

MINERAÇÃO DE DADOS ENQUANTO FERRAMENTA DE PESQUISA: UMA INVESTIGAÇÃO NOS PROGRAMAS DE PÓS-GRADUAÇÃO

DATA MINING AS A RESEARCH TOOL: AN INVESTIGATION IN GRADUATE PROGRAMS

JOÃO CLÁUDIO GRYSZSZYN

Universidade Estadual do Centro-Oeste (UNICENTRO)

E-mail: joaoclaudiogrys@yahoo.com

MYLLER GOMES

Universidade Estadual do Centro-Oeste (UNICENTRO)

E-mail: myller@unicentro.br

RESUMO

No presente contexto social, o volume de dados disponíveis vem aumentando, as técnicas de mineração de dados se apresentam como uma solução para se obter informações precisas e rápidas em meio a essas grandes quantidades de dados. Considerando a necessidade das instituições de ensino superior a estarem atualizadas com essas novas tendências, a presente pesquisa tem como objetivo central demonstrar a aplicabilidade prática e o contexto pelo qual as teses de doutorado em programas de pós-graduação nota 7 na avaliação CAPES e sua relação a mineração de dados. O número total de teses analisadas foi de 58 documentos. Com a análise, percebeu-se que a USP é a instituição responsável pelo maior número de publicações. E o programa em que mais defendeu teses foi o de Ciências da computação. Quanto a aplicabilidade prática, as teses foram aplicadas em um maior número nas áreas de educação, informática e análise de características de solos. Sendo assim a mineração de dados é uma técnica capaz de explorar conjuntos de dados, considerando sua relevância para se obter informações subjacentes no cerne do processo de descoberta do conhecimento.

Palavras-chave: Mineração de dados; Pós-graduação; Teses.

ABSTRACT

In the present social context, the volume of available data is increasing, data mining techniques are presented as a solution to obtain accurate and fast information in the midst of these large amounts of data. Considering the need for higher education institutions to be up-to-date with these new trends, the main objective of this research is to demonstrate the practical applicability and the context in which doctoral these in graduate programs grade 7 in the CAPES assessment and their relationship to data mining. The total number of theses analyzed was 58 documents. With the analysis, it was noticed that USP is the institution responsible for the largest number of publications. And the program in which defended the most theses was Computer Science. As for practical applicability, the theses were applied greater number in the areas of education, informatic and analysis of soil characteristics. Thus, data mining is a technique capable of exploring data sets, considering their relevance to obtain underlying information at the heart of the knowledge discovery process.

Keywords: Data mining; Postgraduate studies; Theses.

1 INTRODUÇÃO

As Instituições de Ensino Superior (IES) exercem grande influência no desenvolvimento econômico das regiões em que estão inseridas. Essas instituições, em especial as Universidades, fornecem a função de ensino, pesquisa e extensão, e quanto mais o território explorar estas práticas, maior a tendência de desenvolvimento socioeconômico. Vale destacar que quanto mais interações os agentes econômicos, tanto privadas como públicos, tiverem com as IES, mais transferência de tecnologia e inovação vai ocorrer. Entretanto, as Universidades devem estar atualizadas com as novas necessidades das empresas e do mercado, pois caso contrário não haverá essa disponibilização de conhecimento (REGO; CALEIRO, 2012).

Por outro lado, no contexto das organizações, a complexidade da globalização faz com que a informação seja uns dos bens mais valiosos que uma organização possui, ou seja, ter informações rápidas e precisas é de grande importância para sua estratégia corporativa. Diante desse contexto a tecnologia da informação proporciona ferramentas e metodologias para as organizações lidarem com essa gama de informações, e assim se manterem ativas e competitivas no mercado em constante transformação (COSTA *et al.* 2009). Para isto, existem as técnicas de Mineração de Dados, que segundo Gaya *et al.* (2017), é um processo que permite extrair informações implícitas, que sejam a princípio desconhecidas, mas que são potencialmente relevantes com base no objetivo da investigação.

Analisando o contexto geral e, considerando que as IES devem sempre estar atualizadas com as necessidades do mercado, que a Mineração de Dados faz parte desse novo contexto econômico e o uso da bibliometria é importante para a identificação de produção das Universidades, o objetivo do presente estudo é analisar o contexto das teses de Doutorado publicadas entre os anos de 2015 a 2018, que possuem relação com a campo da Mineração de Dados. Com atenção especial ao seu contexto empírico de aplicação.

