j.trc.2014.03.005](https://doi.org/10.1016/j.trc.2014.03.005).
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for Dummies*. New York, NY: John Wiley & Sons.
- Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P., & Monostori, L. (2015). Advanced predictive-analysis-based decision support for collaborative logistics networks. *Supply Chain Management*, 20(4), 369–388, doi: [10.1108/SCM-10-2014-0323](https://doi.org/10.1108/SCM-10-2014-0323).
- Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núñez, A., Dollevoet, R., Schutter, B.D., & Li, Z. (2018). A decision support approach for condition-based maintenance of rails based on big data analysis. *Transportation Research Part C: Emerging Technologies*, 95(June), 185–206, doi: [10.1016/j.trc.2018.07.007](https://doi.org/10.1016/j.trc.2018.07.007).
- Janković, S., Mladenović, S., Mladenović, D., Vesković, S., & Glavić, D. (2018). Schema on read modeling approach as a basis of Big Data analytics integration in EIS. *Enterprise Information Systems*, 12(8–9), 1180–1201, doi: [10.1080/17517575.2018.1462404](https://doi.org/10.1080/17517575.2018.1462404).
- Jin, D.H., & Kim, H.J. (2018). Integrated understanding of big data, big data analysis, and business intelligence: A case study of logistics. *Sustainability (Switzerland)*, 10(10), doi: [10.3390/su10103778](https://doi.org/10.3390/su10103778).
- Jukic, N., Jukic, B., Sharma, A., Nestorov, S., & Korallus Arnold, B. (2017). Expediting analytical databases with columnar approach. *Decision Support Systems*, 95, 61–81, doi: [10.1016/j.dss.2016.12.002](https://doi.org/10.1016/j.dss.2016.12.002).
- Khan, U., Aadil, F., Ghazanfar, M.A., Khan, S., Metawa, N., Muhammad, K., Mehmood, I., & Nam, Y. (2018). A robust regression-based stock exchange forecasting and determination of correlation between stock markets. *Sustainability (Switzerland)*, 10(10), 1–20, doi: [10.3390/su10103702](https://doi.org/10.3390/su10103702).
- Khansa, L., Forcade, J., Nambari, G., Parasuraman, S., & Cox, P. (2012). Proposing an intelligent cloud-based electronic health record system. *International Journal of Business Data Communications and Networking*, 8(3), 57–71, doi: [10.4018/jbdcn.2012070104](https://doi.org/10.4018/jbdcn.2012070104).
- Lan, S., Zhang, H., Zhong, R.Y., & Huang, G.Q. (2016). A customer satisfaction evaluation model for logistics services using fuzzy analytic hierarchy process. *Industrial Management & Data Systems*, 116(5), 1024–1042, doi: [10.1108/IMDS-09-2015-0389](https://doi.org/10.1108/IMDS-09-2015-0389).
- Lee, H., Aydin, N., Choi, Y., Lekhavat, S., & Irani, Z. (2018). A decision support system for vessel speed decision in maritime logistics using weather archive big data. *Computers and Operations Research*, 98, 330–342, doi: [10.1016/j.cor.2017.06.005](https://doi.org/10.1016/j.cor.2017.06.005).
- Li, J., Li, X., & Zhu, B. (2016). User opinion classification in social media: A global consistency maximization approach. *Information and Management*, 53(8), 987–996, doi: [10.1016/j.im.2016.06.004](https://doi.org/10.1016/j.im.2016.06.004).
- Li, F., Yang, J., Jin, C., & Guo, C. (2017). A new effect-based roughness measure for attribute reduction in information system. *Information Sciences*, 378, 348–362, doi: [10.1016/j.ins.2016.08.056](https://doi.org/10.1016/j.ins.2016.08.056).
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323, doi: [10.1016/j.tourman.2018.03.009](https://doi.org/10.1016/j.tourman.2018.03.009).
- Long, Q. (2017). A framework for data-driven computational experiments of inter-organizational collaborations in supply chain networks. *Information Sciences*, 399, 43–63, doi: [10.1016/j.ins.2017.03.008](https://doi.org/10.1016/j.ins.2017.03.008).
- Ludwig, N., Feuerriegel, S., & Neumann, D. (2015). Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems*, 24(1), 19–36, doi: [10.1080/12460125.2015.994290](https://doi.org/10.1080/12460125.2015.994290).

- REGE McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-68.
- Mello, R., & Martins, R.A. (2019). Can big data analytics enhance performance measurement systems?, *IEEE Engineering Management Review*, 47(1), 52–57, doi: [10.1109/EMR.2019.2900645](https://doi.org/10.1109/EMR.2019.2900645).
- Miah, S.J., Vu, H.Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. *Information and Management*, 54(6), 771–785, doi: [10.1016/j.im.2016.11.011](https://doi.org/10.1016/j.im.2016.11.011).
- Moore, J. (2017). Data Visualization in support of executive decision making. *Journal of Information, Knowledge, and Management*, 12, 125–138.
- Mortenson, M.J., Doherty, N.F. & Robinson, S. (2015), Operational research from taylorism to terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3), 583–595.
- Murdoch, T.B., & Detsky, A.S. (2013). The inevitable application of big data to health care. *JAMA - Journal of the American Medical Association*, 309(13), 1351–1352, doi: [10.1001/jama.2013.393](https://doi.org/10.1001/jama.2013.393).
- Neaga, I., Liu, S., Xu, L., Chen, H., & Hao, Y. (2015). Cloud enabled big data business platform for logistics services: A research and development agenda. *Lecture Notes in Business Information Processing*, 216, 22–33, doi: [10.1007/978-3-319-18533-0](https://doi.org/10.1007/978-3-319-18533-0).
- Nguyen, T., Larsen, M.E., O'Dea, B., Nguyen, D.T., Yearwood, J., Phung, D., Venkatesh, S., & Christensen, H. (2017). Kernel-based features for predicting population health indices from geocoded social media data. *Decision Support Systems*, 102, 22–31, doi: [10.1016/j.dss.2017.06.010](https://doi.org/10.1016/j.dss.2017.06.010).
- Nimmagadda, S.L., Reiners, T., & Wood, L.C. (2018). On big data-guided upstream business research and its knowledge management. *Journal of Business Research*, 89(September 2016), 143–158, doi: [10.1016/j.jbusres.2018.04.029](https://doi.org/10.1016/j.jbusres.2018.04.029).
- Osuszek, L., Stanek, S., & Twardowski, Z. (2016). Leverage big data analytics for dynamic informed decisions with advanced case management. *Journal of Decision Systems*, 25(June), 436–449, doi: [10.1080/12460125.2016.1187401](https://doi.org/10.1080/12460125.2016.1187401).
- Papagiannidis, S., See-To, E.W.K., Assimakopoulos, D.G., & Yang, Y. (2018). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age?. *Computers and Operations Research*, 98, 355–366, doi: [10.1016/j.cor.2017.06.010](https://doi.org/10.1016/j.cor.2017.06.010).
- Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research*, 252(2), 687–698, doi: [10.1016/j.ejor.2016.01.052](https://doi.org/10.1016/j.ejor.2016.01.052).
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.
- Petticrew, M., & Roberts, H. (2008). *Systematic Reviews in the Social Sciences: A Practical Guide*. John Wiley and Sons.
- Pettit, C., Bakelman, A., Lieske, S.N., Glackin, S., Hargroves, K.C., Thomson, G., Shearer, H., Dia, H., & Newman, P. (2018). Planning support systems for smart cities. *City, Culture and Society*, 12(October), 13–24, doi: [10.1016/j.ccs.2017.10.002](https://doi.org/10.1016/j.ccs.2017.10.002).
- Power, D.J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, 25, 578–588, doi: [10.1080/12460125.2016.1187420](https://doi.org/10.1080/12460125.2016.1187420).
- Power, D.J., Cyphert, D., & Roth, R.M. (2019). Analytics, bias, and evidence: The quest for rational decision making. Special issue on cognitive bias. *Journal of Decision Systems*, 28, 1–18, doi: [10.1080/12460125.2019.1623534](https://doi.org/10.1080/12460125.2019.1623534).
- Rasiulis, R., Ustinovichius, L., Vilutiene, T., & Popov, V. (2016). Decision model for selection of modernization measures: Public building case. *Journal of Civil Engineering and Management*, 22(1), 124–133, doi: [10.3846/13923730.2015.1117018](https://doi.org/10.3846/13923730.2015.1117018).

- Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2019). Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. *Management Decision*. doi: [10.1108/md-07-2018-0821](https://doi.org/10.1108/md-07-2018-0821).
- Rönnqvist, M., Svenson, G., Flisberg, P., & Jönsson, L.E. (2017). Calibrated route finder: Improving the safety, environmental consciousness, and cost effectiveness of truck routing in Sweden. *Interfaces*, 47(5), 372–395, doi: [10.1287/inte.2017.0906](https://doi.org/10.1287/inte.2017.0906).
- Sasson, E., Ravid, G., & Pliskin, N. (2015). Improving similarity measures of relatedness proximity: Toward augmented concept maps. *Journal of Informetrics*, 9(3), 618–628, doi: [10.1016/j.joi.2015.06.003](https://doi.org/10.1016/j.joi.2015.06.003).
- Sathiaraj, D., Punkasem, T., Wang, F., & Seedah, D.P.K. (2018). Data-driven analysis on the effects of extreme weather elements on traffic volume in Atlanta, GA, USA. *Computers, Environment and Urban Systems*, 72(June), 212–220, doi: [10.1016/j.compenvurbsys.2018.06.012](https://doi.org/10.1016/j.compenvurbsys.2018.06.012).
- Schnase, J.L., Duffy, D.Q., Tamkin, G.S., Nadeau, D., Thompson, J.H., Grieg, C.M., McInerney, M.A., & Webster, W.P. (2017). MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. *Computers, Environment and Urban Systems*, 61, 198–211, doi: [10.1016/j.compenvurbsys.2013.12.003](https://doi.org/10.1016/j.compenvurbsys.2013.12.003).
- Schneider, D., & Seelmeyer, U. (2019). Challenges in using big data to develop decision support systems for social work in Germany. *Journal of Technology in Human Services*, 37(2–3), 113–128, doi: [10.1080/15228835.2019.1614513](https://doi.org/10.1080/15228835.2019.1614513).
- Seles, B.M.R.P., Jabbour, A.B.L.D.S., Jabbour, C.J.C., Fiorini, P.D.C., Mohd-Yusoff, Y., & Thomé, A.M.T. (2018). Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for Big Data management towards a low carbon society. *Journal of Cleaner Production*, 189, 763–774, doi: [10.1016/j.jclepro.2018.04.113](https://doi.org/10.1016/j.jclepro.2018.04.113).
- Semanjski, I., Bellens, R., Gautama, S., & Witlox, F. (2016). Integrating big data into a sustainable mobility policy 2.0 planning support system. *Sustainability (Switzerland)*, 8(11), 1–19, doi: [10.3390/su8111142](https://doi.org/10.3390/su8111142).
- Shrestha, A., Cater-Steel, A., & Toleman, M. (2016). Innovative decision support for IT service management. *Journal of Decision Systems*, 25(June), 486–499, doi: [10.1080/12460125.2016.1187424](https://doi.org/10.1080/12460125.2016.1187424).
- Sivamani, S., Choi, J., & Cho, Y. (2018). A service model for nutrition supplement prediction based on Fuzzy Bayes model using bigdata in livestock. *Annals of Operations Research*, 265(2), 257–268, doi: [10.1007/s10479-017-2490-7](https://doi.org/10.1007/s10479-017-2490-7).
- Stein, N., Meller, J., & Flath, C.M. (2018). Big data on the shop-floor: Sensor-based decision-support for manual processes. *Journal of Business Economics*, 88(5), 593–616, doi: [10.1007/s11573-017-0890-4](https://doi.org/10.1007/s11573-017-0890-4).
- Suh, J.H. (2019). SocialTERM-Extractor: Identifying and predicting social-problem-specific key noun terms from a large number of online news articles using text mining and machine learning techniques. *Sustainability (Switzerland)*, 11(1), doi: [10.3390/su11010196](https://doi.org/10.3390/su11010196).
- Tashkandi, A., Wiese, I., & Wiese, L. (2018). Efficient in-database patient similarity analysis for personalized medical decision support systems. *Big Data Research*, 13(May), 52–64, doi: [10.1016/j.bdr.2018.05.001](https://doi.org/10.1016/j.bdr.2018.05.001).
- Tian, J., Zhang, Y., & Zhang, C. (2018). Predicting consumer variety-seeking through weather data analytics. *Electronic Commerce Research and Applications*, 28, 194–207, doi: [10.1016/j.elerap.2018.02.001](https://doi.org/10.1016/j.elerap.2018.02.001).
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222.
- Vera-Baquero, A., Colomo-Palacios, R., Molloy, O., & Elbattah, M. (2015). Business process improvement by means of big data based decision support systems: A case study on call centers. *International Journal of Information Systems and Project Management*, 3(1), 5–26, doi: [10.12821/ijispdm030101](https://doi.org/10.12821/ijispdm030101).
- Wang, G., Gunasekaran, A., Ngai, E.W.T., & Papadopoulos, T. (2016). Big Data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98–110, doi: [10.1016/j.ijpe.2016.03.014](https://doi.org/10.1016/j.ijpe.2016.03.014).

