

A IMPLEMENTAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS NA ÁREA DE CONTROLE EXTERNO DAS CONTAS PÚBLICAS

Mônica Pagno da Silva da Rosa¹;

Aires José Rover²;

Edimeia Liliani Schnitzler³;

Neri dos Santos⁴;

Victor Corrêa Faraon⁵

Abstract: *this article presents research carried out, based on a systematic literature review, in the Scopus database, based on data mining techniques, with the objective of verifying the predictions about the compliance of Santa Catarina municipalities, in what concerns respect for the goals determined by the legislation in force. The results allowed us to conclude that it is feasible to predict whether a municipality will extrapolate the goals imposed by legislation, information that would allow the State Court of Auditors, as an external control body of public accounts, to act proactively, making inspection activities more effective.*

Keywords: *Data Mining; Public Organizations; Brazil*

Resumo: o presente artigo apresenta uma pesquisa realizada, a partir de uma revisão sistemática de literatura, na base de dados *Scopus*, baseada nas técnicas de mineração de dados, com o objetivo de verificar as previsões acerca do cumprimento dos municípios catarinenses, no que diz respeito às metas determinadas pela legislação vigente. Os resultados permitiram concluir que é viável prever se um município irá extrapolar as metas impostas pela legislação, informação que permitiria o Tribunal de Contas do Estado, como órgão de controle externo das contas públicas, atuar de forma proativa, tornando mais efetivas as atividades de fiscalização.

Palavras-chave: Mineração de Dados; Organizações Públicas; Brasil

1. INTRODUÇÃO

¹ Programa de Pós-graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-8432-6686>. e-mail: monicapagno@yahoo.com.br

² Programa de Pós-graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0003-1070-5357>. e-mail: aires.rover@gmail.com

³ Programa de Pós-graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-5118-4431>. e-mail: edililica@yahoo.com.br

⁴ Programa de Pós-graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-0356-6750>. e-mail: nerisantos@gmail.com

⁵ Programa de Pós-graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-0766-0058>. e-mail: victorfaraon@hotmail.com

Convivemos com uma avalanche de dados, provenientes das mais diversas fontes, e o mesmo acontece com as organizações públicas e privadas. Com isso, se torna mais trabalhoso, e mais dificultoso, extrair dados que possam se transformar em informações relevantes.

No intuito de facilitar essa tarefa a computação nos oferece técnicas automatizadas de análise de dados, dentre elas, o presente artigo destaca a mineração de dados. Pois, de acordo com Fraga (2017), é frequente a análise de grande volume de dados utilizando tarefas de aprendizagem de máquina e *big data*.

Nesta perspectiva, Fayyad *et al.* (1996) definem mineração de dados como a aplicação de algoritmos que descubrem e analisam os dados, sendo essa uma parte do processo de descoberta de conhecimento em bancos de dados (KDD), que diante de limitações computacionais aceitáveis produz uma exposição de padrões sobre os dados utilizados.

Fraga (2017) utilizou algoritmos de mineração de dados nos dados do Tribunal de Contas do Estado da Paraíba para averiguar indícios de conluio nas licitações dos municípios. Os resultados do estudo forneceram informações importantes que podem ser utilizadas como norteadoras para a otimização dos processos de auditoria e fiscalização pelos órgãos responsáveis. A Constituição Estadual de Santa Catarina em seu artigo 59 *caput*, incisos e parágrafos ditam competências, a semelhança de outros Tribunais de Contas de nosso País.

Portanto, o objetivo deste artigo é realizar uma revisão sistemática de literatura para identificar quais são as técnicas de mineração de dados que podem ser úteis para utilização dos tribunais de contas brasileiros.

2. FUNDAMENTAÇÃO TEÓRICA

Um modelo tradicional de transformação de dados em informação, e depois em conhecimento, segundo Fayyad *et al.* (1996) incidia em um processamento feito manualmente de todas as informações coletadas por especialistas que, na sequência, produziam relatórios para serem analisados. Na maior parte dos casos se torna-se impraticável devido a enorme volume de dados. Segundo o mesmo autor o KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento nas Bases de Dados) é uma tentativa de solucionar o problema causado pela chamada “era da informação”: a sobrecarga de dados. Fayyad *et al.* (1996) definem KDD como “um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis”.

Hoje, diversos processos definem e uniformizam as fases e atividades da Mineração de dados, bem como, evoluiu para a Mineração de Estruturas Complexas. No início a mineração de dados foi idealizada para utilizar-se de repositórios estruturados de dados como: Banco de Dados, *Data Warehouse*, Arquivos e outros. Entretanto, atualmente os dados são representados por diversos formatos: Estruturado, Não estruturado, Espacial e Temporal, Multimídia, Web etc. E assim, cada vez mais, há a necessidade de mineração desses dados. Com isto, uma área que vem sendo bastante pesquisada é a Mineração de Dados em estruturas complexas. Han e Kamber (2012), citam algumas dessas estruturas: mineração de relacionamentos, mineração na *internet*.