2 DEFINIÇÕES E APLICABILIDADES DA MINERAÇÃO DE DADOS

O termo Mineração de Dados (MD) originou-se pelo fato de haver uma semelhança entre a extração de minérios valiosos na natureza, com a busca por informações valiosas em grandes volumes de dados. Pois ambos os processos necessitam de que haja uma divisão de grande quantidade de materiais ou, uma análise, para que se possa encontrar o que realmente

tem valor (CANCHUMANI, 2015). “A mineração de dados trata de extrair ou minerar conhecimento de grandes volumes de dados” (SANTOS, 2015, p. 95).

Os dados que são capturados em nosso ambiente, são a base para obter um modelo do universo em que estamos inseridos. A forma tradicional de transformar dados em informação geralmente se ampara na análise manual. Com o passar do tempo essa forma de análise, se torna subjetiva, lenta e cara diante do aumento dos bancos de dados. Sendo assim a solução é automatizar parcialmente o processo de análise de dados (FAYYAD; SHAPIRO; SMYTH, 1996a).

Extrair informações em um grande volume de dados é uma tarefa difícil, para facilitar esse processo, o reconhecimento de padrões existentes, facilita a tomada de decisão no gerenciamento de sistemas (MONZANI, 2016). Ao longo do tempo a ideia de encontrar padrões úteis obteve várias nomenclaturas, entre essas, está a extração de conhecimento, mineração de dados, processamento de padrões, coleta de informações etc. (FAYYAD; SHAPIRO; SMYTH, 1996a).

A MD surge com a evolução e sofisticação da área da computação, ligada com o desafio de lidar com grandes volumes de dados que estão armazenados em bancos, por meio da identificação de padrões. Esse desafio surgiu para preencher a lacuna existente em técnicas antigas de análise, ou seja, surgiu para atender as novas características dos dados, sejam elas em relação ao formato, complexidade, armazenamento, atributos ou distribuição (NHACUONGUE, 2015).

Assim a MD visa encontrar padrões existentes e com isso contribuir para a análise de um grande conjunto de dados. Como por exemplo, uma de suas tarefas é a predição, em que, por meio do encontro de padrões é almejado predizer o comportamento de uma entidade, baseando-se em características de sua história (SILVEIRA JÚNIOR, 2018).

Há muito tempo existe a prática de aprendizado a partir de dados, mas o interesse pela MD começou recentemente em nossa história e se tornou rapidamente intenso. A principal causa desse fato é que o gerenciamento de banco de dados começou há pouco tempo. Dados em grandes quantidades, estão presentes em sistemas de gerenciamento de banco de dados (FRIEDMAN, 1997). Desde 1990 o acesso a bases de dados foi facilitado pela tecnologia da informação e comunicação, permitindo assim o processamento com um viés de descoberta do conhecimento (CANCHUMANI, 2015).

Porém, de acordo com as ideias de Friedman (1997) a MD é um campo cuja definição não é exata e, depende do ponto de vista do autor que está a definindo. O que pode ser

comprovado a partir das várias definições que esse autor encontrou na literatura, conforme apresentado no Quadro 1.

QUADRO 1 – Conceitos de mineração de dados

AUTOR	CONCEITO
Fayyad, Shapiro e Smyth (1996a)	A Mineração de Dados é o processo não trivial de identificar dados válidos, novos potencialmente úteis e, finalmente, sob padrões aceitáveis nos dados.
Zekulin (1997)	A Mineração de Dados é o processo de extrair informações previamente desconhecidas, compreensíveis e acionáveis de grandes bancos de dados e usá-las para tomar decisões cruciais de negócios.
Ferruzza (1996)	A Mineração de Dados é um conjunto de métodos usados no processo de descoberta de conhecimento para distinguir relacionamentos e padrões anteriormente desconhecidos nos dados.
John (1997)	A Mineração de Dados é o processo de descobrir padrões vantajosos nos dados.
Parsaye (1997)	A Mineração de Dados é um processo de suporte à decisão em que procuramos em grandes bancos de dados padrões desconhecidos e inesperados de informações.

Fonte: Adaptado de Friedman (1997).

Esse termo está sendo utilizado principalmente por comunidades de sistemas de informação gerencial e analistas de dados. Teve também, uma significativa usualidade no campo de banco de dados. Para enfatizar a ideia de que o conhecimento é oriundo de dados, foi usada a frase “*Knowledge Discovery in Databases*” (KDD), ou para o português, “Descoberta de Conhecimento em Banco de Dados” no primeiro *Workshop* KDD em 1989. Vale destacar que, a MD está presente nesse processo (FAYYAD; SHAPIRO; SMYTH, 1996a).

