

APLICAÇÃO DO PROCESSO DE DESCOBERTA DE CONHECIMENTO EM DADOS SOBRE MORTALIDADE INFANTIL DO ESTADO DE PERNAMBUCO

APPLICATION OF THE KNOWLEDGE DISCOVERY PROCESS IN DATA ON INFANT MORTALITY DATA IN THE STATE OF PERNAMBUCO - BRAZIL

Donathan da Silva Ferreira
Universidade Federal Rural de Pernambuco (UFRPE)
Bacharel em Administração Pública
donathansistema@gmail.com

Rafaella Leandra Souza do Nascimento
Universidade Federal de Pernambuco (UFPE)
Doutoranda em Ciência da Computação
rafaellalsn@gmail.com

Liliane Aparecida da Silva Santos
Universidade Federal de Pernambuco (UFPE)
Mestre em Gestão Pública para o Desenvolvimento do Nordeste
lilianeasantos@ymail.com

Kaline Maria Tenório Salviano
Universidade Federal de Pernambuco (UFPE)
Mestre em Administração
kalinesalviano@gmail.com

RESUMO

A Taxa de Mortalidade Infantil representa um dos grandes desafios humanitários a serem superados. No contexto nacional, o estado de Pernambuco tem logrado êxitos em diminuí-la, porém, a quantidade de óbitos infantis relacionados a causa evitáveis continua alta. Portanto, este estudo objetivou descobrir associações entre variáveis relacionadas com mortalidade infantil pernambucana de menores de um ano de idade e associar as descobertas com recomendações de políticas públicas. Utilizou-se a base de dados do Sistema de Informação sobre Mortalidade de 2018, fornecida abertamente pelo Ministério da Saúde. Na metodologia, utilizou-se o processo de descoberta de conhecimento em banco de dados com o suporte do software WEKA e do algoritmo Apriori. Ao analisar os resultados, pode-se sugerir melhorias no sistema público de saúde estadual, principalmente no que diz respeito a mitigar as causas evitáveis por atenção às gestantes.

Palavras-chave: *Dados Abertos, Mineração de Dados Públicos, Mortalidade Infantil.*

ABSTRACT

The Infant Mortality Rate represents one of the most significant humanitarian challenges to be overcome. In the national context, the state of Pernambuco has succeeded in reducing it; however, the number of infant deaths related to preventable causes remains high. Therefore, this study aimed to discover associations between variables related to infant mortality in Pernambuco under one year of age and associate the findings with public policy recommendations. The 2018 Mortality Information System database was used, openly provided by the Ministry of Health. In the methodology, the knowledge discovery process in a database was used with the WEKA software and the Apriori algorithm. By analyzing the results, improvements in the state's public health system can be suggested, especially concerning mitigating the avoidable causes for the care of pregnant women.

Keywords: *Open Data, Public Data Mining, Infant Mortality.*

1. INTRODUÇÃO

A Taxa de Mortalidade Infantil (TMI) é um indicador importante de desenvolvimento social. Com ela é possível caracterizar não só qualidade dos serviços de saúde prestados à população, mas também a influência dos fatores socioeconômicos sobre bem-estar da gestante e do recém-nascido. O cálculo da TMI é bem simples: a cada mil crianças nascidas vivas, procura-se saber quantas delas vieram a óbito antes de completar os primeiros doze meses de vida. Considera-se também o ano em que os fatos ocorreram bem como os limites geográficos da região do natalino (BRASIL, 2019a).

Nesse sentido, globalmente, em 2018, o número de crianças e adolescentes mortos principalmente por causas evitáveis foi de cerca de 6,2 milhões (UNICEF *et al.*, 2019). Desse total, aproximadamente 5,3 milhões eram de crianças menores de cinco anos de idade, sendo que quase metade delas eram recém-nascidas (UNICEF *et al.*, 2019). Portanto, esse quadro demonstra que a mortalidade infantil ainda é um dos grandes problemas sociais a serem superados pela humanidade.

O Brasil, por sua vez, teve em 2018 uma TMI igual a 12,2; 0,2 menor que a TMI do ano do anterior (FUNDAÇÃO ABRINQ, c2015), no entanto, em 2019 essa taxa voltou a crescer para 12,4. Mesmo assim, comparando o resultado alcançado em 2018 com as 16 mortes registradas para cada mil crianças nascidas vivas no ano 2010, percebe-se uma redução importante de 3,6 na TMI (FUNDAÇÃO ABRINQ, c2015). Logo, constata-se que houve algum progresso em termos da redução da TMI no cenário nacional.

Regionalmente, dentre os estados brasileiros, destaca-se que Pernambuco, também, teve redução na TMI. No ano de 2015, por exemplo, a taxa de mortalidade infantil pernambucana foi de 13,26, em 2016 esse número reduziu em 0,56 vindo a cair novamente no ano 2017 em mais 0,6 até

chegar em 11,7 no ano de 2018 (BRASIL, 2016a, 2017, 2018, 2019b). No contexto nacional, por exemplo, Pernambuco ficou apenas na nona posição entre os estados brasileiros com menores taxa de mortalidade infantil, ou seja, a cada mil crianças pernambucanas nascidas vivas, 11,7 faleceram antes mesmo de completarem o primeiro ano de vida, considerando-se os resultados divulgados para o ano de 2018 (BRASIL, 2019b).

Uma das causas que culminaram na redução da TMI em Pernambuco foi a instituição do Programa Mãe Coruja Pernambucana (PMCO), uma política pública criada pelo Decreto n. 30. 859, de 4 de out. de 2007 (PERNAMBUCO, 2007). Este programa não só se preocupa com a atenção integral à gestante, do parto ao puerpério, mas também com a promoção da alfabetização das mães com pouca escolaridade; além disso, ressalta-se que essa política pública é destinada às usuárias do Sistema Único de Saúde (SUS) residentes no Estado de Pernambuco e traz como benefício a proteção social e a segurança alimentar e nutricional da mãe e do bebê (PERNAMBUCO, 2009).

As contribuições desse programa sobre as taxas de mortalidade infantil e materna foram investigadas em um estudo de autoria de Cabral (2016), e comprovou-se que o PMCO além de aumentar o número de consultas pré-natais, teve efeito positivo na redução da mortalidade infantil; contudo, não houve muita influência quanto à redução da taxa de mortalidade materna, que é expressa pela razão entre o número de óbitos maternos e o número de nascidos vivos ocorridos na mesma área e período. Apesar da redução, é relevante afirmar que a TMI registrada em Pernambuco e demais estados, além da própria média do Brasil, está muito elevada se comparada a países como Japão, Noruega, Singapura e Suécia, em que duas crianças, a cada mil nascidas vivas, falecem antes de completar o primeiro ano de vida (UNICEF *et al.*, 2019).

A composição da TMI sob os três níveis de gestão – federal, estadual e municipal - provém dos dados coletados pelos sistemas de informação do Sistema Único de Saúde (SUS) (CUNHA; VARGENS, 2017). Esses dados sobre saúde pública, quando bem analisados e tratados, podem revelar associações entre os acontecimentos registrados na base de dados e, dessa forma, contribuir para prevenção e promoção da saúde materna-infantil. Por isso é bom entender que “cada vez mais os órgãos públicos vêm tomando decisões com base nos dados, seja para detecção de anomalias, monitoramento de indicadores ou melhoria de processos” (BRASIL, 2020a, p. 5).

Para Cunha e Vargens (2017, p. 72) “dados e informação são bases e fundamentos para a organização, a gestão e a efetivação das ações e serviços do setor de saúde”. Nesse contexto, a metodologia de descoberta de conhecimento em bases de dados de Fayyad *et al.* (1996) vem sendo

utilizada por muitos autores com o propósito de descobrir tendências ocultas em grandes volumes de dados. Portanto, a aplicação destas técnicas objetiva revelar algum conhecimento útil, não perceptível antes da utilização desta metodologia nos dados.

Pelo exposto, viu-se nos parágrafos anteriores que, apesar do relativo progresso alcançado, o estado de Pernambuco nunca conseguiu reduzir a TMI abaixo dos dois dígitos. Por isso, essa pesquisa tem por finalidade conhecer as relações entre as variáveis que podem contribuir para a manutenção desse quadro no estado. Para isso, fez-se o uso da metodologia da descoberta de conhecimento em bases de dados, a qual consiste em um processo que revela associações úteis nos dados, que podem promover melhorias no processo de tomada de decisão e na formulação de políticas públicas destinadas à promoção da saúde das gestantes e dos recém-nascidos.