Na área de mineração de relacionamentos na área de criminalística administrativa pública, esse tipo de mineração, em que as redes sociais concebem o relacionamento (*links*) entre os institutos envolvidos, tem sido muito utilizada devido a grande quantidade de dados produzido pela *Web* e pela diversidade de padrões que podem ser obtidos (Han e Kamber, 2012). Muitos artigos têm trabalhado com mineração de redes sociais na *internet*, que se utiliza de técnicas de inteligência social, e um dos artigos obtidos nesta pesquisa utiliza as redes sociais (Camilo e Silva, 2009).

Nesse mesmo sentido, Toledo, *et al.*, (2020), valeu-se das redes criminosas na busca de novos questionamentos e respostas, frente a demanda da sociedade por justiça. Foi empregada uma análise inovadora da ocorrência de crimes no Estado de Minas Gerais, seus dados foram coletados em um banco de dados de Inteligência de Segurança Pública, usando técnicas de estatística e de redes multiplex, proporcionando identificar grupos criminosos de indivíduos e a estrutura das organizações criminosas. Recorreu-se de 89% ferramentas de *desktop*; em termos de plataformas *web*, os autores analisam a corrupção em contratos em estratégias públicas e no processo de compra de medicamentos. O caso específico das redes multiplex, segundo o autor, conceitua-se como “diversidade de variedade de configurações de conectividade que os elementos que constituem a rede possuem”.

Battiston, *et al.* (2017) asseverou que as redes multiplex obedecem a um caso particular de redes multicamadas em que a estrutura é conformada pelo mesmo conjunto de nós em todas as camadas, e em que as interações entre eles só ocorrem com uma replicação nas diferentes camadas, oferecendo uma representação real dos acontecimentos.

Torres, *et al.* (2020) afirmam que a “*Transparency Internacional* orçou os custos da corrupção nas compras públicas que atingem entre 20 e 25% da importância descrita no

contrato, podendo atingir a 40–50%”. Em seu estudo analisando distintos “tipos de corrupção como (suborno, desfalque por conluio, apropriação indébita, fraude, abuso de discricção, favoritismo, nepotismo)” e que se utilizam seis tipos de “técnicas de Inteligência artificial (classificação, regressão, agrupamento, predição, detecção de outlier e visualização)”. O mesmo autor assevera sobre as diversas utilizações de técnicas de inteligência social como: Teorema de *Bayes*, redes neurais, *Support Vector Machines* (SVM), árvores de decisão, *Random Forest* e regressão logística e linear, redes convolucionais, teoria dos conjuntos difíceis, gráficos de processamento de linguagem natural, *kmeans*, *AdaBoost* e algoritmos. Utilizaram a metodologia proposta por Torres-Carrion, que utiliza as ferramentas tecnológicas e as metodologias e técnicas de *data mining*.

Outra área com grande demanda por processos de mineração de dados é a Vigilância Sanitária (VS), a exemplo da Agência Nacional de Vigilância Sanitária (ANVISA), que é o assunto do artigo de Voos e Marques (2020), que se utiliza a mineração da *internet* em dados das páginas do *Facebook*, e *internet*.

A mineração da *internet* tem sido muito pesquisada nos últimos tempos e nos artigos selecionados na presente revisão, constatou isso. Pois, grande parte é dos tipos de estruturas complexas e poucos de estruturas simples. Além disso, com o volume gigantesco de dados de *Big Data*, acolhendo diversas necessidades em quase todas as áreas permitem pesquisar diversos conteúdos. A mineração da *internet* (ou *Web Mining*), consiste em minerar as estruturas de ligação, o conteúdo, os padrões de acesso, classificação de documentos, entre outras (Camilo e Silva, 2009).

O artigo de Voos e Marques (2020) trata da VS no *Facebook*. As páginas no *Facebook* da Agência Nacional de Vigilância Sanitária (Anvisa Oficial) e da Vigilância Sanitária do município do Rio de Janeiro (Vigilância Sanitária Rio) foram selecionadas para análise de conteúdo. Houve a demonstração das potências e fragilidades da comunicação do risco sanitário na esfera digital. O aplicativo *Netvizz* foi empregado na mineração dos dados. Fundamentos da análise de redes sociais e da análise de conteúdo guiaram o modelo analítico proposto.

Foram estabelecidas métricas para as páginas selecionadas do *Facebook* durante seis meses, em que foram utilizados o número total de publicações nesse período, a contagem de comentários, compartilhamentos, as reações do público, as curtidas e o engajamento por meio da análise de conteúdo. O trabalho evidencia a necessidade de comunicação digital como

alternativa contemporânea para incrementar ferramentas de proteção social, fortalecer o direito à informação e, conseqüentemente, consolidar o direito à saúde (Voos e Marques, 2020).