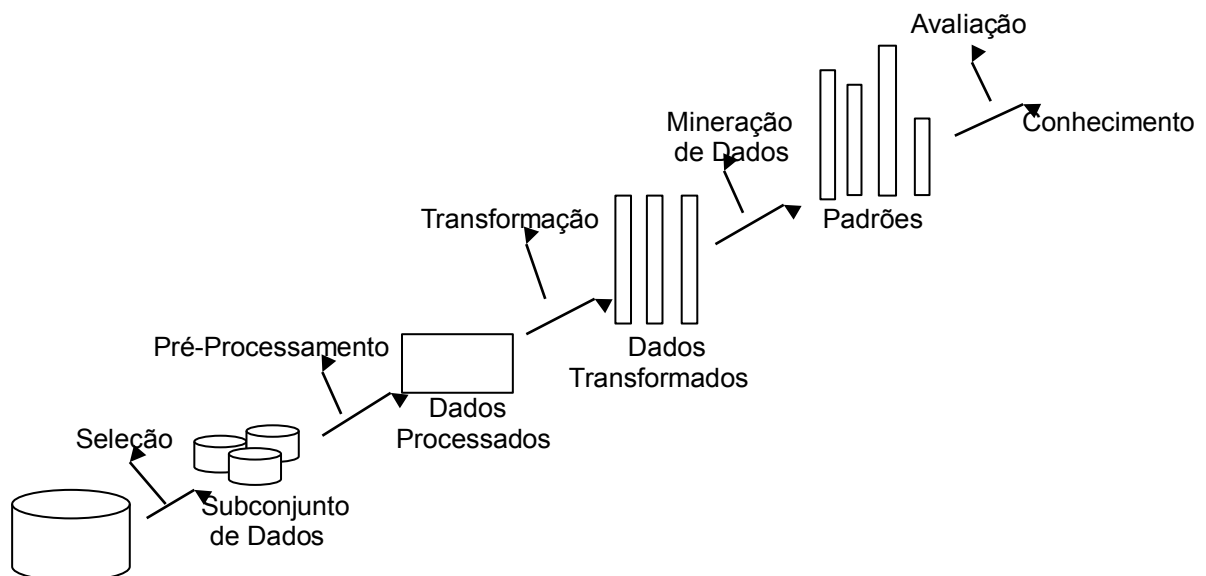


Figura 1 – Processo KDD

Fonte: Adaptado de Fayyad, Shapiro e Smyth (1996b).

Comumente o termo Mineração de Dados é confundida como sinônimo do processo KDD (SANTOS, 2015). O processo KDD está relacionado a descoberta de conhecimento de uma forma geral, no qual a Mineração de Dados é somente uma parte desse processo. Sendo que o processo envolve outras etapas e a Mineração de Dados é apenas uma delas (FAYYAD; SHAPIRO; SMYTH, 1996a; SANTOS, 2015). Dentro do processo de descoberta de conhecimento, a Mineração de Dados é a parte mais importante da análise dos dados (MANHÃES, 2015).

O processo de Descoberta de Conhecimento em Banco de Dados ou processo KDD envolve o pré-processamento dos dados, a mineração e o por último a apresentação dessas informações sob determinadas regras, no qual a MD fica encarregada da principal fase do processo, ou seja, ela possui a função de identificar um método capaz de encontrar padrões existentes nos dados (MONZANI, 2016).

Muitas vezes a MD também é confundida com processos de recuperação de informação, mas na verdade, são duas áreas diferentes em que uma pode complementar a outra. A recuperação da informação está relacionada com técnicas para localizar, acessar e organizar informações. E a Mineração de Dados possui um papel que vai além, que é converter dados brutos em informações úteis, ou descobrir conhecimento em banco de dados (NHACUONGUE, 2015).

3 TÉCNICAS DE MINERAÇÃO DE DADOS: UMA VISÃO GERAL

Para explorar essa grande quantidade de dados a MD consiste em técnicas automáticas que auxiliam nesse processo, com objetivo de descobrir novas relações e padrões entre esses dados, sendo que se fossem analisados “a olho nu”, pelo ser humano, não seriam facilmente descobertos (CARVALHO, 2005). Fayyad, Shapiro e Smyth (1996a) concordam com essa definição ao afirmarem que a MD consiste na aplicação de algoritmos para a extração de padrões presentes nos dados.