Assim, procurou-se responder ao seguinte problema de pesquisa: de que forma a aplicação do processo de descoberta de conhecimento em dados abertos pode auxiliar no entendimento da TMI no Estado de Pernambuco? Desta forma, este estudo utilizou a base do Sistema de Informação sobre Mortalidade (SIM) de 2018, que é disponibilizada abertamente pelo Ministério da Saúde. Objetivou-se, então, analisar as variáveis e extrair padrões úteis nos dados do SIM de 2018, no âmbito do estado de Pernambuco, tendo-se a pretensão de contribuir com o entendimento do cenário estudado e com isso apresentar discussões e recomendações que almejam a prevenção e promoção da saúde do binômio mãe-filho.

2. REFERENCIAL TEÓRICO

2.1 DADOS ABERTOS

No Brasil, a Lei n. 12.527/2011 regula o acesso às informações mantidas por União, Estados, Distrito Federal e Municípios (BRASIL, 2011). Assim, como forma de promover a transparência na Administração Pública Federal e fomentar o desenvolvimento do controle social, o governo federal criou o Portal Brasileiro de Dados Abertos (BRASIL, 2020b). Trata-se de uma ferramenta tecnológica de livre acesso, alimentada pelo governo com dados e informações públicas.

Esses dados são organizados de maneira a permitir sua reutilização em aplicativos desenvolvidos pela sociedade. Além disso, eles apresentam diversas utilidades relacionadas com a administração pública, em especial, no tocante à saúde, à segurança e aos gastos governamentais. Por isso, os dados abertos possibilitam a participação da sociedade no planejamento e no

acompanhamento de políticas públicas (BRASIL, 2020b). Ademais, eles podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa. Por conseguinte, não são restringíveis em termos de acesso, no máximo, pode ser exigida referência da fonte e do compartilhamento (OPEN KNOWLEDGE BRASIL, 2020).

Segundo Junior *et al.* (2014) os dados abertos estão disponíveis na web e são legíveis por máquinas computacionais. Ademais, no contexto público, tais dados podem subsidiar a construção de indicadores acerca de políticas públicas, projetos e programas governamentais. Os aludidos autores, por exemplo, analisaram os dados abertos do Exame Nacional do Ensino Médio (ENEM) aplicando a mineração de dados; desse modo, concluíram que fatores como escolaridade dos pais, lugar de residência do discente e o acesso à internet influenciam na nota final do inscrito.

Para Correa *et al.* (2018), dados abertos também podem ser usados como suporte à saúde pública. Segundo esses pesquisadores, por meio da divulgação de informação de interesse coletivo, tal como o preço de remédios em plataformas webs acessíveis e gratuitas, a tendência é que a população fiscalize ainda mais eventuais aumentos indevidos nos preços desses produtos. Portanto, os dados abertos são uma das maneiras de estimular a participação da sociedade nas questões de interesse coletivo.

Finalmente, é inegável que a instituição da Política de Dados Abertos do Poder Executivo Federal por meio de Decreto nº 8.777, de 11 de maio de 2016 (BRASIL, 2016b) trouxe ganhos em termos de transparência pública à população. No entanto, ainda existem dificuldades, principalmente, relacionadas a pouca padronização das plataformas de bases de dados. Por isso, recomenda-se mais investimentos para o desenvolvimento de recursos tecnológicos e aplicações mais interativas e acessíveis a fim de facilitar e estimular a participação da sociedade nos negócios de interesse coletivo (BRASIL, 2016b; SILVA; GALVÃO, 2018).

Em últimas palavras, os dados abertos são fontes importantes de consulta para a sociedade. Com eles pode-se conhecer, por exemplo, informações sobre número de óbitos ou de nascimentos de crianças menores de um ano de idade, a quantidade de profissionais de saúde atuando na atenção básica hospitalar e a identificação de doenças por espaço geográfico (SOUZA *et al.*, 2019; SARTORELLI, *et al.*, 2017).

2.2 Mineração de Dados Públicos

Em primeiro lugar, precisa-se entender que a Descoberta de Conhecimentos em Banco de Dados é um processo não trivial, que revela informações úteis dificilmente identificáveis de um outro modo (FAYYAD *et al.*,1996). Já a Mineração de Dados (MD) é uma das etapas da Descoberta de Conhecimentos em Banco de Dados (FAYYAD *et al.*, 1996). Desse modo, a MD é entendida como um processo em que há aplicação de algoritmos inteligentes com a finalidade de extrair padrões interessantes de dados, revelando um novo conhecimento (FAYYAD *et al.*,1996).

Nesse contexto, muitos autores vêm defendendo que a MD pode auxiliar no processo de tomada de decisão em várias áreas do setor público. Assim, em um estudo sobre fraudes em licitações públicas na modalidade pregão eletrônico, Souto, Arruda e Araújo (2019), por exemplo, constataram que a MD pode identificar a presença de *bots* agindo em sites de pregões eletrônicos e concluíram que MD pode contribuir na identificação de ações suspeitas nas plataformas online. No trabalho de Nascimento *et al.* (2018) a MD foi utilizada para classificar e identificar perfis de empresas com potencial de irregularidade em relação a legislação dos impostos estaduais.

Já no contexto educacional, pesquisas feitas por Balaniuk (2010); Silva, Morino e Sato (2014) e Gonçalves e Beltrame (2019), mostraram ser possível a aplicação da MD em apoio a controle externo, bem como para buscar associações entre o desempenho de estudantes no Exame Nacional do Ensino Médio e os fatores socioeconômicos e ainda para analisar o cenário da evasão escolar estudantil em cursos de nível superior. Ainda, Silva *et al.* (2019) estudaram o desempenho de alunos utilizando informações dos anos de 2015 e 2017 do Sistema Nacional de Avaliação da Educação Básica (SAEB). Estes dados são disponibilizados abertamente no portal do Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP).

Analisando-se os dados abertos obtidos do portal #DataPoa, Santos *et al.* (2018) empregaram a MD a fim de conhecerem a quantidade de acidentes de trânsito envolvendo ciclistas na cidade de Porto Alegre-RS. Com isso chegaram à conclusão de que a maioria dos acidentes, aproximadamente 1260, aconteceram durante o dia. Eles também destacaram que grande parte de acidentes ocorreram nas regiões Sul (463) e Norte (365) da cidade e a quantidade de vítimas fatais foi considerada pequena pelos autores (37) no período de 2011 a 2015.

Galvão e Marin (2009) consideram que a MD pode ser aplicada na área da saúde pública como forma de agilizar a obtenção de conhecimentos. Correlaciona-se com isso o fato de que é possível analisar os bancos de dados hospitalares na pretensão de prever tendências, o que ajuda na identificação do melhor tratamento ao paciente. Esse fato foi comprovado em uma pesquisa feita no

estado do Paraná. Nessa investigação os autores defendem que a MD aplicada em bancos de dados de saúde pública é um processo importante na construção de mapas de risco de mortalidade infantil (VIANNA *et al*, 2010).

Como se vê, a MD pode ser usada em várias áreas da Administração Pública, tais como na saúde, na educação e até na modalidade de licitação pública denominada de pregão eletrônico. Além disso, segundo Silva, Peres e Boscaroli (2016), a MD se apropria de conceitos provenientes da Inteligência Artificial, Aprendizado de Máquina, Estatística e Banco de Dados. Assim, a MD se vale de softwares e outras ferramentas tecnológicas para analisar grandes conjuntos de dados na busca de padrões e relações que demonstrem algum conhecimento relevante, logicamente, isso seria incompatível com capacidade de processamento e assimilação de informação pelos seres humanos (SOUTO; ARRUDA; ARAÚJO, 2019).

2.3 MORTALIDADE INFANTIL

A Taxa de Mortalidade Infantil (TMI) representa um importante informativo sobre o grau de desenvolvimento social e econômico de um país ou região (KROPIWIEC; FRANCO; AMARAL, 2017). A TMI está associada às condições de vida da população e à qualidade da assistência de saúde direcionada ao binômio materno-infantil, da gestação ao parto, e durante o primeiro ano de vida do recém-nascido (GAVA; CARDOSO; BASTA, 2017).