Big data as a value generator in DSS

REGE

- Wang, Y., Kung, L.A., & Byrd, T.A. (2018a). Big Data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13, doi: 10.1016/j.techfore.2015.12.019.
- Wang, Y., Geng, S., & Gao, H. (2018b). A proactive decision support method based on deep reinforcement learning and state partition. *Knowledge-Based Systems*, 143, 248–258, doi: 10.1016/j.knosys.2017.11.005.

About the authors

Gustavo Grander Professional Doctoral Student in Project Management from Nove de Julho University, Master in Business Administration from Western Paraná State University, Specialist in Project Management from Fundação Getúlio Vargas and Production Engineer by the Federal Technological University of Paraná. Has professional experience in the pharmaceutical industry having acted as Project Manager and Project Analyst, metallurgical industry having acted as Production Supervisor and furniture industry having acted as Production Engineer. It has CAPM certification (Certified Associate in Project Management). Gustavo Grander is the corresponding author and can be contacted at: grandergustavo@gmail.com

Luciano Ferreira da Silva Doctorate in Business Administration at Pontifical Catholic University, Master in Business Administration, Communication and Education at São Marcos University, Specialist in Organizational Psychology, Specialist in Human Resources Management and graduated in Business Administration. Professor and researcher at the Graduate Program in Project Management at Nove de Julho University. Experience in organizations in market assessment, sales projection, survey of new business opportunities and training. Author of the book "Legal Revolutions - Brazilian law and politics", as well as other academic works in the area of operations, sustainability, People Management, Project Management, among others. Scientific Editor at "Gestão e Projetos" magazine.

Ernesto Del Rosario Santibañez Gonzalez Associate Professor in the Department of Industrial Engineering, Executive Director and Founding Council of Industry-University – Talca University. He is associate editor of the *Journal of Cleaner Production* and *Journal of Intelligent Manufacturing*, on the Editorial Board of the *Journal of Power Sources* and the *International Journal of Hydrogen Energy* and is a guest editor of six special editions of ISI and Scopus journals, including the *International Journal of Research, Sustainability, European Journal of Operational Research, Clean Production Journal*, and Guest Editor for two special editions of ISI journals, including the *International Journal of Production Economics, Computers and Industrial Engineering*.

Associate Editor: Luis Pinochet

APÊNDICE B

International Journal of Business Analytics

Volume 9 • Issue 1

A Patent Analysis on Big Data Projects

Gustavo Grandes, Universidade Nove de Julho, Brazil

Luciano Ferreira da Silva, Universidade Nove de Julho, Brazil

 <https://orcid.org/0000-0001-6482-8729>

Ernesto D. R. Santibanez Gonzalez, University of Talca, Chile

ABSTRACT

Studies concerning big data patents have been published; however, research investigating big data projects is scarce. Therefore, the objective of this study was to conduct an exploratory analysis of a patent database to collect information about the characteristics of registered patents related to big data projects. The authors searched for patents related to big data projects in the Espacenet database on January 10, 2021 and identified 109 records. The textual analysis detected three word classes interpreted as (1) a direction to cloud computing, (2) optimization of solutions, and (3) storage and data sharing structures. The results also revealed emerging technologies such as blockchain and the internet of things, which are utilized in big data project solutions. This observation demonstrates the importance that has been given to solutions that facilitate decision-making in an increasingly data-driven context. As a contribution, they understand that this study endorses a group of researchers that has been dedicated to academic research on patent documents.

KEYWORDS

Big Data, Blockchain, Cloud, Intellectual Property, IoT, Patent, Project, Technology

1. INTRODUCTION

Big Data is characterized as a large data set, which is challenging to store, process, analyze, and understand using traditional database processing tools (S. Huang & Chaovatwongse, 2015). The advancement in the use of Big Data has driven a generation of technologies and architectures designed to extract economic value through analysis (Gantz & Reinsel, 2011). Thus, decisions oriented to data analysis stand out compared to decisions based on intuition (McAfee & Brynjolfsson, 2012). With the increase in the generation of unstructured data, the analysis and interpretation through structured tables with rows and columns becomes increasingly complex.

A Big Data project, in its turn, can be defined as a data-intensive project that presents large-scale problems with restrictions of volume, variety, speed (Becker, 2017), and still veracity of the data (Barham & Daim, 2018). Therefore, the culture of data-driven decision-making is essential to the success of a Big Data project (Dutta & Bose, 2015). Big data projects, using cutting-edge analysis involving artificial intelligence and machine learning, condition computers to identify what these data represent by identifying patterns more assertively.

Decision-making based on data depends on the successful implementation of Big Data projects (Dutta & Bose, 2015). However, the excessive focus on data and the neglect of adequate decision-

DOI: 10.4018/IJBAN.288516

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

making has been evidenced as an aspect that reflects the failure of Big Data projects (Chiheb, Boumahdi, & Bouarfa, 2019). Big Data project characteristics reflect potential highly uncertain inputs (for example, the relevance of data) and also potentially highly uncertain results (for example, information may be inaccurately derived from data analysis) (J. Saltz, Shamshurin, & Connors, 2017). Big Data projects have also provided a good discussion regarding success factors. In a broad perspective, there is a convergence between studies that investigated success factors in big data projects and that encompass five predominant characteristics: Governance, Objectives, Processes, Data, Team, and Tools (Gao, Koronios, & Selle, 2015; J. S. Saltz & Shamshurin, 2016).

Despite the evolution made in the Big Data area in recent years, practitioners reinforce the need for research related to guides, models, or even methodologies (Lara, De Sojo, Aljawarneh, Schumaker, & Al-Shargabi, 2020). As a result, opportunities arise for project management professionals to contribute by identifying appropriate measures and types of data, helping to interpret the data, and placing them in an evaluation context (Olsson & Bull-Berg, 2015). In addition to identifying technological trends in specific fields (Abbas, Bilal, Zhang, & Khan, 2015), patent analysis can identify recurring problems in a specific area, as patent registrations are intended to solve a practical problem.

Previous patent analysis studies of Big Data technology have been performed. Examples include abstract analysis of Chinese Big Data patents (Z. Yang et al., 2017), hot classification fields of Big Data technology (Gui, Liu, Bai, & Zhang, 2017), methods technology assessment using quantitative patent analysis for technology transfer in Big Data marketing (Jun, Park, & Jang, 2015), business interests and activities around Big Data (Y. Huang et al., 2016) and an analysis of the patenting activities of global jurisdictions in the Big Data field (Saheb & Saheb, 2020). While these studies analyze Big Data patents, they do not directly address the specific area of Big Data projects.

Singh (2019) stated that Big Data projects involve research in purely technical areas, such as ubiquitous information and integration in Big Data ecosystems. Thus, given the relevance of Big Data projects and the scarcity of studies characterizing the patents generated in this area, we took this as an opportunity for new research. Herein, an exploratory analysis of the Espacenet patent database was performed on January 10, 2021, to identify and subsequently describe registered Big Data projects patents, based on an initial sample of 109 records.