As várias técnicas que são utilizadas, devem ser escolhidas de acordo com a situação dos dados. Ou seja, com base nos tipos de dados e no conhecimento que se almeja disponibilizar a partir desses dados. Sendo assim, somente a partir da identificação das características dos dados, é que se pode escolher a melhor técnica para analisá-los e, descobrir padrões e informações a partir destes (MANHÃES, 2015).

Várias técnicas de MD podem ser utilizadas juntas, entre essas técnicas está a árvore de decisão, algoritmo de priori e agrupamento. A árvore de decisão está relacionada com o

processo de classificação de dados dentro de processo de tomada de decisão. Já o algoritmo de priori busca essencialmente por padrões nos dados. E os algoritmos de agrupamentos podem ser por deslocamento médio e por número de grupos (MONZANI, 2016).

Existem duas categorias em que são divididas as técnicas de MD que, são as supervisionadas e as não supervisionadas (TAN; STEINBACH; KUMAR, 2009). As técnicas supervisionadas possuem como objetivo elaborar modelos para definir um atributo especificado. E as técnicas não supervisionadas visam encontrar padrões e tendências nas bases de dados (PELOIA, 2015).

Existem também algumas tarefas de MD como, análise exploratória de dados, recuperação de conteúdo, classificação, mineração de associação, entre outras tarefas. Para desenvolver cada uma dessas tarefas existe um, ou, vários algoritmos, que podem ser aplicados em uma ou mais tarefas (HRISTIDIS *et al.*, 2010).

Existem diversas áreas de aplicação do processo KDD nos negócios, finanças, *marketing*, detecção de fraude, telecomunicações, manufatura etc. Também pode ser aplicada em outras áreas como a astronomia e saúde (FAYYAD; SHAPIRO; SMYTH, 1996a). Pode ser utilizado no âmbito comercial ou científico (MONZANI, 2016). Os avanços das técnicas de Mineração de Dados estimulam o desenvolvimento de várias áreas da ciência, sendo assim as pesquisas relacionadas a inteligência computacional possui caráter multidisciplinar (FERREIRA, 2018). No Quadro 2 são apresentados alguns exemplos de aplicações de técnicas de Mineração de Dados.

Quadro 2 – Exemplos de aplicabilidade de MD

SITUAÇÕES DE APLICABILIDADE DE MD
MD vem sendo usada para a gestão de desastres naturais, em que com a utilização das tarefas de MD nessa área tem como objetivo encontrar padrões, correlações e tendências nos dados, para assim entender mudanças e poder evitar recorrência de futuros fenômenos indesejáveis (HRISTIDIS <i>et al.</i> , 2010).
Os autores Pradhan <i>et al.</i> (2008) utilizaram a mineração de dados para mapear áreas de risco de deslizamento de terra por meio da análise de fotos de obtidas por satélite.
Ferreira (2018) utilizou a MD na indústria petrolífera para identificar padrões de dados multidimensionais dessa indústria. Pois nesse contexto o a precisão e eficiência dos dados são importantes devidos aos altos valores implicados. E pequenas otimizações no consumo de óleo de gás geram efeitos significativos para a economia. Assim a MD permite a decisão baseadas em dados, e oferece acesso efetivo e rápido a grandes bancos de informações.

Fonte: Dados da pesquisa (2022).

Conforme observado a MD apresenta semelhanças nas suas definições, mas em essência a sua ligação está no desenvolvimento do processo de descoberta do conhecimento por meio de técnicas analíticas avançadas.

4 METODOLOGIA

A presente pesquisa, com o objetivo de analisar o contexto empírico em que os estudos de doutorado, possuem relação com a MD, considerou-se na coleta dados, somente as teses que possuem relação com o tema e que foram publicadas no território brasileiro. Também foi realizada classificação manual, que excluiu as teses publicadas em línguas estrangeiras, devido a limitação de programas tradutores que realizem uma tradução compreensível. Sendo assim serão consideradas somente as teses publicadas no idioma português brasileiro.

Foram analisadas somente as teses que foram publicadas por meio de IES que possuem nota 7 (avaliação da CAPES). Essa nota é atribuída aos programas de pós-graduação das IES, e é realizada pela CAPES, considerando alguns critérios específicos. Tal escolha pelas publicações exclusivas de IES com nota 7, é pelo fato de que essa é a nota mais alta da avaliação realizada pela CAPES em programas de pós-graduação. Ou seja, com esse critério, é possível focar os objetivos propostos somente no âmbito das publicações das IES mais renomadas do Brasil. Para encontrar os resultados mais atualizados referentes aos objetivos do estudo, foram consideradas somente as publicações do ano de 2015 a 2019. Essa delimitação temporal é definida considerando a necessidade apresentar uma realidade mais recente das teses.