Lins *et al.* (2019) utilizaram Mineração de Dados e a Análise Envoltória de Dados (DEA) para desenvolver o seu artigo. DEA é baseada em variáveis quantitativas de mapas conceituais, os quais são aplicados no contexto do mapeamento conceitual, lançando luz sobre os fatores quantitativo e qualitativo que influenciam no desempenho em saúde. A teoria e a prática têm evidenciado que na maioria dos problemas complexos são melhor solucionados usando uma combinação de métodos, sendo os dois, um B difícil e B *soft*. Tal abordagem tem sido denominada de multimetodologia.

Waszak e Sines (2003) argumentam que a multimetodologia já pode ser considerada essencial em alguns programas típicos de educação e saúde. O objetivo foi propor uma metodologia de indicadores de desempenho para amparar a formulação de políticas de saúde brasileiras, utilizando indicadores quantitativos, no caso, mapas conceituais definem a conjuntura, estruturam o problema, trazem diferentes perspectivas das partes interessadas e lançam o levantamento de informações básicas municipais divulgado anualmente pelo IBGE, abarcando 16 temas dos 5.565 municípios brasileiros: recursos humanos, legislação e instrumentos de planejamento municipal, educação, cultura, esporte, habitação, transporte, saúde e meio ambiente, etc.

O artigo de Ullah *et al.* (2020), que trata dos Padrões de espaço temporais de visitantes em parques verdes urbanos por mineração de *Big Data* de mídia social com base em relatórios da Organização Mundial da Saúde (OMS). Este tipo de mineração analisou o número de *check-ins* em vários parques de Xangai usando dados de *check-in* de redes sociais com *geo-tags*, examinaram três anos de dados Geomarcados. Estimativa de densidade do *kernel* (KDE) é um método para considerar a divulgação em qualquer ponto dentro de uma estrutura espacial, independentemente dos limites administrativos. Uma comparação detalhada entre duas áreas povoadas de forma semelhante foi realizada usando o KDE para estudar as disparidades na saúde ou quaisquer outras questões relacionadas à saúde pública.

Foi realizado um estudo descritivo usando dados de mídia social. Analisou-se o conjunto de dados de geo-localização com base no *Weibo* (semelhante ao *Twitter*) em 10 distritos de Xangai/China. Realizou-se uma análise do comportamento de *check-in* que envolveu Metodologia de Análise de Dados (LBSN); pré-processamento e limpeza de dados, análise

espacial e temporal dos dados LBSN e avaliação estatística para fornecer a significância dos dados LBSN (Ullah *et al.*, 2020).

Na mesma linha utilizando mineração de dados geo-espacial, Jayashree *et al.* (2018) desenvolveram um artigo sobre Aplicação do Mapa Cognitivo *Fuzzy* para a previsão geo-espacial de risco de surto de dengue em regiões tropicais do sul da Índia. O trabalho busca classificar o risco de surto de dengue em uma determinada região geográfica por meio do Mapa Cognitivo *Fuzzy* (FCM). Neste estudo, especialistas médicos contribuíram para construir o modelo FCM para avaliar o risco de dengue. Este modelo foi treinado usando o algoritmo de aprendizagem não linear baseado em dados *Hebbian* (DD-NHL), para 71 amostras de dados coletados em áreas suburbanas da cidade de Chennai, uma das regiões tropicais do sul da Índia.

Silva *et al.* (2017) também focaram seu estudo na área da saúde. Os autores utilizaram ferramentas de bioinformática e um *data warehouse* com a aplicação de algoritmos de aprendizado de máquina, com o objetivo de recuperar informações de alta precisão, classificar gêneros e estruturas de leitura aberta de genomas de geminivírus e demarcar espécies. Como resultado, os autores proporcionam a recuperação de informações sobre o geminivírus, como também, a descoberta de conhecimento sobre o mesmo através de um ambiente amigável e de simples utilização (Silva *et al.*, 2017).

Ainda na área da saúde, Von Wangenheim *et al.* (2013) implementaram o aplicativo DICOM SR – Relatório Estruturado (SR) de Imagem Digital e Comunicações em Medicina (DICOM), utilizando vocabulários estruturados em uma rede pública estadual de telemedicina, com o objetivo de atender a demanda por redes de telemedicina em grande escala. O trabalho teve como resultado de seu estudo uma mineração de dados eficiente, tornando possível a rastreabilidade de diagnósticos, morbidades e informações clínicas do paciente (Von Wangenheim *et al.*, 2013).

A área da educação também faz uso da mineração de dados. A maior parte dos dados de uma instituição educacional é armazenada de forma semiestruturada e não-estruturada, por meio de textos, *e-mail*, artigos, documentos (atas, memorandos, ofícios, relatórios e outros). A busca de padrões de conhecimento nestes documentos é muito comum. Porém, nem sempre, os resultados obtidos são precisos: documentos não relacionados, volume muito alto de informações dispensáveis, entre outros. A mineração de textos, visa ajudar neste processo (Camilo & Silva, 2009).