Our results advance the discussion on Big Data projects since this study fills a gap in intellectual property knowledge. This study also endorses the researchers who promote patent research to search for information on solving technical problems without being limited to scientific articles. Finally, based on this study's main findings, we suggest extending this analysis from a Big Data project management perspective in the future.

In the following sections, we present the Theoretical background (Section 2), Methodological procedures (Section 3), Results and Discussion (Section 4), and Conclusions (Section 5) of this study.

2. THEORETICAL BACKGROUND

Previous studies reported an approach for organizations to create Big Data projects, encompassing phases of project planning, implementation, and post-implementation (Mousannif, Sabah, Douiji, & Sayad, 2014, 2016). Moreover, another study proposed implementing Big Data projects based on the strategic basis, data analysis, and implementation phases (Dutta & Bose, 2015). These two proposals, which aim to establish effective Big Data solutions, initially discuss the strategic aspect of the project and the attendance of business problems, reinforcing the holistic characteristic of all Big Data projects.

Interestingly, the strategic alignment of Big Data projects with the organizational vision has been considered a success factor in Big Data projects (Gao et al., 2015; J. S. Saltz & Shamshurin, 2016). Indeed, the proper use of data visualization technologies is crucial for successful data interpretation and must be considered by the decision-makers at the strategic level, such as data analysts and executives (Moore, 2017).

Big data and cloud computing are among the technological revolutions of the moment and have led to significant transformations in information technology and impacts on scientific research (Bouthoum, Hanine, Agouti, & Tikniouine, 2016). Cloud computing provides vital support to address the challenges related to shared computing resources such as computing, storage, networking, and analytical software. Notably, the application of these resources has led to impressive advances in Big Data (C. Yang, Huang, Li, Liu, & Hu, 2017).

Some emerging technologies rely on Big Data. For example, the Internet of Things (IoT) is a network of heterogeneous devices with sensors with different functionalities connected to a private or public network (Miorandi, Sicari, De Pellegrini, & Chlamtac, 2012). Its ability to react to the surrounding environment depends on Big Data. Additionally, Blockchain, a distributed public accounting technology initially intended for cryptocurrencies, utilizes Big Data technologies. It has been reported that systems that combine blockchain technology and the IoT benefit from lower operating costs, decentralized resource management and robustness against threats and attacks. Therefore, the convergence of the IoT and Blockchain technologies could overcome significant challenges currently faced by the IoT (Ferrag et al., 2018).

In the information age, data has become an essential strategic resource. In this sense, Big Data, especially in the IoT era, has become a practical tool for enterprise and government decision-making, specifically impacting economic and social life (Wang, 2020).

Data mining is the computational process of discovering patterns in large data sets. It involves methods at the intersection of Artificial Intelligence (AI), machine learning, statistics, and database systems (Majumdar, Naraseyappa, & Ankalaki, 2017). According to Fayyad, Piatetsky-Shapiro, and Smyth (1996), data mining applies specific algorithms or techniques to extract knowledge patterns from data sets. The patterns discovered should be significant insofar as they lead to some advantage in decision-making (Ahmed, Aziz, Tezel, & Riaz, 2018).

Decision-making models for Big Data projects have been proposed to reduce problems of limited rationality in the context of complex decisions. Chiheb et al. (2019) presented a conceptual model called BD-Da that describes the levels to be considered to develop a Big Data project to solve the often neglected decision-making problem. The model helps companies make faster and more intelligent decisions and gain a competitive advantage.

Elgendi and Elragal (2016) developed a B-DAD structure that integrates Big Data tools, architecture, and analysis. The structure was developed and tested to identify promotional products that should be offered for sale. It was also shown to make decision-making more assertive through advanced Big Data analytical techniques that reveal hidden insights. Later, Poletto, De Carvalho, and Costa (Poletto, De Carvalho, & Costa, 2017) presented an integrated model of Big Data, Business Intelligence, and Decision Support System to apply during the decision-making phases. The authors state that organizations are essentially composed of processes, people, and technology; however, technology alone is useless without one of the first two.

More specific solutions within Big Data projects have also been developed. Examples include a decision-making approach based on the diffuse AHP-TOPSIS methodology to select an appropriate cloud solution to manage Big Data projects (Bouthoum, Hanine, Agouti, & Tikniouine, 2017). Additionally, a project management and execution support platform with an emphasis on analytics and reporting was reported to facilitate successful project decision-making (Pondel & Pondel, 2016). It is also evident that there is a need for greater criticality for the data storage and analysis steps, especially concerning the high risk of collecting low-quality data and the steep learning curve of the initial phase of Big Data projects (Lara et al., 2020).

Large data sets originating from heterogeneous to distributed sources likely have different data formats and models. They may not be in the same data format or model of a target Data Warehouse or contain missing, low quality, redundant or data. Thus, Data Warehouse architecture software, known as Extraction-Transformation-Loading (ETL), is employed (Ali & Wrembel, 2017). The emergence

of Big Data has caused more and more companies to modify their data management strategies to optimize performance (Jianmin, Wenbin, Tongrang, Shilong, & Hongwei, 2020).

After highlighting the relevant discussion based on the studies cited in the context of Big Data projects, in the next section, we describe the methodological procedures of our research and provide details about data collection and analysis.

3. METHODOLOGICAL PROCEDURES

According to Clarke (2018), depending on the questions you are being asked (i.e., state of the art, prior art/novelty/patentability, freedom to operate, opposition, and (in)validity), there are different types of patent searches. Herein, we sought to collect relevant documents published in a particular technical field and extract data and descriptive information.

To achieve our goal of carrying out an exploratory analysis of the characteristics of registered Big Data projects patents, we searched the Espacenet database on January 10, 2021, using the following search string: “Big data project*”. The search retrieved 109 records, and after a structured analysis of the documents, six (06) patents were eliminated from the sample because they did not have a Big Data project scope or cited the search term in a decontextualized way. We also identified two cases of repetitions, in which the patents were registered in two offices. It should be pointed out that for the descriptive analysis, we did not consider the six patents outside the scope of the research. For the textual lexical analysis, we did not consider the six patents outside the scope or the two repeated patents to not bias the quality of the results (Table 1).