O processo de identificação das Teses, consistiu-se em 2 etapas, na primeira foram identificadas as IES que possuem nota 7, no site do “Cursos recomendados/reconhecidos” da CAPES. E a segunda etapa identificou as Teses, no *site* do “Catálogo de Teses e Dissertações” da CAPES. O passo a passo das etapas 1 e 2, são apresentadas a seguir.

Etapa 1 - Busca pelos “cursos recomendados/reconhecidos” pela CAPES → No campo “consulta” foi usado por “nota” → Inserido na Plataforma Sucupira usou-se a nota “7” → Busca no nome da área de avaliação, para encontrar as áreas do conhecimento → Busca no nome na área do conhecimento, para identificar o nome das Instituições de Ensino → Coleta do nome das universidades para realização da segunda etapa.

Etapa 2 - Busca pelo “catálogo de teses e dissertação” da CAPES → Inserção no campo de busca pela palavra-chave “mineração de dados”, e clicado no espaço “buscar” → Para refinar os resultados foram usados os filtros: “Doutorado (tese)”; os anos de “2015”, “2016”, “2017”, “2018” e “2019” e as instituições que possuem nota 7 (encontradas na etapa 1). E na sequência, clicar em “refinar meus resultados” → Buscar *link* “detalhes” de cada tese → Baixar a tese para a análise.

5 RESULTADOS

Para encontrar as teses de doutorado em relação a mineração de dados, foram estabelecidas 2 etapas de coleta de dados. Sendo que na primeira seriam encontradas as IES com nota 7, e na segunda etapa seriam encontradas as teses. Na primeira etapa foram encontradas 47 IES com a nota 7 da CAPES, conforme Quadro 3.

Quadro 3 - IES com nota 7 na avaliação da CAPES

(continua)

INSTITUIÇÕES DE NÍVEL SUPERIOR COM NOTA 7/CAPES
INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA (IMPA)
CENTRO BRASILEIRO DE PESQUISAS FÍSICAS (CBPF)
FIOCRUZ (CENTRO DE PESQUISAS RENÉ RACHOU) (FIOCRUZ-CPqRR)
FUNDAÇÃO GETÚLIO VARGAS (RJ) (FGV/RJ)
FUNDAÇÃO GETÚLIO VARGAS (SP) (FGV/SP)
FUNDAÇÃO OSWALDO CRUZ - FIOCRUZ (FIOCRUZ)
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE)
PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO (PUC/SP)
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO (PUC-RIO)
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL (PUC/RS)
UNIVERSIDADE DE BRASÍLIA (UNB)
UNIVERSIDADE DE SÃO PAULO (RIBEIRÃO PRETO) (USP/RP)
UNIVERSIDADE DE SÃO PAULO (SÃO CARLOS) (USP/SC)
UNIVERSIDADE DE SÃO PAULO (USP)
UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO (UERJ)
UNIVERSIDADE DO VALE DO RIO DOS SINOS (UNISINOS)
UNIVERSIDADE ESTADUAL DE CAMPINAS (UNICAMP)
UNIVERSIDADE ESTADUAL DE CAMPINAS, PIRACICABA (UNICAMP-PI)
UNIVERSIDADE ESTADUAL DE LONDRINA (UEL)
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO - INSTITUTO DE FÍSICA TEÓRICA (UNESP-IFT)
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO (ARARAQUARA) (UNESP-ARAR)
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, JABOTICABAL (UNESP-JAB)
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, PRESIDENTE PRUDENTE (UNESP-PP)

Quadro 3 - IES com nota 7 na avaliação da CAPES

(conclusão)

INSTITUIÇÕES DE NÍVEL SUPERIOR COM NOTA 7/CAPES
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, SÃO JOSÉ DO RIO PRETO (UNESP-SJRP)
UNIVERSIDADE FEDERAL DA BAHIA (UFBA)
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE (UFCG)
UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)
UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)
UNIVERSIDADE FEDERAL DE PELOTAS (UFPEL)
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)
UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)
UNIVERSIDADE FEDERAL DE SANTA MARIA (UFSM)
UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)
UNIVERSIDADE FEDERAL DE SÃO PAULO (UNIFESP)
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)
UNIVERSIDADE FEDERAL DE VIÇOSA (UFV)
UNIVERSIDADE FEDERAL DO CEARÁ (UFC)
UNIVERSIDADE FEDERAL DO PARANÁ (UFPR)
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ)
UNIVERSIDADE FEDERAL DO RIO GRANDE (FURG)
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE (UFRN)
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)
UNIVERSIDADE FEDERAL FLUMINENSE (UFF)
UNIVERSIDADE SÃO FRANCISCO (USF)
USP (CENTRO DE ENERGIA NUCLEAR NA AGRICULTURA) (USP/CENA)
USP (ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ) (USP/ESALQ)

Fonte: Dados da pesquisa (2022).