Tecnicamente, TMI refere-se à quantidade de óbitos de menores de um ano de idade, acontecidos em determinado espaço geográfico, no ano considerado, por mil nascidos vivos (BRASIL, 2019a). Vale ressaltar que TMI é dividida nos seguintes componentes: neonatal precoce (0-6 dias de vida), neonatal tardio (7-27 dias) e pós-neonatal (28 a 364 dias) (BRASIL, 2019a). Com a TMI, portanto, é possível fazer uma estimativa da probabilidade de os nascidos vivos falecerem antes de completarem o primeiro ano de vida (BRASIL, 2019a).

Nesse contexto, o Sistema de Informação sobre Mortalidade (SIM) começou a entrar em operação no Brasil, em 1975 (BRASIL, 2001a). Basicamente, o SIM é um mecanismo de coleta e disponibilização de dados sobre as mortalidades registradas em todo território nacional (BRASIL, 2001a), sendo importante para mensurar a qualidade dos serviços de saúde ofertados à população.

Por outro lado, o Sistema de Informação sobre Nascidos Vivos (SINASC) só foi implantado no Brasil 15 anos depois do início do SIM (BRASIL, 2001b). O objetivo principal do SINASC é coletar

dados sobre os nascimentos ocorridos em todo território nacional (BRASIL, 2001b). Com esse conhecimento, é possível apresentar a taxa de natalidade nos diferentes estados da federação, bem como estimar o crescimento demográfico do país para os próximos anos.

Desse modo, é interessante destacar que as análises em bancos de dados do SIM e do SINASC podem apresentar situações que antes eram desconhecidas. Marinho *et al.* (2019), por exemplo, investigaram os dados provenientes do Censo Demográfico 2010 e do SIM e descobriram que, proporcionalmente, nas microrregiões do Brasil a mortalidade infantil de indígenas é maior do que a dos não indígenas. Nesse trabalho, eles identificaram também que a maioria das mortes infantis de indígenas ocorreram no período pós-neonatal; enquanto na população de não indígenas, os falecimentos dos menores de um ano de idade, vieram a acontecer, em grande parte, no período neonatal precoce.

Em um outro estudo, Sartorelli *et al.* (2017) usaram a MD nos dados do SINASC e do SIM de uma cidade paranaense e constataram que a inadequada assistência à mulher durante a gestação ainda é uma das principais causas de mortalidade infantil evitável. Para França e Lansky (2016, p. 9) “não há dúvida de que os maiores riscos de mortalidade neonatal são geralmente associados à qualidade da atenção pré-natal, ao atendimento ao parto e aos cuidados recebidos pelo recém-nascido”. Logo, a diminuição da mortalidade infantil na variável neonatal está relacionada com a superação desses problemas.

Já Maia, Souza e Mendes (2020) identificaram que grande parte dos óbitos infantis estão associados a fatores biológicos tais como, o baixo peso ao nascer, a prematuridade, a asfixia grave ou a moderada, a gestação múltipla e a presença de malformações congênitas nos bebês. Ademais, os referidos autores usaram as informações do SIM e do SINASC, de residentes de capitais brasileiras e detectaram que o número insuficiente e a baixa qualidade das consultas pré-natais contribuem significativamente para a manutenção desse quadro.

Outros destaques importantes apontados na investigação de Maia, Souza e Mendes (2020) dizem respeito às iniquidades sociais e econômicas. Situações como a pouca escolaridade materna, a não ocupação remunerada da mãe, os extremos etários, ou seja, ser adolescente ou mulher com 35 anos ou mais e até a raça da gestante são fatores que dificultam o acesso materno-infantil aos serviços de saúde, principalmente, os mais especializados (MAIA, SOUZA, MENDES, 2020).

Neste contexto, Silva, Paes e Silva (2018) analisaram não só a fecundidade, mas também os efeitos de dois programas governamentais sobre a mortalidade infantil no semiárido brasileiro, no

período de 2005 a 2010. Trata-se do Programa Bolsa Família (PBF) e a Estratégia Saúde da Família (ESF). Segundo os referidos autores, a transferência de renda proporcionada pelo PBF depende de alguns requisitos, tais como: cartão de vacina do recém-nascido em dia e que a grávida faça exames pré-natais e tenha acompanhamento nutricional durante todo o período de gestação. O ESF, também, traz ganhos na promoção da saúde do binômio materno-infantil, já que facilita o acesso de famílias de baixa renda aos serviços de saúde. Dessa forma, os supracitados autores concluíram que a aplicação desses dois programas contribui de maneira significativa para diminuição dos óbitos infantis no semiárido brasileiro.

Em últimas palavras, outra situação que também contribuiu para a redução da taxa de mortalidade infantil no semiárido brasileiro foi a redução da fecundidade. Como bem observado por Silva, Paes e Silva (2018), no período de 2005 a 2010, a média da Taxa de Fecundidade Total (TFT) da região declinou. Segundo esses autores, a menor quantidade de filhos pode fazer com que a qualidade da assistência aos recém-nascidos melhore, impactando de forma positiva na qualidade de saúde materno-infantil.

3. PROCEDIMENTOS METODOLÓGICOS

Inicialmente, recorreu-se aos estudos já realizados com o propósito de averiguar o estado da arte das contribuições científicas acessíveis pelas principais plataformas de consulta de estudos acadêmicos. Por isso, quanto aos objetivos, este estudo fez uso da pesquisa exploratória. Nesse sentido, objetivou-se investigar o significado de conceitos desconhecidos bem como orientar o delineamento do estudo (SAMPIERI; CALLADO; LUCIO, 2013).

Este estudo utilizou-se da base de dados sobre a mortalidade geral de 2018, arquivo disponibilizado pelo Ministério da Saúde no Portal Brasileiro de Dados Abertos. Assim, além de nomear e classificar as variáveis presentes em tal banco de dados, o momento seguinte passou a ser a interpretação delas de acordo com o contexto do estado de Pernambuco. Portanto, esta pesquisa também é classificada como descritiva, em que a intenção principal não é explicar, mas sim demonstrar os fatos.

No que tange à abordagem, essa pesquisa é classificada como quantitativa, pois objetivou a quantificação e classificação dos principais resultados. “A pesquisa quantitativa, que tem suas raízes no pensamento positivista lógico, tende a enfatizar o raciocínio dedutivo, as regras da lógica e os

atributos mensuráveis da experiência humana” (GERHARDT; SILVEIRA, 2009. p.33). Dado o exposto, procurou-se focar na análise objetiva dos dados sobre a mortalidade infantil no Estado de Pernambuco para o ano de 2018.

Pelo exposto, quanto à finalidade, esta pesquisa é classificada como aplicada, tipo que, para Kauark, Manhães e Medeiros (2010), envolve a produção de conhecimentos para solucionar problemas específicos. Desse modo, essa pesquisa tem a pretensão de encontrar conhecimentos úteis num banco de dados abertos, visando a utilização dos mesmos em um contexto determinado.

3.1. A COLETA E A DESCRIÇÃO DOS DADOS

Em primeiro lugar, efetuou-se a pesquisa bibliográfica com a finalidade de conhecer as principais abordagens acerca do assunto tratado neste trabalho. Desse modo, este estudo perseguiu recomendação de Lamy (2020), no que se refere à coleta bibliográfica, ou seja, procurou-se por trabalhos acadêmicos em fontes confiáveis, definindo-se os termos de pesquisas de acordo com os principais assuntos referentes à problemática da investigação.

Assim sendo, a seleção dos trabalhos foi realizada nos Portais Google Acadêmico e Periódicos CAPES. Nesse sentido, constatou-se a importância de definição de termos com a finalidade de tornar o processo de busca mais eficiente. Logo, palavras chaves como: mortalidade infantil, dados abertos, mineração de dados, mortalidade infantil no Estado de Pernambuco, mortalidade infantil e mineração de dados; em associação com os termos: administração pública ou gestão pública; foram as frases digitadas nos supracitados mecanismos de busca online. Assim sendo, a filtração pelos trabalhos científicos mais recentes e a escolha daqueles que têm mais similitude com a temática desta investigação foram aspectos tratados na última etapa da coleta bibliográfica, a seleção do universo amostral. Depois de concluídos esses estágios de coleta bibliográfica, procedeu-se a interpretação e contextualização das informações encontradas.

A coleta documental de dados (dados secundários) aconteceu num segundo momento, no qual consultou-se os dados abertos sobre a mortalidade geral referente ao ano de 2018. Cabe destacar que esses dados abertos tiveram como origem o SIM de 2018. Nesse sentido, pontua-se que é o Departamento de Informática do Sistema Único de Saúde (DATASUS), órgão pertencente à Secretaria Executiva do Ministério da Saúde, o gerenciador do SIM (FERRAZ, 2009). Reitera-se que o SIM é um mecanismo de coleta de dados sobre mortalidade geral de abrangência nacional. O sistema pode ser utilizado como um indicador importante para mensurar a qualidade de vida da

população. Além disso, é possível utilizá-lo com a finalidade de avaliar a eficácia e eficiência das políticas públicas destinadas ao melhoramento do sistema público de saúde.