Lima *et al.* (2018), por meio da Análise de Conteúdo (AC) de exames de avaliação de alunos, determinou indicadores de desempenho que são usados para avaliar a qualidade da educação; a mineração de dados no *SysEnade* é realizada usando WEKA 9, que é um *software* de código aberto, desenvolvido em 1993, na Universidade de Waikato, Nova Zelândia. O WEKA possui uma API que permite sua incorporação a qualquer aplicativo, permitindo que as tarefas de mineração de dados sejam realizadas de forma automatizada a partir do próprio *software*. O WEKA usa arquivos ARFF (*Attribute-Relation File Format*) como entrada. Porém, dependendo do algoritmo que será utilizado, outras informações, como atributos de classe, devem ser definidas. Assim, os dados selecionados devem ser transformados e adaptados para serem usados.

Os dados utilizados são obtidos nos exames do ENAD, coletados por 3 anos e a mineração desenvolvida em fases: 1º - Identificação das questões do exame a serem realizadas, sua classificação em tópicos antecipadamente definidos, tal classificação das questões pode ser feita manualmente ou por especialista ou automaticamente usando técnicas de mineração de texto; 2º- Calcula-se para cada aluno uma nota para cada tópico. Isso é feito identificando o resultado do aluno para cada questão do exame, agrupando as questões por tópico e calculando a pontuação conseguida por aquele aluno naquele tópico; 3º- Analisa-se os resultados que nessa fase já estão transformados e agrupados, pode-se aplicar técnicas de análise de dados, como estatística descritivas e inferenciais (Lima *et al.*, 2018).

A área de tecnologia da informação e comunicação também apresenta contribuições na utilização de mineração de dados. Antolínez (2018), utilizou o método de mineração de textos para categorizar a correspondência eletrônica do Departamento Administrativo de *La Función Pública en Colombia*, com o propósito de codificar o conhecimento. Para tanto, o autor classificou as correspondências eletrônicas e utilizou algoritmos de aprendizagem de máquina do tipo *Naive*, *bayesiano*, máquinas de suporte vetorial e *Boosting*, incluindo *benchmarking*. Como resultados Antolínez (2018) obteve uma aplicação semiautomática com uma acurácia superior a 90%, o autor também concluiu que o modelo pode até ser ampliado para um sistema de interação com os cidadãos manejando grandes quantidades de dados em tempo real.

Seguindo na área de tecnologia da informação e comunicação, Brito *et al.* (2018), monitoraram e avaliaram as intervenções de treinamento em tecnologias da informação e comunicação em larga escala, tendo como cenário todas as regiões brasileiras. Em seu estudo, Brito *et al.* (2018) utilizaram a análise de redes sociais e técnicas de mineração de dados para

associar os dados da análise de redes sociais com os dados obtidos das instituições executoras, indicadores locais de infraestrutura e condições socioeconômicas locais. Para identificar a associação entre os atributos das instituições e dos municípios onde operam os telecentros de treinamento, os autores utilizaram redes *bayesianas*. Os autores utilizaram os registros de interações, na plataforma de aprendizagem, entre os participantes durante o treinamento em tecnologias da informação e comunicação e para obtenção dos dados pessoais dos participantes e dados sobre a utilização das tecnologias da informação e comunicação foram aplicados questionários individuais. Como resultado, o estudo possibilitou concluir que ainda existe uma lacuna socioeconômica, cultural, geográfica e demográfica entre aqueles que planejam as intervenções e seus beneficiários, o que afeta a eficácia da intervenção (Brito *et al.*, 2018).

Faria e Monteiro (2018), desenvolveram outro estudo na área de tecnologia da informação e comunicação, com o objetivo de propiciar aos auditores de segurança de rede ou sistemas de prevenção de intrusão subsídios relevantes e compreensíveis à tomada de decisão. Para isso, os autores desenvolveram um trabalho teórico onde utilizaram algoritmos supervisionados para mineração de dados e obtiveram como resultado indicadores, através de matrizes de confusão. A utilização das técnicas citadas se mostrou eficaz na detecção de intrusões, identificando acessos normais e anômalos (Faria & Monteiro, 2018).

Por fim, utilizando a mineração de processos em seu estudo, Gandulfo *et al.* (2016) analisaram o processo de controle de férias de uma organização pública jurídica, o Tribunal Federal. Como os processos apresentaram pouca estruturação, os autores utilizaram algoritmos de mineração de dados para gerar modelos chamados “espaguetes”. O estudo teve como objetivo a redução de custos através da utilização do método de mineração de processos. Como resultado, o estudo apresentou o método aqui mencionado (Gandulfo *et al.*, 2016).

3. PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos utilizados neste artigo são baseados em uma revisão sistemática da literatura sobre tipos de mineração de dados utilizados nas organizações públicas brasileiras. Greenhalgh (1997) afirma que uma revisão sistemática é um apanhado de estudos primários que utiliza métodos e objetivos claros, conduzido por uma metodologia explícita e reprodutível.