Na segunda etapa, ao digitar a palavra-chave “mineração de dados”, no buscador “Catálogo de Teses e Dissertações” da CAPES, foram encontrados 1.762 resultados. Logo após aplicar os filtros disponíveis na plataforma para se obter as teses de interesse da pesquisa, foram encontrados 77 documentos que foram possíveis de serem baixados. Mas em uma última seleção, feita de forma manual, foram descartados 19 documentos que eram em

língua estrangeira, ou que eram apenas um resumo da tese, e não o documento completo. Sendo assim, chegou-se a 58 teses.

Com esse mapeamento, foi possível de realizar algumas análises no que diz respeito as aplicações práticas das teses, bem como das suas características contextuais, por meio das quais foram desenvolvidas e publicadas. Em relação a sua aplicabilidade prática, foi possível perceber que a mineração de dados foi aplicada com destaque nas áreas envolvendo educação, informática e análise de características de solos. Conforme apresentado no Gráfico 1 desta pesquisa.

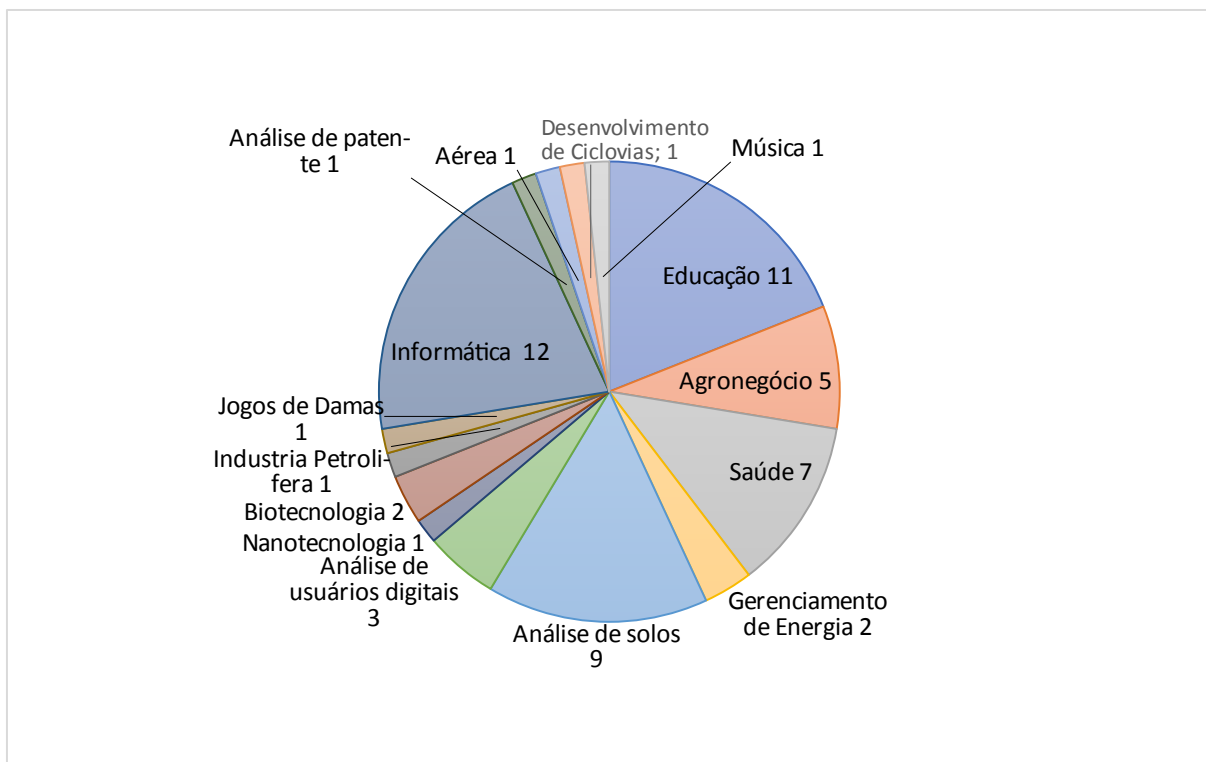


Gráfico 1 - Área de aplicação das teses – 2015 a 2019

Fonte: Dados da pesquisa (2022).