Table 1. Patents eliminated and repeated from the sample

	Patent number	Patent name
Eliminated patents	CN108446206A	AR technology-based application availability display system
	US2020324135A1	Device and method for treating eight cancers with ultraviolet radiation
	CN107177625A	Artificial carrier system with site-specific mutagenesis and site-specific mutagenesis method
	CN107671020A	Automatic cleaning brush of seabed observation instrument equipment and work principle of automatic cleaning brush
	CN104484582A	Method and system for automatically analyzing bio information items through modular selection
	WO2019100012A1	Data collection & analytics based on detection of biological cells or biological substances
Repeated patents	US9384264B1	Analytic systems, methods, and computer-readable media for structured, semi-structured, and unstructured documents
	EP3179387A1	
	WO2020014399A1	Decentralized Cybersecure Privacy Network For Cloud Communication, Computing And Global e-Commerce
	US2019386969A1	

For data analysis, we used Iramuteq version 0.7, free and open-source software that enables the textual analysis of a corpus prepared by the researchers. We organized patent summaries according to the available information in the documents. After processing the corpus, 230 text segments and 7.890 occurrences were identified. Only adjectives, verbs, and common names were used as analysis keys.

An analysis of the textual corpus, consisting of 230 text segments and 7,890 occurrences from 101 textual patents, identified 567 hápax or words that appear only once in the entire textual corpus. A summary of the corpus is presented in Table 2.

Table 2. Summary of corpus analysis

Summary of corpus analysis	Total
Number of texts	101
Number of text segments	230
Occurrences	7,890
Average occurrence per text	78.90
Number of forms	1,269
Number of hápax	567*

*44.68% of form, 7.19% of occurrences.

Next, we used the Descending Hierarchical Classification Method (DHCM) to obtain text segment classes. At the same time, DHCM detects text segment vocabulary that is similar and different in the classes (Camargo, 2005). We also performed a Factorial Correspondence Analysis (FCA), a feature of Iramuteq that produces graphical representations in a Cartesian plane. FCA helps visualize the proximity and distances of words from the crossing of the textual corpus vocabulary. The procedures performed in this analysis include the calculation of frequencies, chi-square correlation values for each word in the corpus, and the execution of the FCA in a contingency table that crosses the active forms and the variables (Salviati, 2017).

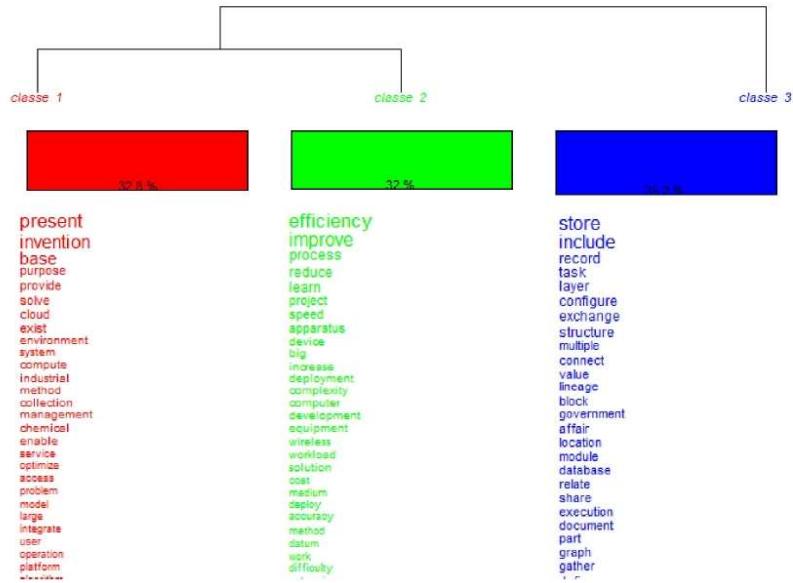
4. RESULTS AND DISCUSSIONS

As shown in Table 3, China displayed an expressive number of registrations in China with 81 patents, corresponding to 78.64% of the sample. China registered one patent in 2015 and has exhibited year-to-year growth until 2020. The United States of America (USA) is the country with the second most registrations with ten records, and the World Intellectual Property Organization (WIPO) is third with six records. Moreover, the Republic of Korea (ROK) had three records, the European Patent Office (EPO) had two, and Japan had one. The large number of registrations and steady increase in productivity over the years highlight China's interest in the topic.

Table 3. Intensity map of the number of records in countries/regions related to big data projects

Country/Region	2015	2016	2017	2018	2019	2020	Total
China	1	2	6	13	17	42	81
USA	2	1	1		3	3	10
WIPO		1		1	1	3	6
ROK				1		2	3
EPO			1		1		2
Japan						1	1
Total	3	4	8	15	22	51	103

Figure 1. Dendrogram of DHCM



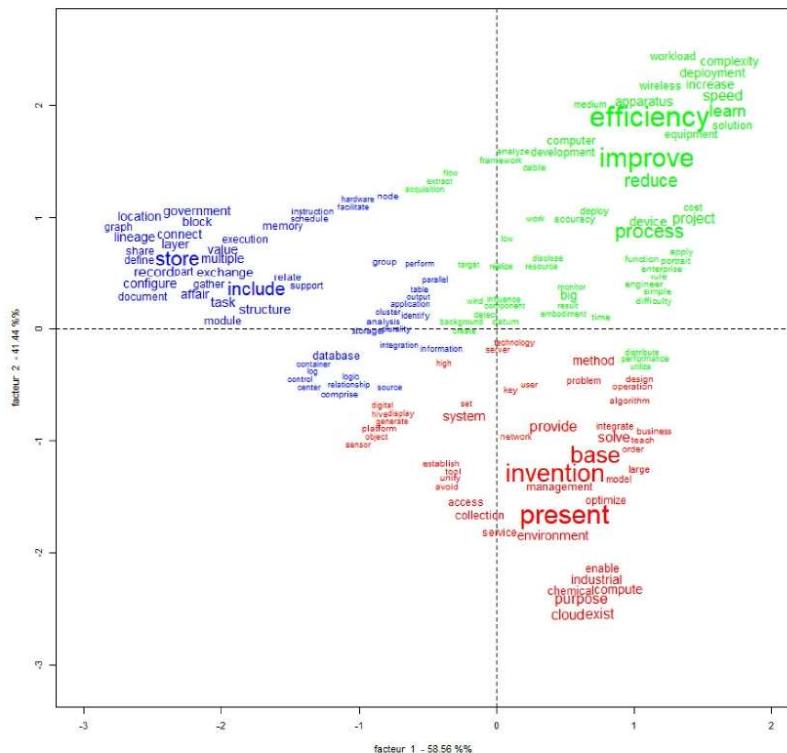
When studying Big Data patents in China, Z. Yang et al. (2017) showed that from 1980–2016 the increase in Big Data patent registrations in China occurred after 2005, and the claimants' patent applications were mainly from universities and companies. Teixeira and Queirós (2016) highlight the 'Intermediate Intellectual Property Service', which since 2008 has been a strategy for technological development to make China a country with a high level of intellectual property creation. Indeed, our intensity map illustrates the predominance of registrations in China, confirming the success of this strategy.