A revisão sistemática da literatura foi realizada nas bases de dados *Scopus* e *Web of Science*, na data de 22 de novembro de 2020. Nesta busca foram utilizadas as palavras-chave: “*data mining*” e “*public organizations*” e “*Brazil*”, não foi adotada limitação temporal quanto à data de publicação dos trabalhos e nem ao tipo das publicações. Como resultado foi obtido um total de 44 (quarenta e quatro) publicações.

Foram lidos os títulos, as palavras-chave e os resumos das 44 (quarenta e quatro) publicações para averiguar a relação com o tema pesquisado. A partir desta leitura foi possível eliminar 18 (dezoito) publicações, restando 26 (vinte e seis) publicações para possível acesso e leitura na íntegra. Das 26 (vinte e seis) publicações selecionadas, após obtido acesso e delineamento de acordo com o tema, restaram 16 (dezesesseis) publicações.

As 16 (dezesesseis) publicações restantes foram, devidamente, analisadas considerando-se a seguinte questão norteadora da pesquisa: **quais os tipos de mineração de dados que poderão ser utilizados por técnicos de dados dos Tribunais de Contas Estaduais que efetuam a fiscalização dos órgãos públicos?**

4. ANÁLISE DOS RESULTADOS DA BUSCA SISTEMÁTICA

Uma das áreas públicas que mais necessita de estudos para o controle de riscos é a área da criminalística, da Segurança Pública, que evidencia pesquisas sobre sistemas de prevenção de fraudes e demonstrou áreas de fraudes eletrônicas (e-fraude): com cartão de crédito, telecomunicações, seguro de saúde, seguro de automóveis e leilão *online*. as formas de mineração mais utilizadas na atualidade na área de Segurança Pública, a Mineração de Dados são as em estruturas complexas.

Por meio das técnicas de inteligência social, a ordem para a diversidade global do sistema multiplex, apontou vários crimes. Assim, foi possível executar a prisão de vários indivíduos, ao se utilizar a análise multiplex, que possibilita o planejamento de estratégias preventivas e repressivas por meio da avaliação da corrupção.

A análise de risco de conluio e corrupção usando classificadores *Naive Bayes*, dos autores Balaniuk *et al.* (2013), também trata de corrupção em compras públicas e agências governamentais. Neste trabalho evidencia-se a prática que pode ser utilizada pelos Tribunal de Contas Estaduais e outros Tribunais, já que são as instituições que tratam das fiscalizações em órgãos públicos e possuem bancos de dados contendo número considerável de registros de

operações governamentais. Esses registros podem ser ligados com técnicas e conhecimentos dos auditores de tribunais de contas brasileiros e assim elucidar perfis de fraudes, conluíus e outros tipos de crimes contra a administração pública. É possível também aumentar a responsabilidade das equipes de auditoria governamental, bem como, aumentar a economia, eficiência e eficácia.

Voos e Marques (2020) tem como objetivo revelar potências e fragilidades da comunicação do risco sanitário por meio do *Facebook*, a partir de páginas da Vigilância Sanitária nesta plataforma social. Neste país, nos últimos dez anos, tem-se vivenciado uma elevação progressiva da utilização de mídias sociais, possibilitando modos de vida que se engendram a partir das trocas entre os universos *online* e *off-line*, que se justapõem e se complementam. As autoras afirmam ser imprescindível que as comunicações se materializem por intermédio das mídias sociais, e se a saúde pública não se apropriar deste espaço outros o farão. Neste caso, é muito utilizada a técnica de mineração de dados e na sequência *clustering*.

Ao analisar as páginas da VS contatou-se sua proatividade, pois, a VS está operando por meios alternativos de informar e dialogar com a coletividade informando a respeito dos riscos sanitários, e assim promovendo a saúde. Verificou-se também que a esfera municipal da VS se utiliza desta plataforma social de forma mediata, sendo pouco usada na esfera estadual. Na área federal a VS preferiu utilizar o *Twitter* negligenciando a mídia social mais popular no Brasil, o *Facebook*. Observando o engajamento dos cidadãos ao acessarem a *internet*, que refletem as diferenças socioeconômicas e culturais. Levando em contas tais diferenças sociais é preciso ser considerada no planejamento das políticas públicas que se destinem à comunicação digital (Voos & Marques, 2020).

A ANVISA aproveitou o espaço comunicativo cedido pelo *Facebook* para implementar mecanismos de participação social, convocando os usuários para participar de consultas públicas conexas a doenças raras. Já a Vigilância Sanitária Rio assinala para a importância dada pelos cidadãos vinculados ao tema do controle de zoonoses, pois houve um crescimento acentuado de cães e gatos, ocasionado por ações de defesa dos animais e a importância dada pelas obras da VS, na defesa da sociedade em relação aos riscos da raiva e do esporo tricose. Percebe-se que ambas as VS se preocupam com a reputação de seus trabalhos, pois, precisam assegurar uma comunicação do risco que seja confiável, honesta e consistente. Ainda mais hoje com a propagação de notícias falsas (*fake news*), a era da midiatização do risco sanitário

suscitam o poder de articulação do Estado, e a confiabilidade dos estabelecimentos públicos é posta em xeque (Voos & Marques, 2020).