Outro fator relevante a ser considerado são os programas de pós-graduação que são responsáveis pela maior quantidade de publicações. Sendo observado que os programas que mais tiveram teses publicadas no período, são os de ciências de computação e matemática computacional, ciência da computação, engenharia agrícola e engenharia elétrica. Conforme é demonstrado no Quadro 4.

Quadro 4 - Programas de pós-graduação com número de teses – 2015 a 2019

PROGRAMA DE PÓS-GRADUAÇÃO	NÚMERO DE TESES
Informática	12
Informática na Educação	11
Ciência da Computação	6
Engenharia Agrícola	4
Engenharia Elétrica	4
Ciências de Computação e Matemática Computacional	4
Engenharia Mecânica	3
Cardiologia	1
Ciência, Tecnologia E Sociedade	1
Ciências Odontológicas	1
Ciência e Engenharia de Materiais	1
Ciências da Saúde	1
Ciência do Solo	1
Ciência do Sistema Terrestre	1
Computação Aplicada	1
Engenharia Florestal	1
Engenharia de Sistemas E Computação	1
Engenharia Elétrica e da Computação	1
Engenharia de Transportes	1
Geociências Aplicadas	1
Geografia	1
Total	58

Fonte: Dados da pesquisa (2022).

Pode ser observado que nesse caso, há uma concentração de pesquisas em mineração de dados nos programas das ciências tecnológicas, e menos nas áreas das ciências sociais. Uma das possibilidades desse fato ocorrer, é que como visto por Nhacuongue (2015), a MD surgiu devido ao aumento de dados na área da computação. Fayyad, Shapiro e Smyth (1996a) afirmam que existem outras áreas que podem se apropriar das técnicas de MD.

Como a mineração de dados visa extrair informações em grandes quantidades de dados (SANTOS, 2015). As áreas das ciências sociais poderiam explorar mais essas técnicas em suas pesquisas científicas. Pois nas ciências sociais está inserido o curso de administração. Sendo assim, existe uma relevância da MD para o curso de administração, diante da grande

quantidade de dados que o novo contexto social oferece e a importância das empresas se manterem atualizadas e com informações precisas e rápidas.

Vale ressaltar que as análises aqui realizadas, consideram somente as teses que foram possíveis de serem baixadas na íntegra. Essa impossibilidade de acesso ao restante das teses, se dá ao fato de que as mesmas poderiam estar em processo de patenteamento, ou que por algum outro motivo não foram autorizadas as suas divulgações. Sendo assim as análises foram realizadas somente com as teses possíveis de serem baixadas, que somam um total de 58 documentos.

6 CONSIDERAÇÕES FINAIS

Considerando os objetivos propostos pelo estudo, que consistem em identificar o contexto que as teses de doutorado, em relação a mineração de dados, foram publicadas. Em específico a identificação das universidades com programas de pós-graduação de nível 7, e também encontrar as teses dessas IES, e após realizar a análise do seu conteúdo, foi observado que existem 47 IES que possuem programas de pós-graduação com nota 7. E que após as buscas realizadas chegou-se ao total de 58 teses, e que todas foram analisadas.

Nas análises realizadas nas teses foi verificado que, em relação as IES que mais publicam teses de doutorado em relação a mineração de dados são a USP, UFSM e UFRGS. Sendo assim, nota-se que as pesquisas estão sendo mais publicadas na região Sul e Sudeste, pois são onde se encontram tais universidades. Outro fato identificado é que dentre as IES, que possuem teses analisadas, 19 são instituições públicas. Assim é reforçada a importância do ensino público gratuito e de qualidade para o desenvolvimento da ciência impactando na sociedade em termos econômicos, sociais e ambientais.

Em relação a análise dos programas de pós-graduação, por meio dos quais as teses foram desenvolvidas, se observa que a maioria das teses foram publicadas por programas de pós-graduação ligados a computação e informática. Ficando evidente que a mineração de dados ainda não se expandiu de forma significativa para outros campos, como por exemplo, as ciências sociais, onde está inserido o curso de administração. Isso é um ponto negativo, pois durante o estudo ficou evidente a importância das técnicas de mineração de dados, para se obter informações precisas e rápidas.