When generating the dendrogram (Fig. 1), which is a tree diagram that displays the groups formed by clustering at their similarity levels, three classes were identified and analyzed according to the relationships between the words. Class 1 represents 32.8% of the corpus and has words like "system", "method", "model" and "management" that are related to the corpus as procedures or forms of organization. Class 1 also contains words like "cloud", "compute", "platform" and "service" that can indicate a direction for the use of cloud computing services. Class 2 represents 32% of the corpus and consists of words such as "efficiency", "improve", "process", "reduce", "speed" and "solution" suggesting the corpus is directed towards the optimization of solutions. Additionally, Class 2 also contains words like "apparatus", "device", "equipment" and "wireless", referring to objects that facilitate solution optimization. Lastly, Class 3 represents 35.2% of the corpus and is characterized by words like "store", "database", "block", "record", "structure", "module" and "share", which are indicative of data storage and sharing structures.

It should be pointed out that since Classes 1 and 2 have equivalence in terms of word means compared to Class 3; thus, they could have been combined into one class.

The FCA (fig. 2) made it possible to visualize the textual corpus distributed in the Cartesian plane. As shown in Figure 2, the three distinct classes do not overlap. The textual corpus was formed by the

Figure 2. FCA of the sample



words in the patent summary, and each word had to be repeated at least ten times to be considered in the FCA. The interpretation of word placement is subjective, and it is up to researchers to understand why the words have this distribution. Thus, after the software generated the FCA, we carried out a deep analysis directed at all patents by their respective words to identify relationships between them and form a common corpus that justifies each word group.

When interpreting the distances between words and distributions along the plane, the x-axis appears to refer to data collection and storage technologies (left) and processing and performance techniques (right). In other words, patents related to Big Data design applications. On the other hand, the y-axis indicates a tangibility relationship such that the lower and upper parts of the plane represent intangible and tangible solutions, respectively.

Due to continuous development, IoT technology and Big Data have become analytical tools widely applied in many technical fields (Wang, 2020). The rapid evolution of decision support systems has made experts realize that putting Big Data on the cloud is a real challenge that businesses must

consider. In this sense, the selection of cloud solutions is recognized as an essential research issue for Big Data projects (Boutkhoum et al., 2016).

A significant challenge faced by this sector is that Big Data generated from the various IoT sensors contains rich spatiotemporal information (C. Yang et al., 2017). We identified a Big Data IoT project in the cloud (Patent number CN112001023A) which provides a method for analyzing the degree of safety from pollution by measuring reflected light from a glass curtain wall. According to the authors, the invention helps to continuously monitor the light pollution phenomenon on the glass curtain wall. Furthermore, utilizing Big Data, the patent objectively analyzes and visualizes the security of the building's light environment for protecting the eyesight and improving the productivity of the employees.

It has been reported that the advancement of IoT is expected to provide easy public access to valuable results from Big Data and cloud computing (C. Yang et al., 2017). However, access to the best information needs to be provided to the public for better decision-making (Abbas et al., 2015). Herein, we found Patent number CN111338302A that refers to the technical field of Smart factories, and more specifically, to a chemical modeling processing system based on industrial Big Data and IoT. For inventors, the invention aims to integrate and implement AI and Big Data technology in an independent and controllable way into traditional production lines, helping production companies achieve synergistic data, production, and control flows, improve production efficiency, reduce production costs. In other words, this technology could contribute to the development of intelligent, independent and controllable manufacturing.

It is predicted that Blockchain technology will soon overcome significant challenges faced in this area of research and development. In this context, a method for extracting and managing project results generated by specialists, executed by a server and stored in a Blockchain has been developed (Patent number WO2020175753A1). Moreover, we also identified a Blockchain consensus algorithm and platform, Blockchain encryption technology, and evolutionary network developed to reduce cryptocurrency transaction costs and time (Patent number KR20200045089A). Alkhamisi and Alboraei (2019) state that with the rise of Blockchain technology, recent research has focused on adopting Blockchain to address security issues associated with IoT. Furthermore, Ferrag et al. (2018) classified threat models considered by Blockchain protocols on IoT networks into five main categories: identity-based attacks, manipulation-based attacks, cryptanalytic attacks, reputation-based attacks, and service-based attacks.

In the sample, we also detected patents related to a remote guidance control system for corporate services (patent number CN107748764A). This invention makes querying a company's service demand information more efficient and convenient by efficiently combining the client, the platform cloud data, and third-party service agencies. Despite the numerous advantages, it is challenging to find recommendation systems with a specific formula that directly provide optimized parameters. This setback is mainly attributed to the fact that more and more parameters are generated as the machine learning models become more and more complex. Notably, it has been proposed that a minimum gradient descent algorithm based on the Gene, Behavior, and Phenotypic (GBP) dimension could mitigate this problem (patent number CN111581524A). Recommendation systems collect historical user classifications, interactions, portraits, social networks, and context and analyze historical interests to recommend relevant information for users to meet needs and recommendation algorithms, which are the core of recommendation systems.

Data mining plays a crucial role in analyzing large data sets. Patent number US10282603B2 refers to a method implemented by a computer to analyze technical documents concerning a corpus. According to the inventors, several algorithms can perform data searches; however, information retrieval is often an inaccurate process with the rise of Big Data projects. Therefore, improvements in information retrieval are necessary. We also found a series of patents related to performance optimization. For example, patent number CN103823881A describes a method and device for optimizing the performance of a distributed database. Additionally, patent number CN106130801A

provides a method for updating and monitoring Big Data, which solves the problems associated with components during traditional distributed data collection.

When an ETL process is implemented as a workflow, where tasks process data and are connected by data flows, ETL tools must be developed to execute these workflows (Ali & Wrembel, 2017). Patent number CN107145576A is an ETL tool that addresses a Big Data ETL scheduling system by supporting visualization and process flow, eliminating complicated background operations, improving ETL development speed and efficiency, and reducing implementation costs. It has been reported that the ETL architecture strategy processes the data uniformly (Jianmin et al., 2020). Moreover, patent CN107330028A discloses an Apache NiFi application extension method for importing source data into a database. According to the method, different access tools are not required when multiple data sources have access to the database. Consequently, the data access process is unified. An access plug-in can also be extended automatically.

The culture of data-driven decision-making has promoted structural changes in processes, people, and technology (Elgendi & Elragal, 2016). The value built from the use of Big Data in the decision-making process lies in the ability of decision-makers to allocate resources and make adjustments to the process more assertively. Indeed, some patents identified in our sample potentially improve decision-making. However, as Boutkhoum et al. (2017) pointed out, selecting the most appropriate solution for Big Data projects is complex and requires a comprehensive evaluation process.

Patent CN106096733A presents an analytical model that evaluates Big Data for water conservation. This method applies an analytical hierarchy process and a fuzzy mathematical theory to Big Data, comprehensively considering various schemes of water service. Once the schemes are classified, the best decision-making scheme is selected. Another patent (CN111444446A) refers to a method and system for collecting engineering cost data based on Big Data, computer equipment, and a computer-readable storage medium. According to the inventors, the method has the advantage that different decision-makers can master the cost data for each stage in each engineering field conveniently. The total engineering cost is obtained as a whole, and, finally, a reasonable decision is made.