Na mesma área da saúde, Lins *et al.* (2019) se utilizam da multi metodologia que considera que o foco está em um problema complexo do mundo real. Portanto, é necessária uma combinação de métodos quantitativos e qualitativos, para enfrentar as várias naturezas do problema. Isso se torna apropriado para apoiar decisões de políticas, nos setores de educação, saúde, meio ambiente e segurança. Ao tratar de problemas complexos e reais, e se abordando a respeito de pessoas, pontos de vista e valores, não se podiam ser representados matematicamente. Desenvolvimento de uma gama de não quantitativos, métodos suaves, que eram, no entanto, rigorosos e sistêmicos. Os exemplos são metodologia de sistemas suaves (SSM), mapeamento cognitivo, análise de escolha estratégica e apresentação e teste de suposições estratégicas (SAST).

A integração de indicadores quantitativos com descrições qualitativas de um contexto é uma questão compreensível de diversas disciplinas científicas, pois, contribui para vincular abordagens teóricas e práticas à resolução de problemas, a exemplos de métodos de estruturação de problemas, o pensamento sistêmico e a multi metodologia. Podem fortalecer a compreensão mútua por analistas e *stakeholders* do problema em questão, se forem utilizados os resultados metodológicos e analíticos. Já os resultados quantitativos admitem que os “municípios ineficientes” compreendam as causas para obter sua eficiência geral em termos de baixas eficiências parciais de DEA específicas combinadas com altas taxas de mortalidade. O DEA que é fundamentado em variáveis quantitativas de mapas conceituais, é então utilizado para calcular os indicadores de desempenho e instituir referências municipais de saúde em agrupamentos selecionados (Lins *et al.*, 2019).

Visitas a Parques Verdes *versus* Saúde Pública foi uma investigação, descrita por Ullah *et al.*, (2020), do comportamento espaço-temporal de visitantes em mais de 115 parques verdes em 10 distritos de Xangai com aproximadamente 250.000 *check-ins*. Utilizou-se mineração de *Big Data* de mídia social, estudando-se o comportamento dos populares em relação às visitas a estes parques, usando o *Weibo* (maior blog de mídia social da China). O objetivo da pesquisa foi a conscientização a respeito da associação dos parques verdes com a saúde do público, auxiliar os ordenadores de políticas públicas sobre as vantagens de fornecer acesso a espaços verdes para residentes urbanos e prover às cidades urbanas novas abordagens sistemáticas para

quantificar e monitorar seu acesso a espaços verdes para aumentar as atividades físicas saudáveis entre os cidadãos.

Foi averiguado os padrões espaciais, temporais e afetivos de 3 anos de visitas a parques públicos usando dados do Weibo para pesquisar a saúde de frequentadores em Xangai e comparar os *check-ins* de cada ano com base em prazos semanais e mensais para ressaltar como o comportamento dos visitantes está mudando a cada ano. Os resultados indicam que o número de visitantes está ampliando ano a ano. Esta pesquisa baseada em LBSN conclui que frequentadores do sexo feminino têm maior probabilidade de visitar parques verdes do que frequentadores do sexo masculino (Ullah *et al.*, 2020).

Segundo Jayashree *et al.* (2018) a precisão da abordagem de classificação baseada em Mapa Cognitivo *Fuzzy* – FCM - é melhor do que os algoritmos de aprendizado de máquina *benchmark*, que apresentam deficiência em trabalhar com pequenos conjuntos de dados e sem serem capazes de usar o conhecimento de especialistas. O trabalho apresentado busca classificar o risco de surto de dengue em determinada região geográfica antecipadamente por meio de previsões, sendo importantíssima para determinação de políticas públicas em relação a saúde.

Lima *et al.* (2018), colocam que além dos especialistas da educação, toda sociedade deve contribuir para melhorar a qualidade do ensino, que só pode acontecer por meio de pesquisas que contribuam para a eficiência e eficácia de avaliações permanentes. A metodologia que se utilizou de técnicas de mineração de textos, permitiu o desenvolvimento de uma arquitetura que pode ser adaptada a qualquer exame de avaliação. Isso permite uma análise automática de testes e dados educacionais.

De uma forma geral, os artigos da área de tecnologia da informação e comunicação demonstraram a eficiência da aplicação dos métodos de mineração de dados em seus trabalhos. Os artigos demonstram que nesta área a mineração de dados foi utilizada para estruturação de dados, principalmente. Antolínez (2018) utilizou o método de mineração de textos para codificar o conhecimento; Brito *et al.* (2018) utilizaram análise de redes sociais, mineração de dados e redes *bayesianas* em seu trabalho. E Faria e Monteiro (2018) utilizaram algoritmos supervisionados para mineração de dados.