Nesse contexto, importa mencionar que o banco de dados sobre mortalidade geral do ano de 2018 está acessível – de maneira aberta – no Portal Brasileiro de Dados Abertos. O Ministério da Saúde (MS) é o mantenedor e o divulgador dessas informações à coletividade. Logo, essa atitude visa facilitar a transparência das informações públicas sob guarda dos órgãos governamentais.

Destaca-se que essa base de dados contém 87 variáveis ou colunas e 1.316.719 observações ou linhas que abrange o território nacional e informações de mortalidade geral, disponibilizadas em formato *.csv*. Entretanto, como este estudo foca na análise dos dados sobre mortalidade infantil no Estado de Pernambuco, delimitou-se a quantidade de colunas e linhas ao respectivo espaço geográfico, ou seja, os registros referentes a outros estados e as faixas etárias maiores que 1 ano de idade foram suprimidos da observação final. Nesse contexto, foi identificado um total de 1.690 registros considerados para análise referentes a Pernambuco.

Além disso, fatores evidenciados em outros estudos como relevantes para a mortalidade infantil, também, foram considerados nesta pesquisa de maneira a afunilar ainda mais o processo de descoberta do conhecimento. Tem-se com isso, mais uma etapa de delimitação da pesquisa. Isso é importante porque nem todas as variáveis contidas na base são relevantes para esta pesquisa. Assim, características econômicas, sociais e biológicas da mãe, especialmente, aquelas apontadas como influenciadoras da mortalidade infantil e já reportadas em outros estudos foram analisadas nesta investigação. Portanto, a escolaridade da mãe, o tipo de trabalho exercido pela mãe e a realização de exames pré-natais são algumas das variáveis que foram apuradas neste estudo.

3.2. A ANÁLISE DOS DADOS: O PROCESSO DE DESCOBERTA EM BANCO DE DADOS

Nessa parte do estudo, realizou-se a análise das variáveis selecionadas para o processo de MD. Na prática, teve início a descoberta de conhecimento em bases de dados, uma metodologia interativa envolvendo muitas etapas (FAYYAD *et al.*, 1996; HAN; KAMBER; PEI, 2012), como demonstrado na Figura 1.

Figura 1 - Mineração de dados como parte integrante do processo de descoberta do conhecimento.



Fonte: Autoria própria.

Observa-se que o processo de descoberta de conhecimento é composto das seguintes fases:

- Limpeza de dados: com a finalidade de suprimir conteúdos desnecessários;
- Integração de dados: nessa fase os dados de fontes diferentes podem ser combinados;
- Seleção de dados: onde os dados considerados relevantes são recuperados;
- Transformação dos dados: onde os dados são consolidados ou transformados de forma a se permitir a mineração;
- Mineração de dados, basicamente, consiste na aplicação de métodos inteligentes com o propósito de extrair padrões de dados;
- Avaliação do padrão: é nesse ponto que se tenta identificar os padrões relevantes para a obtenção do conhecimento;
- Apresentação do conhecimento: nesse estágio os mecanismos de visualização e exibição dos conhecimentos são empregados a fim de apresentar o conhecimento extraído ao usuário.

Ainda conforme Han, Kamber e Pei (2012), as quatro primeiras fases do processo de descoberta de conhecimento em banco de dados são distintas formas de pré-processamento de dados. Assim, objetiva-se nessas quatro etapas iniciais a preparação dos dados para a MD. Vale destacar que a proposta da MD tem utilidade para descobrir padrões interessantes de dados, originados de banco de dados, armazém de dados e a Web (HAN; KAMBER; PEI, 2012). Portanto, busca-se com isso encontrar conhecimentos importantes sobre determinado assunto.

Aplicando-se essas distintas maneiras de pré-processamento de dados ao objeto de investigação deste estudo, realizou-se uma pré-seleção de dados. Desse modo, das 87 colunas da base de dados do

SIM de 2018, rejeitou-se 64. O primeiro motivo para tais descartes foi a constatação de que havia uma grande quantidade de dados faltantes em muitas das colunas da base de dados, dificultando a utilização de alguma técnica para identificação de padrão.

Outro critério de exclusão adotado nesta pesquisa é a relação da variável em análise com o tema central desta investigação. Ora, não faria sentido algum manter variáveis como: OCUP- tipo de trabalho que o indivíduo mais desenvolveu antes de vir a óbito; ESC- escolaridade em anos do falecido e o ACIDTRAB - motivo da morte do indivíduo é consequência de acidente no trabalho. Vale ressaltar que este trabalho aborda a mortalidade infantil pernambucana, portanto, presume-se que essas variáveis fogem em muito do escopo deste estudo.

Com base na variável CAUSABAS duas novas variáveis foram incluídas na base de dados com o objetivo de agregar conhecimento ao estudo. Segundo a definição de causas de óbitos infantis (MALTA *et al.*, 2007) identificou-se tais causas são evitáveis ou não (TIPOCAUSA) e as classes de evitabilidade (CLCAUSA). Para algumas variáveis criou-se classes com base em intervalo de valores, para facilitar a identificação de grupos nos dados como em IDADEMAE e PESO. Um último método de pré-processamento foi a aplicação da técnica da conversão. Segundo Provost e Fawcett (2016), ela tem o propósito de preparar os dados, desconsiderando-se os dados duplicados e os valores ausentes. Para tratar as variáveis que ainda continham valores ausentes, abordou-se a estratégia de criar uma categoria dentro das variáveis identificando os campos ausentes (NP – não preenchido).

Assim, no final, a nova base de dados ficou com 25 variáveis e 1.690 registros. O Quadro 1 expõe as 25 variáveis resultantes da fase de pré-processamento dos dados:

Quadro 1. Variáveis selecionadas para estudo.

N.º	Nome	Definição
1	SEXO	Sexo do falecido. (M – masculino; F – feminino; I - ignorado).
2	RACACOR	Cor informada pelo responsável pelas informações do falecido. (1 – Branca; 2 – Preta; 3 – Amarela; 4 – Parda; 5 – Indígena).
3	LOCOCOR	Local de ocorrência do óbito. (1 – hospital; 2 – outros estabelecimentos de saúde; 3 – domicílio; 4 – via pública; 5 – outros; 6 - aldeia indígena; 9 – Ignorado).
4	IDADEMAE	Idade da mãe. (Números).
5	ESMAE	Escolaridade da mãe em anos. (1 – Nenhuma; 2 – de 1 a 3 anos; 3 – de 4 a 7 anos; 4 – de 8 a 11 anos; 5 – 12 anos e mais; 9 – Ignorado).
6	ESMAE2010	Escolaridade 2010. Nível da última série concluída pela mãe. (0 – Sem escolaridade; 1 – Fundamental I (1ª a 4ª série); 2 – Fundamental II (5ª a 8ª