We also highlight patent number CN111652090A, which belongs to the mechanical vision field. This invention discloses a digital management method for planting hemp in China. China's hemp planting companies do not monitor the main control points, and manual operation often generates some hysteresis information. Consequently, the timing of decision-making and deviation correction actions are negatively affected. This invention uses a video library database of plant growth images and monitors the images to build a model library and knowledge base. The system can perform batch image processing and recognition by configuring parameters to assist decision-making and improve productivity.

Big Data in modern businesses allows companies to make faster and wiser decisions and achieve competitive advantages. However, many Big Data projects provide disappointing results that do not meet the needs of the decision-maker. These poor results are because they neglect the decision of these projects, which is an aspect of the decision-making process. (Chiheb et al., 2019). It has been stated that with the growing demand for using Big Data to take advantage of their opportunities, organizations are looking for clear and simple solutions and guidelines for Big Data management (Elgendi & Elragal, 2016). As highlighted by Poletto, Carvalho, and Costa (2017), the use of Big Data will only be successful if there is a data integration strategy for the generation of relevant information and knowledge management.

Finally, we highlight applicants with more than one record in our sample. Altogether nine applicants fit this requirement, including Limited Companies with 11 records, followed by universities with two, a research institute with one, and an individual with one (Table 4).

Patent data presents valuable insights for inventors, engineers, companies, and decision-makers (Akers, 2003). It is also considered an indicator of innovation and reflects technological and scientific changes (Archibugi, 1992). In this sense, the mapping of inventions and, consequently, the applicants can reveal strong indicators of technological trends and, in some cases, mastery of a technique or

Table 4. Applicants with more than one registered patent

Applicants	Patent number	Patent name
Chongqing Socialcredits Big Data Tech Co Ltd	CN109408712A	A method for constructing a multi-dimensional information portrait of a travel agency user
	CN109345348A	Recommendation method of multi-dimensional information portrait based on travel agency users
Electric Power Res Inst	CN109190381A	A detection method for Hadoop security vulnerability
	CN109961376A	Distributed energy storage equipment management and control system and method
H3C Big Data Tech Co Ltd	CN108062399A	Data processing method and device
	CN109344163A	Data verification method, apparatus, and computer-readable medium
Lenovo Beijing Co Ltd	CN108804233A	Memory space recycling method, device, electronic equipment and storage medium
	CN106293929A	Data processing method and first electronic equipment
Shenzhen OneConnect Intelligent Technology Co Ltd	CN111984685A	Data tilt detection method, device, computer equipment and readable storage medium
	CN111984630A	Log association method, device and computer equipment
	CN111158642A	Data construction method and device, computer equipment and storage medium
South China University of Technology	CN111444236A	Mobile terminal user portrait construction method and system based on Big Data
	CN107817787A	Intelligent production line manipulator fault diagnosis method based on machine learning
Xiamen University	CN207423929U	Novel air quality intelligence prediction system
	CN111310868A	Water-writing handwritten character recognition method based on convolutional neural network
Wang Yujue	CN109650186A	Communication cable taking-up and paying-off device for communication Big Data project
	CN109650187A	Communication cable taking-up and paying-off device for communication Big Data project
Xinhua Sanda Data Tech Co Ltd	CN107548151A	Method and device for marking wireless access point (AP)
	CN107547519A	Wireless Big Data deployment method and device

technology. Therefore, both academia and industry must recognize the emerging patterns of big data technologies, which are of primary importance in the growth of data-driven enterprises (Saheb & Saheb, 2020).

5. CONCLUSION

When carrying out an exploratory analysis with patent data to understand the characteristics of patent registration related to Big Data projects, we observed a significant number of patents registered in China. This result indicates substantial importance this country gives to the subject. Notably, the dendrogram analysis and FCA identified three classes of words interpreted as a direction to cloud computing, optimization of solutions, and storage and data sharing structures.

The main limitation of our study was that despite the Espacenet database being one of the primary databases, with more than 120 million patent documents, incremental research with other databases could complement the study sample and possibly produce results complementary to those presented herein.

The present study also found that emerging technologies such as Blockchain and IoT are being used in Big Data project solutions. Moreover, we observed the importance given to solutions that help decision-making in an increasingly data-driven context. Indeed, this is perhaps the most significant practical contribution of this study.

As an academic contribution, this study endorses a body of researchers that have dedicated their academic research to patent documents, making patent studies in the academic environment more noticeable.

Since Big Data research, and more specifically, Big Data projects, are areas of incipient knowledge, there are many opportunities for new research on the topic. This observation is the main implication of this paper. We suggest that future research extend this discussion to Big Data project management through a theoretical lens or even applied research for empirical validation of this study's findings. Based on the identified patent clusters, it is plausible that empirical research could explain the Big Data project phenomenon.