5. CONCLUSÃO

Por meio da análise multiplex, descrita por autores estudados, é possível concretizar o planejamento de estratégias preventivas e repressivas e resolver crimes decorrentes da corrupção. É possível prender muitos criminosos e no futuro talvez prever e prevenir crimes de corrupção. Nesta mesma linha, a análise de risco de conluio e corrupção usando classificadores *Naive Bayes*, que tratam de corrupção em compras públicas e agências governamentais podem ser resolvidos. Evidenciam-se as práticas que podem ser utilizadas para prevenir e detectar crimes de corrupção, por meio de capacitação de servidores usando as técnicas e ferramentas utilizadas pelos autores estudados. As formas de mineração mais utilizadas na atualidade na área de Segurança Pública, geralmente se concretizam com a Mineração de Dados em Estruturas Complexas.

Já na área da vigilância Sanitária (VS) ao assumir a conduta de mediatizar os riscos sanitários encontrados em uma inspeção, como faz a VS municipal do Rio de Janeiro no *Facebook*, e a ANVISA que se utilizou primeiramente do *Twitter* depois do *Facebook*, assumem o compromisso de problematizar questões de relevância em saúde pública e mediar o debate. Os órgãos fiscalizadores precisam se modernizar, capacitando seus servidores, com cursos que se destinem à comunicação digital, em especial mineração de dados, clusterização, e as várias técnicas, metodologias e ferramentas necessárias para trabalhar com *big data* e redes sociais.

A multimetodologia aplicada à avaliação da atenção à saúde em municípios brasileiros, por meio dos seus resultados metodológicos e analíticos podem intensificar a compreensão mútua de analistas e *stakeholders* a respeito dos problemas encontrados. Podem também empregar a avaliação de desempenho por meio de *benchmarks* podendo contribuir para a regulação da área pública de saúde, visando à melhoria de seus serviços.

Outra maneira de incentivar a melhoria na saúde é por meio de se exercitar fisicamente por meio de caminhadas em parques verdes, levar as crianças para brincarem e praticarem atividades que desenvolvam o prazer pelos esportes. Potencializar atividades saudáveis entre os residentes urbanos, é imprescindível na medida em que fornece às cidades urbanas novas abordagens sistemáticas para quantificar e monitorar cada vez mais o acesso a áreas verdes, beneficiando a saúde humana.

Utilizar a ferramenta Mapa Cognitivo *Fuzzy* (FCM) enfatizando sua capacidade para prever risco de surto de dengue (ou outro tipo endêmico) em uma dada região levando em conta condições meteorológicas combinadas ao território, os fatores demográficos e de estilo de vida

dos residentes é imprescindível como alerta para o desenvolvimento precoce de políticas públicas amenizando a gravidade dos casos.

Na análise da qualidade do ensino, a prática da metodologia, utilizada nos exames do ENADE, revelou-se viável e adequada para ser reutilizadas em outros cursos, notadamente para a análise de exames que alcançam vários anos e são realizados por um número significativo de alunos. Porém, para a análise de outros testes, que podem ter estruturas completamente diferentes, a metodologia pode ser reutilizada, mas nova adaptação seria necessária.

Na área de tecnologia da informação e comunicação é importante atentar para a utilização da mineração de dados, análise de redes sociais e diversos algoritmos de aprendizagem de máquina para estruturar e codificar o conhecimento disponível, facilitando assim a interpretação e utilização dos dados para os mais diversos fins.

Os Órgãos Estaduais e Municipais devem estar atentos aos trabalhos de Mineração de Dados provenientes dos campos político, social, econômico e cultural que é de uma complexidade extraordinária, porém, de vital importância para organizar as prioridades principalmente no sentido de capacitar servidores em todas as áreas citadas neste trabalho, e assim evidenciar em suas auditorias e monitorias conhecimentos para a conquista e efetivação de direitos sociais, bem como o direito à saúde, e educação de toda a comunidade.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

- Antolínez, E. A. V. (2018) Modelado basado en datos para la clasificación semiautomática de correspondencia electrónica: caso de estudio para la Administración Pública Colombiana. 2018. Trabajo de grado (Maestría em Getión de Información), Escuela Colombiana de Ingeniería Julio Garavito.
- Balaniuk, R., Bessiere, P., Mazer, E., Cobbe, P. (2013) Collusion and corruption risk analysis using naïve bayes classifiers. *Communications in Computer and Information Science*, 246.
- Battiston, F., Nicosia, V., Latora, V., San Miguel, M. (2017) Layered social influence promotes multiculturalism in the Axelrod model. *Scientific Reports*, 7 (1809).