Vale ressaltar que, na pesquisa do estado da arte, foi observado uma incidência de artigos relacionados a mineração de dados aplicados na área do sensoriamento remoto. Já nas

teses, foi observado uma incidência de teses aplicadas na área da educação, análise de dados, e análise de características de solos. Com esses resultados, pode ser observado que a aplicabilidade prática dos estudos em relação a mineração de dados pode variar de acordo com o tipo de pesquisa. Sendo claro que os resultados aqui encontrados, com base em teses de doutorado, não podem ser generalizados para outros tipos de documentos científicos.

A presente pesquisa durante seu processo de execução, passou por algumas limitações. Uma delas diz respeito a disponibilidade parcial das teses no catálogo de teses e dissertações. Pois grande parte dessas teses não eram disponibilizadas por questão de direitos autorais. Sendo que também algumas teses não eram disponibilizadas de forma integral, sendo disponibilizado somente a carta de apresentação, ou o resumo da tese, devido ao seu processo de patenteamento. Outra limitação desse estudo é que foi considerado somente teses publicadas entre os anos de 2015 a 2019, e de IES com programas de pós-graduação de nível 7. Impossibilitando a generalização dos resultados aqui encontrados, para todas as IES avaliadas pelas CAPES.

REFERÊNCIAS

ARAÚJO, R. F; ALVARENGA, L. A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007. **Revista Eletrônica Biblioteconomia**, v. 16, n. 31, p. 51-70, 2011.

CARVALHO, L. A. V. **Data mining**: a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Editora Ciência Moderna Ltda., 2005.

CANCHUMANI, R. M. L. **Domínios Científicos na UFRJ**: mapeamento de áreas de conhecimento. 185 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro-RJ, 2015.

COSTA, C. N. *et al.* Descoberta de conhecimento em bases de dados. **Revista Eletrônica da Faculdade São José**. v. 2, p. 1-10, 2009.

FAYYAD, U; SHAPIRO, G. P; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, 1996a.

FAYYAD, U; SHAPIRO, G. P; SMYTH, P. Knowledge discovery and data mining: towards a unifying framework. **All rights reserved**. p. 82-88. 1996b.

FERREIRA, F. G. **Classificação de petróleo**s. 105 f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, 2018.

FRIEDMAN, J. H. **Data mining and statistics: what's the connection?** 1997. Disponível em: <http://www-stat.stanford.edu/~jhf/ftp/dm-stat.pdf>. Acesso em: 25 de out 2021.

GAYA, J. F. O. *et al.* Análise da mudança de processo seletivo para ingresso no ensino superior usando ferramentas de mineração de dados. **Revista Scientia Plena**. v. 13, n. 04, p. 1-10, 2017.

HRISTIDIS, V. *et al.* Survey of data management and analysis in disaster situations. **The Journal of Systems and Software**. v. 83, p. 1701-1714, 2010.

MANHÃES, L. M. B. **Predição do desempenho acadêmico de graduandos utilizando mineração de dados educacionais**. 140 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, 2015.

MONZANI, R. C. **Monitoramento não invasivo e identificação do consumo de energia individual de cargas residenciais**. 148 p. Tese (Doutorado em Engenharia Elétrica) – Universidade Estadual de Campinas, Campinas-SP, 2016.

NHACUONGUE, J. A. **O campo da Ciência da Informação: contribuições, desafios e perspectivas da mineração de dados para o conhecimento pós-moderno**. 194 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo-SP, 2015.

PELOIA, P. R. **Indução de regras de decisão para análise descritiva de padrões de produtividade em talhões de cana-de-açúcar**. 122 f. Tese (Doutorado em Planejamento e Desenvolvimento Rural Sustentável) - Universidade Estadual de Campinas, Campinas-SP, 2015.

PRADHAN, B. *et al.* Application of a data mining model fo landslide hazard mapping. *In: PROCEEDINGS OF ISPRS*, 37., Beijing, 2008. **Anais [...]**. Beijing: ISPRS, p. 187-196, 2008.

REGO, M. C.; CALEIRO, A. Em torno do contributo das instituições de ensino superior para a dinâmica regional de crescimento económico. **DRd – Desenvolvimento Regional em debate**, n. 2, p. 124-138, 2012.

SANTOS, D. B. S. **Concepção de uma solução escalável para maximização de influência ciente de tópicos em redes sociais**. 102 f. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Campina Grande, Campina Grande-PE, 2015.

SILVEIRA JUNIOR, C. R. **Mineração de regras de associação espaço-temporais temáticas aplicada a imagens de explosões solares**. 140 f. Tese (Doutorado em Ciência da Computação) – Universidade Federal de São Carlos, São Carlos-SP, 2018.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining**. Rio de Janeiro: Ciência Moderna, 2009.