		série); 3 – Médio (antigo 2º Grau); 4 – Superior incompleto; 5 – Superior completo; 9 – Ignorado).
7	OCUPMAE	Tipo de trabalho habitualmente exercido pela mãe, de acordo com a Classificação Brasileira de Ocupações – CBO 2002. No caso da mãe do falecido(a) ser “aposentada”, preencher com a ocupação habitual anterior.
8	QTDFILVIVO	Números de filhos vivos. (Número; 9 - ignorado).
9	QTDFILMORT	Número de filhos mortos. Não incluir a criança cujo óbito se notifica na respectiva DO. (Número; 9 - ignorado).
10	GRAVIDEZ	Tipo de gravidez. (1 – única; 2 – dupla; 3 – tripla e mais; 9 – ignorada).
11	SEMAGESTAC	Semanas de gestação com dois algarismos. (Números com dois algarismos; 9 - ignorado).
12	GESTACAO	Faixas de semanas de gestação (1 - Menos de 22 semanas; 2 - 22 a 27 semanas; 3 - 28 a 31 semanas; 4 - 32 a 36 semanas; 5 - 37 a 41 semanas; 6 - 42 e + semanas).
13	PARTO	Tipo de parto. (1 – vaginal; 2 – cesáreo; 9 – ignorado).
14	OBITO PARTO	Momento do óbito em relação ao parto. (1 - antes; 2– durante; 3–depois; 9– Ignorado).
15	ASSISTMED	Se refere ao atendimento médico continuado que o paciente recebeu, ou não, durante a enfermidade que ocasionou o óbito. (1 – sim; 2 – não; 9 – ignorado).
16	CAUSABAS	Causa básica da Declaração do Óbito -DO, de acordo com a Classificação Internacional de Doenças e Problemas Relacionados à Saúde - Códigos CID 10.
17	ATESTANTE	Indica se o médico que assina atendeu o paciente (1:Sim; 2: Substituto; 3:IML;4:SVO;5:Outros).
18	ESCMAEAGR1	Escolaridade da mãe agregada (formulário a partir de 2010). (00 – Sem Escolaridade; 01 – Fundamental I Incompleto; 02 – Fundamental I Completo; 03 – Fundamental II Incompleto; 04 – Fundamental II Completo; 05 – Ensino Médio Incompleto; 06 – Ensino Médio Completo; 07 – Superior Incompleto; 08 – Superior Completo; 09 – Ignorado; 10 – Fundamental I Incompleto ou Inespecífico; 11 – Fundamental II Incompleto ou Inespecífico; 12 – Ensino Médio Incompleto ou Inespecífico).
19	MORTEPARTO	Momento do óbito em relação ao parto. (1 - antes; 2– durante; 3–depois; 9– Ignorado).
20	PESO	Peso ao nascer em gramas. (Número (quatro algarismos))
21	NECROPSIA	Refere-se a execução ou não de necropsia para confirmação do diagnóstico. (1 – sim; 2 – não; 9 – ignorado).
22	TIPOCAUSA	Indica se o óbito é evitável ou não
23	CLCAUSA	Classes de evitabilidade
24	TPPOS	Óbito investigado. (1 – sim; 2 – não).
25	ALTCAUSA	Indica se houve correção ou alteração da causa do óbito após investigação. (1-Sim; 2 – Não).

Fonte: (BRASIL, 2020c).

Como se vê, as variáveis escritas no Quadro 1 referem-se às correlações de fatores que estão, intrinsecamente, relacionados à mortalidade infantil. Sendo assim, ao aplicar a MD nessas variáveis busca-se encontrar padrões de dados que revelem conhecimentos úteis, e assim, descrevê-los.

Desse modo, na fase de MD empregou-se a regra da associação com o propósito de identificar as ocorrências de padrões frequentes na base de dados (COSTA; BERNARDINI; VITERBO FILHO, 2014). As regras de associação são do tipo, se X então Y, sendo X e Y os conjuntos de itens. Desse modo, o que se pretende é aferir a probabilidade de os eventos X e Y acontecerem simultaneamente (CASTRO; FERRARI, 2016). Esse é o segredo da regra da associação: calcular as chances de duas ou mais variáveis manifestarem-se para uma mesma característica ou observação ao mesmo tempo; descobrindo-se, dessa maneira, uma informação útil.

Alerta-se que para determinar a qualidade da regra de associação, avaliando assim o padrão da mesma, usou-se das métricas de desempenho e o grau de confiança, que mensura a ocorrência de X implicar em Y (NASCIMENTO, 2016). Segundo Brasil Neto (2009) as medidas de suporte e de confiança podem tanto reduzir o espaço de busca, como aumentar o número de regras de associação. Desse modo, trabalhou-se neste estudo com o algoritmo mais usado na análise de regras de associação, o Apriori (SHARDA; DELEN; TURBAN, 2019). Com o algoritmo apriori é possível calcular as chances de dois ou mais eventos - não repetitivos - acontecerem ao mesmo tempo, por meio de lógica da associação (VASCONCELOS; CARVALHO, 2018).

Em vista disso, a MD foi feita com auxílio do WEKA, um software que usa algoritmos pré-configurados com a finalidade de facilitar o processo de compreensão dos dados. Essa aplicação é compatível com os seguintes métodos de MD: regressão, classificação, agrupamento, associação e seleção de atributos (WITTEN *et al.*, 2016). Portanto, nesse estudo usou-se o WEKA com propósito de encontrar associações entre as 25 variáveis selecionadas, visando descrição de conhecimentos interessantes acerca da mortalidade infantil pernambucana.

Desse modo, adotou-se a seguinte estratégia a respeito do algoritmo Apriori: em primeiro lugar, fez-se a geração de itens frequentes, posteriormente, a poda dos conjuntos não relevantes para descoberta de itens frequentes; finalmente, calculou-se o suporte (confiança) dos candidatos a regra (CASTRO, 2014).

4. RESULTADOS E DISCUSSÃO

Inicialmente, realizou-se análise das variáveis. Assim sendo, verificando as cinco maiores causas básicas relacionadas à mortalidade infantil encontradas nos dados, elaborou-se o Quadro 2. Em primeiro lugar, percebeu-se um maior número de ocorrências de Septicemia bacteriana não especificada do recém-nascido com 106 registros ; em segundo lugar, o feto e recém-nascidos

afetados por transtornos maternos hipertensivos com 101 ocorrências; já a malformação não especificada do coração ficou em terceiro lugar com 61 ocorrências; a causa referente ao feto e recém-nascido afetados por doenças maternas renais e das vias urinárias apareceu em quarto lugar com 46 ocorrências; finalmente, em quinto lugar, ficou o desconforto respiratório não especificado do recém-nascido com 44 ocorrências.

Quadro 2. As cinco maiores causas básicas de mortalidade infantil encontrada nos dados.

N.º	Código	Ocorrências	Definição
1	P369	106	Septicemia bacteriana não especificada do recém-nascido
2	P000	101	Feto e recém-nascido afetados por transtornos maternos hipertensivos
3	Q249	61	Malformação não especificada do coração
4	P001	46	Feto e recém-nascido afetados por doenças maternas renais e das vias urinárias
5	P229	44	Desconforto respiratório não especificado do recém-nascido

Fonte: Autoria Própria (2021), baseada na Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde CID-10 (MEDICINANET, 2021).

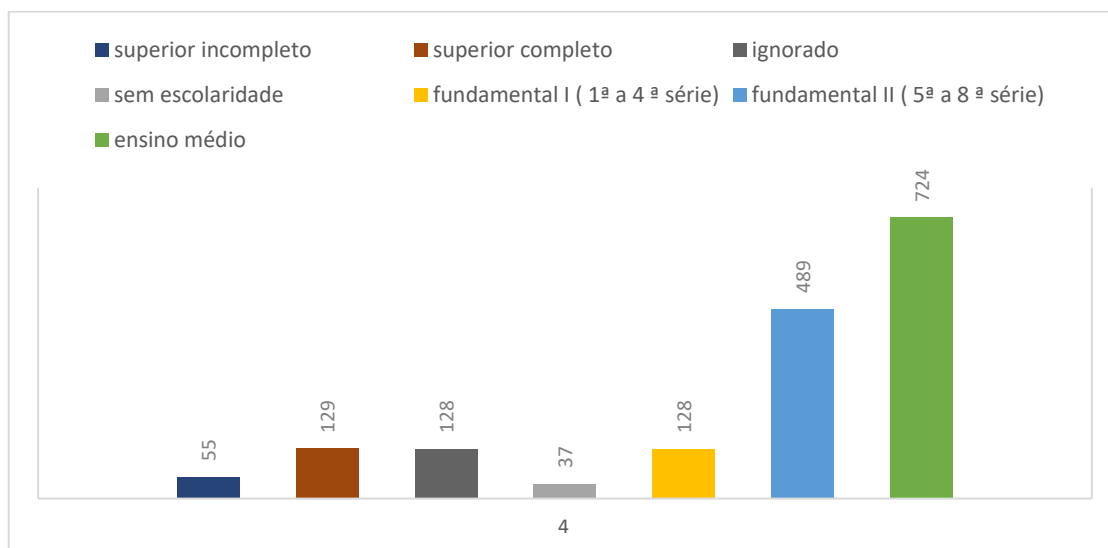
Como se vê, o Quadro 2 mostra que a septicemia bacteriana não especificada do recém-nascido foi a principal causa de mortalidade infantil em território pernambucano no ano de 2018. Trata-se de uma infecção sanguínea provocada pela exposição dos recém-nascidos a determinados tipos de bactérias. Ela acomete principalmente bebês com baixo peso ao nascer, o que pode estar diretamente relacionado à ausência de acompanhamento nutricional da mãe durante o período de gestação; ademais, outros motivos são: o vazamento do líquido que envolve o feto antes do parto bem como a existência de problemas socioeconômicos (TESINI, 2020).