- Brasil. [Constituição (1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República.
- Brito, S. R., Silva, A. S., Mata, E. C., Vijaykumar, N. L., Rocha, C. A. J., Monteiro, M. A., Costa, J. C. W. A., Francês, C. R. L (2018) An approach to evaluate large-scale ICT training interventions. *Information Systems Frontiers*, 20, p. 883–899.
- Camilo, C. O.; & Silva, J. C. (2009) Mineração de dados: conceitos, tarefas, métodos e ferramentas. Instituto de Informática Universidade Federal de Goiás.
- Faria M. M. & Monteiro A. M.(2018) Intrusion detection in computer networks based on KNN, K-Means++ and J48. *In: Arai k.; Kapoor S.; Bhatia R. (org.). Intelligent Systems and Applications. IntelliSys 2018. Advances in Intelligent Systems and Computing*, 868. Springer. https://doi.org/10.1007/978-3-030-01054-6_19.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996) From data mining to knowledge discovery in databases. *AI magazine*, 17 (3), 37-54.
- Fraga, A. A. (2017) Detecção de casos suspeitos de fraudes em licitações realizadas nos municípios da Paraíba: uma aplicação de técnicas de mineração de dados. 2017. Dissertação (Mestrado em Economia) – Programa de Pós-graduação em Economia do Setor Público, Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba, João Pessoa.
- Gandulfo, P. I., Nunes, V. T., Ralha, C. G., Coimbra, E. D. N. (2016) Public organization cost reduction using a process mining method: a case study on vacation control. *In: Simpósio brasileiro de sistemas de informação: sistemas de informação na era da computação em nuvem*, 12., 2016, Florianópolis, SC. Anais [...].
- Greenhalgh, T. (1997) Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ: British Medical Journal*, 315 (7109), 672.
- Han, J & Kamber, M. (2012) Data mining: concepts and techniques. *Elsevier*, 3.
- Jayashree, L. S., Lakshmi Devi, R., Papandrianos, N., Papageorgiou, E. I. (2018) Application of fuzzy cognitive map for geospatial dengue outbreak risk prediction of tropical regions of southern india. *Journal Intelligent Decision technologies*, 12 (2), 231-250.
- Lima, P. S. N., Ambrosio, A. P. L., Felix, I. M., Brancher, J. D., Ferreira, D. J. (2018) Content analysis of student assessment exams. *IEEE Frontiers in Education Conference (FIE)*, 1-9.
- Lins, M. P. E., Netto, S. O. A., Lobo, M. S. C. (2019) Multimethodology applied to the evaluation of Healthcare in Brazilian municipalities. *Health Care Manag Sci*, 22 (2), 197-214.

- Santa Catarina [Constituição (1989)] Constituição do Estado de Santa Catarina de 1989. Santa Catarina, Edição atualizada em fevereiro de 2015.
- Silva, J. C. F., Carvalho, T. F. M., Basso, M. F., Deguchi, M., Pereira, W. A., Sobrinho, R. R., Vidigal, P. M. P., Brustolini, O. J. B., Silva, F. F., Dal-Bianco, M., Fontes, R. L. F., Santos, A. A., Zerbini, F. M., Cerqueira, F. R., Fontes, E. P. B. (2017) Geminivirus data warehouse: a database enriched with machine learning approaches. *BMC Bioinformatics*, 18 (1). <https://doi.org/10.1186/s12859-017-1646-4>
- Toledo, A. S. O., Carpi, L. C., Atman, A. P. F. (2020) Diversity analysis exposes unexpected key roles in multiplex crime networks. *In: complex networks XI springer proceedings in complexity, Anais [...]. Springer.*
- Torres B. Y., Lópes, B. V. F., Torres, C. P., Jimenes, M. G. (2020) Artificial intelligence techniques to detect and prevent corruption in procurement: a systematic literature review. *Communications in Computer and Information Science*, 1194.
- Ullah, H., Wan, W., Haidery, S. A., Khan, N. U., Ebrahimpour, Z., Muzahid, A. A. M. (2020) Spatiotemporal patterns of visitors in urban green parks by mining social media big data based upon WHO reports. *IEEE Access*, 8, 39197-39211.
- Von Wangenheim, A., Jr. Barcellos, C. L., Andrade, R., Giuliano, I. C. B., Borgatto, A. F., Andrade, D. F. (2013) Implementing DICOM structured reporting in a large-scale telemedicine network. *Telemedicine and e-Health*, 19 (7), 535-541.
- Voos, F. L. & Marques, M. C. D. C. (2020) A vigilância sanitária no Facebook: potências e fragilidades da comunicação do risco sanitário na esfera digital. *Saúde e Sociedade*, 29 (1),1-14.
- Waszak, C. & Sines, M. C. (2003) Mixed methods in psychological research. *In: A. Tashakkori e C. Teddley. (Eds.), handbook of mixed methods in behavioral and social research. Anais [...]* 557-576. *Thousand Oaks: Sage Publications.*