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

REFERENCES

- Ahmed, V., Aziz, Z., Tezel, A., & Riaz, Z. (2018). Challenges and drivers for data mining in the AEC sector. *Engineering, Construction, and Architectural Management*, 25(11), 1436–1453. doi:10.1108/ECAM-01-2018-0035
- Akers, L. (2003). The future of patent information - A user with a view. *World Patent Information*, 25(4), 303–312. doi:10.1016/S0172-2190(03)00086-3
- Ali, S. M. F., & Wrembel, R. (2017). From conceptual design to performance optimization of ETL workflows: Current state of research and open problems. *The VLDB Journal*, 26(6), 777–801. doi:10.1007/s00778-017-0477-2
- Alkhamisi, A. O., & Alboraei, F. (2019). Privacy-aware Decentralized and Scalable Access Control Management for IoT Environment. *Journal of King Abdulaziz University Computing and Information Technology Sciences*, 8(1), 71–84. doi:10.4197/Comp.8-1.7
- Archibugi, D. (1992). Patenting as an indicator of technological innovation: A review. *Science & Public Policy*, 19(6), 357–368. doi:10.1093/spp/19.6.357
- Barham, H., & Daim, T. (2018). Identifying critical issues in smart city big data project implementation. *Proceedings of the 1st ACM/EIGSCC Symposium on Smart Cities and Communities, SCC 2018*. doi:10.1145/3236461.3241967
- Becker, D. K. (2017). Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2320–2330. doi:10.1109/BigData.2017.8258186
- Boutkhoun, O., Hanine, M., Agouti, T., & Tikniouine, A. (2016). Selection problem of Cloud solution for big data accessing: Fuzzy AHP-PROMETHEE as a proposed methodology. *Journal of Digital Information Management*, 14(6), 368–382.
- Boutkhoun, O., Hanine, M., Agouti, T., & Tikniouine, A. (2017). A decision-making approach based on fuzzy AHP-TOPSIS methodology for selecting the appropriate cloud solution to manage big data projects. *International Journal of Systems Assurance Engineering and Management*, 8(S2), 1237–1253. doi:10.1007/s13198-017-0592-x
- Camargo, B. V. (2005). *Alceste: Um programa informático de análise quantitativa de dados textuais*. E. Universitária, Ed.
- Chiheb, F., Boumahdi, F., & Bouarfa, H. (2019). A conceptual model for describing the integration of decision aspect into big data. *International Journal of Information System Modeling and Design*, 10(4), 1–23. doi:10.4018/IJISMD.2019100101
- Clarke, N. S. (2018). The basics of patent searching. *World Patent Information*, 54, S4–S10. doi:10.1016/j.wpi.2017.02.006
- Dutta, D., & Bose, I. (2015). Managing a big data project: The case of Ramco cements limited. *International Journal of Production Economics*, 165, 293–306. doi:10.1016/j.ijpe.2014.12.032
- Elgendi, N., & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, 100, 1071–1084. doi:10.1016/j.procs.2016.09.251
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–53.
- Ferrag, M. A., Derdour, M., Mukherjee, M., Derhab, A., Maglaras, L., & Janicke, H. (2018). Blockchain technologies for the internet of things: Research issues and challenges. *ArXiv*, 6(2), 2188–2204.
- Gantz, J., & Reinsel, D. (2011). Extracting Value from Chaos State of the Universe: An Executive Summary. *IDC IVView*, (June), 1–12. Retrieved from <http://idcdocserv.com/1142>
- Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in Big Data analytics projects. *2015 Americas Conference on Information Systems, AMCIS 2015*, 1–14.
- Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal patent analysis for big data technology. In *PICMET 2017 - Portland International Conference on Management of Engineering and Technology: Technology Management for the Interconnected World, Proceedings* (pp. 1–8). doi:10.23919/PICMET.2017.8125461

- Huang, S., & Chaovilitwongse, W. A. (2015). Computational Optimization and Statistical Methods for Big Data Analytics: Applications in Neuroimaging. *The Operations Research Revolution*, (November), 71–88. doi:10.1287/educ.2015.0135
- Huang, Y., Youtie, J., Porter, A. L., Robinson, D. K. R., Cunningham, S. W., & Zhu, D. (2016). Big data and business: Tech mining to capture business interests and activities around big data. In *Proceedings - 2016 IEEE International Conference on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016* (pp. 145–150). doi:10.1109/BDCloud-SocialCom-SustainCom.2016.32
- Jianmin, W., Wenbin, Z., Tongrang, F., Shilong, Y., & Hongwei, L. (2020). An improved join-free snowflake schema for ETL and OLAP of data warehouse. *Concurrency and Computation*, 32(23), 1–11. doi:10.1002/cpe.5519
- Jun, S., Park, S., & Jang, D. (2015). A Technology Valuation Model Using Quantitative Patent Analysis: A Case Study of Technology Transfer in Big Data Marketing. *Emerging Markets Finance & Trade*, 51(5), 963–974. doi:10.1080/1540496X.2015.1061387
- Lara, J. A., De Sojo, A. A., Aljawarneh, S., Schumaker, R. P., & Al-Shargabi, B. (2020). Developing big data projects in open university engineering courses: Lessons learned. *IEEE Access : Practical Innovations, Open Solutions*, 8, 22988–23001. doi:10.1109/ACCESS.2020.2968969
- Majumdar, J., Naraseyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: Application of big data. *Journal of Big Data*, 4(1), 20. Advance online publication. doi:10.1186/s40537-017-0077-4
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 4. PMID:23074865
- Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7), 1497–1516. doi:10.1016/j.adhoc.2012.02.016
- Moore, J. (2017). Data Visualization in support of executive decision making. *Journal of Information, Knowledge, and Management*, 12, 125–138. doi:10.28945/3687
- Mousannif, H., Sabah, H., Doujji, Y., & Sayad, Y. O. (2014). From big data to big projects: A step-by-step roadmap. *Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014*, 373–378. doi:10.1109/FiCloud.2014.66
- Mousannif, H., Sabah, H., Doujji, Y., & Sayad, Y. O. (2016). Big data projects: Just jump right in! *International Journal of Pervasive Computing and Communications*, 12(2), 260–288. doi:10.1108/IJPC-04-2016-0023
- Olsson, N. O. E., & Bull-Berg, H. (2015). *International Journal of Managing Projects in Business*. *International Journal of Managing Projects in Business*, 8. 10.1108/IJMPB-09-2014-0063
- Poleto, T., De Carvalho, V. D. H., & Costa, A. P. C. S. (2017). The full knowledge of big data in the integration of interorganizational information: An approach focused on decision making. *International Journal of Decision Support System Technology*, 9(1), 16–31. doi:10.4018/IJDSST.2017010102
- Pondel, J., & Pondel, M. (2016). The concept of project management platform using BI and big data technology. *ICEIS 2016 - Proceedings of the 18th International Conference on Enterprise Information Systems*, 1, 166–173. doi:10.5220/0005834601660173
- Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *Journal of Big Data*, 7(1), 12. Advance online publication. doi:10.1186/s40537-020-00287-9
- Saltz, J., Shamshurin, I., & Connors, C. (2017). Business Information Systems Workshops. *Lecture Notes in Business Information Processing*, 183–195. doi:10.1007/978-3-319-52464-1_17
- Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. *Proceedings - 2016 IEEE International Conference on Big Data*, 2872–2879. doi:10.1109/BigData.2016.7840936

- Salviati, M. E. (2017). *Manual do Aplicativo Iramuteq* (versão 0.7 Alpha 2 e R Versão 3.2.3). Retrieved from <http://iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>
- Singh, N. (2019). Big data technology: Developments in current research and emerging landscape. *Enterprise Information Systems*, 13(6), 1–31. doi:10.1080/17517575.2019.1612098
- Teixeira, A. A. C., & Queirós, A. S. S. (2016). Economic growth, human capital and structural change: A dynamic panel data analysis. *Research Policy*, 45(8), 1636–1648. doi:10.1016/j.respol.2016.04.006
- Wang, W. (2020). Data analysis of intellectual property policy system based on Internet of Things. *Enterprise Information Systems*, 14(9-10), 1–19. doi:10.1080/17517575.2020.1712744
- Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13–53. doi:10.1080/17538947.2016.1239771
- Yang, Z., Zhang, Z., Zhang, S., Wang, J., Lin, H., & Zeng, B. (2017). Article. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 2116–2122). IEEE. doi:10.1109/FSKD.2017.8393098

APÊNDICE C

Documento confidencial, pois, no momento da defesa, o artigo ainda estava em processo de submissão. Disponibilizado apenas aos professores que compuseram a banca de defesa.