Vale esclarecer que o alcoolismo e tabagismo materno, bem como a relação conjugal e a cobertura pré-natal também são fatores que influenciam sobre a saúde da grávida, por exemplo, o tabagismo materno influencia na prematuridade e isso pode aumentar as chances de infecção sanguínea no recém-nascido (ALVES, 2018). Portanto, o fato preocupante é que as 106 ocorrências de mortalidade infantil por septicemia bacteriana não especificada do recém-nascido são considerados uma quantidade alta de mortes infantis, já que se trata de uma infecção tratável.

Em relação ao grau de instrução, de um total de 1.690 registros de mulheres grávidas, que tiveram filhos não sobreviventes nos primeiros 12 meses de vida no ano 2018, 724 delas já tinham concluído o ensino médio, 489 o ensino fundamental II e 129 o ensino superior completo. Como se

vê, o número de mães com ensino médio é mais de cinco vezes maior que o número de mães com ensino superior. Esses valores podem ser representados no gráfico da Figura 2.

Figura 2: Nível da última série concluída pela mãe.



Fonte: Autoria Própria (2021).

Já no que diz respeito à variável ocupação exercida habitualmente pela mãe, o Quadro 3 mostra as maiores ocorrências na base. Assim sendo, constata-se que 681 dessas mães eram donas de casa, 151 estudantes e 182 trabalhadoras agropecuárias em geral. É bom entender que as grávidas, no desempenho de atividades agropecuárias, ficam expostas a riscos químicos, físicos e biológicos (MARCHIORI; FERRAZ, 2016). A manipulação de agrotóxicos pelas grávidas, por exemplo, pode ocasionar o aborto (BRASIL, 2019c).

Quadro 3. Ocupações da mãe com mais ocorrências na base de dados.

Código	Definição	Quantidade
999992	Dona de casa	681
621005	Trabalhador agropecuário em geral	182
999991	Estudante	151
622020	Trabalhador volante da agricultura	108

Fonte: Autoria Própria (2021), baseado na Classificação Brasileira de Ocupações (CBO) (SÃO PAULO, 2018).

Após a análise descritiva das variáveis, inicia-se a fase da MD. Desse modo, com a execução do algoritmo Apriori no software WEKA encontrou-se algumas regras de associações relevantes, como mostradas no Quadro 4.

Quadro 4: Regras retornadas pelo algoritmo Apriori.

N.º	SE =>	ENTÃO	Confiança
1	PARTO=1 e ASSISTMED=1 e CLCAUSA=2 (274 registros)	TIPOCAUSA=1 (274 registros)	100%
2	PARTO=2 e ASSISTMED=1 e CLCAUSA=2 (172 registros)	TIPOCAUSA=1 (172 registros)	100%
3	QTDFILMORT=1 e NECROPSIA=2 (215 registros)	ASSISTMED=1 (208 registros)	97%
4	GESTACAO=2 e CLCAUSA=2 (261 registros)	OBITOPARTO=3 e PESO=<1000 (236 registros)	90%
5	ESMAE2010=3 e OCUPMAE=999992 (318 registros)	MORTEPARTO=3 (290 registros)	91%
6	OCUPMAE=621005 (registros 182)	MORTEPARTO=3 (registros 174)	96%

Fonte: Autoria Própria (2021).

Analisando-se os dados apresentados no Quadro 4, a regra 1 demonstra a ligação entre as 274 mortes infantis evitáveis por atenção à mulher na gestação associada ao tipo de parto vaginal, mesmo com a assistência médica à mãe e aos recém-nascidos. Já a regra 2 descreve uma redução desse número para 172, quando os partos foram do tipo cesáreas. Cabe destacar que o grau de confiança de ambas as regras é de 100%.

Assim sendo, nos partos do tipo cesáreos, apesar da assistência médica prestada a paciente grávida, ocorreram 172 mortes infantis que poderiam ter sido evitadas por atenção à mulher durante a gestação. Já nos partos do tipo vaginal esse número saltou para 274. Logo, observa-se uma maior frequência de óbitos infantis no parto do tipo vaginal em relação ao parto cesáreo. Por isso, é importante melhorar a atenção pré-natal, garantir o acesso das gestantes aos serviços de saúde e investir na qualificação da assistência médica envolvida nos cuidados da mãe e do bebê (LANSKY *et al*, 2014).

Além disso, os dados demonstram que, quando a quantidade de filhos mortos foi igual a um, em 215 dessas ocorrências não foram realizados exames de necropsia e houve assistência médica durante a enfermidade que provocou o óbito. É isso que apresenta a regra 3 com um nível de confiança

de 97%. Essa falta de investigação prejudica, inclusive, o diagnóstico de confirmação dos óbitos desses menores de um ano de idade. O curioso, no entanto, é que 208 dessas crianças receberam assistência médica durante as enfermidades que ocasionaram os óbitos.

Na regra 4, que tem 90 % de confiança, identificou-se uma associação interessante acerca da quantidade de crianças nascidas antes da gestação completa, especificamente, entre a 22^a e a 27^a semanas de gestação, todos esses recém-nascidos tinham um peso inferior a 1000 gramas. Esses bebês vieram a óbito depois de parto. Contudo, esses falecimentos poderiam ter sido evitados por atenção à mulher na gestação. Vale esclarecer que os bebês que nascem entre a 25^a e a 27^a semanas de gestação são considerados “prematurados extremos”. Nesse sentido, o baixo peso ao nascer é uma outra dificuldade a ser enfrentada pelos recém-nascidos que chegam ao mundo antes do desenvolvimento completo de vários órgãos (SZCZESNY, 2021).

Desse modo, vale alertar que o problema da prematuridade se relaciona com a qualidade da assistência pré-natal e com as condições psicossociais da gestante: o estresse psicológico na gravidez, por exemplo, aumenta as chances de ocorrência da prematuridade (BEZERRA; OLIVEIRA; LATORRE, 2006; FRANÇA; LANSKY, 2016). Logo, é importante o desenvolvimento de políticas públicas que melhorem a qualidade da atenção às gestantes e aos bebês.

Ademais, conforme a regra 5 com 91% de confiança, há associações entre gestantes que trabalham como donas de casa com nível médio completo e a ocorrência de mortes de crianças menores de um ano de idade. Além disso, o exercício de atividades relacionadas à agropecuária pela mãe, também, indica chances altas de ocorrência de mortes infantis de acordo com a regra 6 com 96% de confiança. O fato é que essas mães só com nível médio completo, provavelmente, encontraram dificuldades em conseguir uma boa colocação no mercado de trabalho cada vez mais exigente em termos de qualificação profissional; além disso, empregos na área da agropecuária podem representar riscos à saúde da grávida devido aos riscos de acidentes no ambiente de trabalho e da contaminação por agrotóxicos.

Diante do exposto, este estudo encontrou conhecimentos relevantes a partir das regras de associações geradas entre as variáveis estudadas. Além de revelar padrões nos dados do cenário investigado, os achados demonstram como o processo de descoberta de conhecimento e a MD podem ser úteis para a formação de políticas públicas destinadas à redução da mortalidade infantil pernambucana, pois indicam lacunas no acompanhamento das gestantes que reduziriam as causas de mortes evitáveis.

5. CONSIDERAÇÕES FINAIS

Efetivamente, este estudo possibilitou encontrar conhecimentos úteis na base de dados do Sistema de Informação Sobre Mortalidade de 2018, os quais permitiram a compreensão do cenário investigado e podem contribuir no desenvolvimento de ações e políticas públicas destinadas à redução da mortalidade infantil pernambucana. Nesse sentido, é preciso ressaltar que o Estado de Pernambuco não é inerte em relação ao problema da mortalidade infantil, prova disso é que em 2007 foi criado o Programa Mãe Coruja Pernambucana, uma política pública estadual empregada na atenção integral das gestantes e das crianças, até os cinco anos de idade. Mesmo com essa atitude, persiste-se a dificuldade de reduzir as ocorrências de mortes infantis evitáveis por atenção à mulher na gestação.

Por isso, uma boa maneira de enfrentar esse problema é por meio do aperfeiçoamento do sistema público de saúde estadual e municipal. Ora, recursos governamentais podem ser empregados na construção de novas maternidades ou na restauração das antigas, na aquisição de modernas incubadoras neonatais, bem como na qualificação continuada dos profissionais da saúde envolvidos diretamente com a atenção integral das gestantes e dos recém-nascidos (na assistência médica e realização de exames, por exemplo).

Associações entre os dados mostraram que muitos dos óbitos infantis ocorreram quando as mães desses bebês tinham o ensino médio completo e eram donas de casa. É importante mencionar questões como os riscos da gravidez não planejada, a dificuldade de acesso aos serviços de saúde e a pouca qualificação profissional que restringe as chances de um emprego bem remunerado.

No quesito atividades durante a gravidez, os dados mostraram associações de falecimentos de crianças com menos de um ano, filhos de mães que trabalhavam na agropecuária. Sabe-se que essa profissão pode oferecer riscos à saúde das grávidas, especialmente, em função dos acidentes no trabalho, esforços físicos e de um eventual contato com agrotóxicos. Por isso a importância de acesso à saúde, educação e orientação para esse grupo de mulheres fornecido pelo governo que, bem implementado, ajudaria a mitigar os casos evitáveis.

Como se vê, os dados públicos disponibilizados abertamente por órgãos governamentais viabilizam a transparência pública e estimulam a participação da sociedade nas ações de interesse coletivo. Isso engrandece a gestão democrática dos serviços públicos. Assim sendo, este estudo trouxe, de um desses banco de dados abertos, conhecimentos relevantes acerca da mortalidade infantil em território pernambucano.

Contudo, isso não foi feito de maneira ampla e como trabalhos futuros recomenda-se a delimitação desse estudo a partir das cidades e regiões do Estado de Pernambuco, investigando particularidades dos contextos analisados. Sugere-se, também, que se relacionem os dados provenientes do Sistema de Informação Sobre Mortalidade (SIM) com outras bases de dados abertos, tais como o Sistema de Informações sobre Nascidos Vivos (SINASC), o Número de Equipes de Agentes Comunitários de Saúde (ACS), das Distribuição Unidades Básicas de Saúde em Funcionamento (UBS) e outros dados gerados por secretarias e órgãos da gestão municipal com o propósito de refinar o processo de descoberta de conhecimento.

REFERÊNCIAS

ALVES, J. B. *et al.* Sepsis Neonatal: Mortalidade em município do sul do Brasil, 2000 A 2013. **Revista Paulista de Pediatria**. São Paulo, v. 36, n. 2, p. 132-140, 2018.

BALANIUK, R. A Mineração de Dados como apoio ao Controle Externo. **Revista do TCU**, n. 117, p. 79-86, 2010.

BEZERRA, L. C.; OLIVEIRA, S. M. J. V.; LATORRE, M. R. D. O. Prevalência e fatores associados à prematuridade entre gestantes submetidas à inibição de trabalho de parto prematuro. **Revista Brasileira de Saúde Materno Infantil**. Recife, v. 6, n. 2, p. 223-229, 2006.

BRASIL NETO, C. R. **Previsão de incidência de Dengue nas cidades brasileiras através de Inteligência Artificial**. 2009. 78 f. Dissertação (Mestrado em ciência da computação) -Universidade Federal de Pernambuco. Centro de Informática. Recife, 2009.

BRASIL. **Manual de Procedimentos do Sistema de Informação sobre Mortalidade**. 1º. ed. Brasília/DF: Ministério da Saúde. Fundação Nacional de Saúde, 2001a.

BRASIL. **Manual de instruções para o preenchimento da declaração de nascido vivo**. 3ª. ed. Brasília/DF: Ministério da Saúde: Fundação Nacional da Saúde, 2001b.

BRASIL. Lei n. 12.527, de 18 de nov. de 2011. Regula o acesso à informação. **Diário Oficial da União**, Brasília/DF, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 08 dez. 2020.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Instituto Brasileiro de Geografia e Estatística-IBGE. **Tábua completa de mortalidade para o Brasil – 2015**. Rio de Janeiro: IBGE, 2016a.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. **Diário Oficial da União**, Brasília/DF, 11 maio 2016b. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 22 dez. 2020.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Instituto Brasileiro de Geografia e Estatística-IBGE. **Tábua completa de mortalidade para o Brasil – 2016**. Rio de Janeiro: IBGE, 2017.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Instituto Brasileiro de Geografia e Estatística-IBGE. **Tábua completa de mortalidade para o Brasil – 2017**. Rio de Janeiro: IBGE, 2018.

BRASIL. Ministério da Saúde. **Saúde Brasil 2019 uma análise da situação de saúde com enfoque nas doenças imunopreveníveis e na imunização**. Brasília/DF: Ministério da Saúde, 2019a. 520p.:il.

BRASIL. Ministério da Economia. Instituto Brasileiro de Geografia e Estatística-IBGE. **Tábua completa de mortalidade para o Brasil - 2018**. Rio de Janeiro: IBGE, 2019b.

BRASIL. Ministério da Saúde. Exposição no trabalho e no ambiente ao agrotóxico. **Instituto Nacional Do Câncer -INCA**, 2019c. Disponível em: <https://www.inca.gov.br/exposicao-no-trabalho-e-no-ambiente/agrotoxicos>. Acesso em: 18 jan. 2021.

BRASIL. **Análise de Dados em Linguagem R**. Brasília/DF: Fundação Escola Nacional de Administração Pública, 2020a.

BRASIL. Ministério da Economia. Portal Dados Abertos. **Comissão de Valores Mobiliários**. Brasília/DF, 08 de out. 2020b. Disponível em: <https://www.gov.br/cvm/pt-br/aceso-a-informacao-cvm/dados-abertos>. Acesso em: 14 dez. 2020.

BRASIL. Estrutura do Sistema de Informação sobre Mortalidade - SIM (1979 a 2018). **Portal Brasileiro de Dados Abertos**, 2020c. Disponível em: <https://dados.gov.br/dataset/sistema-de-informacao-sobre-mortalidade-sim-1979-a-2018/resource/4d858661-073a-485f-8a5d-3c496a8c6354>. Acesso em: 05 jan. 2021.

CABRAL, A.L.N. **Avaliação do impacto do Programa Mãe Coruja Pernambucana nas taxas de mortalidade infantil e materna em Pernambuco**. 2016. 49 f. Dissertação (Mestrado em Gestão e Economia da Saúde) - Programa de Pós-graduação em Gestão e Economia da Saúde, Universidade Federal de Pernambuco, Recife, 2016.

CASTRO, L. N.; FERRARI, D. G.; **Introdução a mineração de dados**. São Paulo: Saraiva, 2016

CASTRO, R.F.V. **Análise de desempenho dos algoritmos Apriori e Fuzzy Apriori na extração de regras de associação aplicados a um Sistema de Detecção de Intrusos**. 2014. 100 f. Dissertação (Mestrado em engenharia eletrônica) - Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

CORREA, A. C. *et al.* FarmaJusta: Dados Abertos Como Suporte à Saúde Pública. In: **Anais da V Escola Regional de Sistemas de Informação do Rio de Janeiro**. SBC, 2018. p. 9-15.

COSTA, J. J.; BERNARDINI, F. C.; VITERBO FILHO, J. A mineração de dados e a qualidade de conhecimentos extraídos dos boletins de ocorrência das rodovias federais brasileiras. **AtoZ: novas práticas em informação e conhecimento**, [S.l.], v. 3, n. 2, p. 139-157, dez. 2014

CUNHA, E. M.; VARGENS, J. M. C. Sistemas de informação do Sistema Único de Saúde. In: GONDIM, Grácia Maria de Miranda; CHRISTÓFARO, Maria Auxiliadora Córdova; MIYASHIRO, Gladys Miyashiro (Org.). **Técnico de vigilância em saúde: fundamentos**. Rio de Janeiro: EPSJV, 2017. p. 71-112.

FAYYAD, U. M. *et al.* Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: **KDD**. 1996. p. 82-88.

FERRAZ, L. H. V. C. **O SUS, o DATASUS e a informação em saúde: uma proposta de gestão participativa. 2009**. 109 f. dissertação (Mestre Modalidade Profissional em Saúde Pública) – Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2009.

FRANÇA, E.; LANSKY, S. Mortalidade infantil neonatal no Brasil: situação, tendências e perspectivas. **Anais**, p. 1-29, 2016.

FUNDAÇÃO ABRINQ. Taxa de mortalidade infantil (para cada mil nascidos vivos). **Observatório da Criança e do Adolescente**, São Paulo/SP. c2015. Disponível em: <https://observatoriocrianca.org.br/cenario-infancia/temas/sobrevivencia-infantil-infancia/620-taxa-de-mortalidade-infantil-para-cada-mil-nascidos-vivos?filters=1,234>. Acesso em: 07 jan. 2021.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, v. 22, n. 5, p. 686-690, 2009.

GAVA, C.; CARDOSO, A. M.; BASTA, P. C. Mortalidade infantil por cor ou raça em Rondônia, Amazônia Brasileira. **Revista de Saúde Pública**, v. 51, p. 35, 2017.

GERHARDT, T. E.; SILVEIRA, D. T. (org). **Métodos de Pesquisa**, Porto Alegre: UFRGS, 2009.

GONÇALVES, O. L.; BELTRAME, W. A. R. Mineração de dados e Evasão estudantil: Analisando o curso de nível superior do Ifes. In: **XII CASI - Congresso de Administração, Sociedade e Inovação**. 2019

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3ª. ed. Waltham: Elsevier, 2012.

JUNIOR, R. A. L. *et al.* Mineração de Dados Abertos. Anais da IV Jornada Científica de Sistemas de Informação. Faculdade Maurício de Nassau. Parnaíba - PI. 2014.

KAUARK, F.; MANHÃES, C. F.; MEDEIROS, H. C. **Metodologia da pesquisa: guia prático**. Itabuna: Via Litteral EDITORA, 2010. 88 p.

KROPIWIEC, M. V.; FRANCO, S. C.; AMARAL, A. R. Fatores associados à mortalidade infantil em município com índice de desenvolvimento humano elevado. **Revista Paulista de Pediatria**, v. 35, n. 4, p. 391-398, 2017.

LAMY, M. **Metodologia da pesquisa: técnicas de investigação, argumentação e redação**, 2º ed. São Paulo: Matrioska Editora, 2020.

LANSKY, S. *et al.* Pesquisa Nascido no Brasil: perfil da mortalidade neonatal e avaliação da assistência à gestante e ao recém-nascido. **Cad. Saúde Pública**, Rio de Janeiro, v. 30, supl. 1, p. 192-207, 2014.

MAIA, L. T. S.; SOUZA, W. V.; MENDES, A. C. G. Determinantes individuais e contextuais associados à mortalidade infantil nas capitais brasileiras: uma abordagem multinível. **Cadernos de Saúde Pública**, v. 36, p. 1-19, 2020.

MALTA, D. C. *et al.* Lista de causas de mortes evitáveis por intervenções do Sistema Único de Saúde do Brasil. **Epidemiol. Serv. Saúde**, Brasília, v. 16, n. 4, p. 233-244, dez. 2007.

MARCHIORI, P. M.; FERRAZ, L. Gestante agricultora: agravos e riscos ocupacionais. **Revista de Terapia Ocupacional da Universidade de São Paulo**, 27, n. 2, p. 190-198, 2016.

MARINHO, G. L. *et al.* Mortalidade infantil de indígenas e não indígenas nas microrregiões do Brasil. **Revista Brasileira de Enfermagem**, v. 72, n. 1, 2019.

MEDICINANET. **Lista Geral CID-10.** Disponível em: https://www.medicinanet.com.br/categorias/lista_cid10.htm. Acesso em: 05 jan. 2021

NASCIMENTO, R. L. S. **Mineração de Dados Educacionais e Visualização de Informações Geográficas Utilizando Mapas de Calor.** Trabalho de Conclusão de Curso (Monografia) - Universidade Federal Rural de Pernambuco. Recife-PE, 2016.

NASCIMENTO, R. L. S. *et al.* Mineração de dados na identificação de empresas irregulares quanto ao pagamento de impostos. **Revista de Engenharia e Pesquisa Aplicada**, v. 3, n. 3, 2018.

OPEN KNOWLEDGE BRASIL. **Por que open?** Disponível em: <https://www.ok.org.br/dados-abertos/>. Acesso em: 08 dez. 2020.

PERNAMBUCO. Decreto n. 30.859, de 04 de out. de 2007. Cria o Programa Mãe Coruja Pernambucana, e dá outras providências. **Alepe Legis: Legislação Estadual de Pernambuco**, Recife/PE, 04 out. 2007. Disponível em: <https://legis.alepe.pe.gov.br/texto.aspx?tiponorma=6&numero=30859&complemento=0&ano=2007&tipo=&url=>. Acesso em: 03 jan. 2020.

PERNAMBUCO. Lei n. 13.959, de 15 de dezembro de 2009. Dispõe sobre o Programa Mãe Coruja Pernambucana. **Alepe Legis: Legislação Estadual de Pernambuco**, Recife/PE, 15 dez. 2009. Disponível em: <http://legis.alepe.pe.gov.br/texto.aspx?id=3760>. Acesso em: 07 jan. 2021.

PROVOST, F.; FAWCETT, T. **Data Science para negócios: o que você precisa saber sobre mineração de dados e pensamento analítico de dados.** Rio de Janeiro, RJ: Alta Books, 2016.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, M. P. B. **Metodologia de pesquisa.** 5. ed. Porto Alegre: Penso, 2013. 624 p.

SANTOS, D. C. *et al.* Descoberta de Conhecimento em Dados Abertos sobre Acidentes de Trânsito com Ciclistas em Porto Alegre. **International Journal of Knowledge Engineering and Management (IJKEM)**, v. 7, n. 17, p. 25-48, 2018.

SÃO PAULO. **Tabelas de codificação utilizadas no Sinan-Acidentes e doenças relacionadas ao trabalho.** São Paulo/SP: [s.n.], 2018. Disponível em: https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/man_cod_sinan_2018_versao_0210_1538489641.pdf. Acesso em: 30 jan. 2021.

SARTORELLI, A. P. *et al.* Fatores que contribuem para a mortalidade infantil utilizando a mineração de dados. **Saúde e Pesquisa**, v. 10, n. 1, p. 33-41, 2017.

SHARDA, R.; DELEN, D.; TURBAM, E. **Business Intelligence e Análise de Dados para Gestão do Negócio.** 4. ed. Porto Alegre/RS: Bookman, 2019.

SILVA, E. S. A.; PAES, N. A.; SILVA, C. C. Efeitos dos programas governamentais e da fecundidade sobre a mortalidade infantil do Semiárido brasileiro. **Saúde em Debate**, v. 42, p. 138-147, 2018.

SILVA, L. A.; MORINO, A. H.; SATO, T. M. C. Prática de mineração de dados no exame nacional do ensino médio. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação.** 2014. p. 651.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à mineração de dados: com aplicações em R.** Rio de Janeiro/RJ: Elsevier, 2016.

SILVA, P. *et al.* Modelos de regressão aplicados a predição do desempenho escolar de estudantes do ensino fundamental. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE).** 2019. p. 1621.

SILVA, R. I. Â.; GALVÃO, M. C. A importância do uso de dados abertos pelo poder público. **IX Seminário e Ensino, Pesquisa e Extensão**, Ceres/UFRN, 28,29 e 30 de maio 2018.

SOUTO, H. M.; ARRUDA, E.M; ARAÚJO, W. J. Mineração de dados no contexto dos pregões eletrônicos. **Informação em Pauta**, v. 4, n. especial 2, p. 47-64, 2 nov. 2019.

SOUZA, P. M. *et al.* Repositório DATASUS: Organização e relevância dos dados abertos em saúde para a vigilância epidemiológica. **P2P E INOVAÇÃO**, v. 6, n. 1, p. 50-59, 2019.

SZCZESNY, L. Prematuros: uma luta pela vida. **Associação Brasileira de Pais, Familiares, Amigos e Cuidadores de Bebês Prematuros**, 2021. Disponível em: [https://www.prematuridade.com/index.php/interna-post/prematuros-uma-luta-pela-vida-8150#:~:text=Os%20prematuros%20extremos%20s%C3%A3o%20aqueles,os%20seus%20C3%B3rg%C3%A3os%20est%C3%A3o%20formados](https://www.prematuridade.com/index.php/interna-post/prematuros-uma-luta-pela-vida-8150#:~:text=Os%20prematuros%20extremos%20s%C3%A3o%20aqueles,os%20seus%20C3%B3rg%C3%A3os%20est%C3%A3o%20formados.). Acesso em: 28 jan. 2021.

TESINI, B. L. Sepsis no recém-nascido. **Manual MSD Versão Saúde para a Família**, 2020. Disponível em: <https://www.msmanuals.com/pt-br/casa/problemas-de-sa%C3%BAde-infantil/infec%C3%A7%C3%B5es-em-rec%C3%A9m-nascidos/sepsis-no-rec%C3%A9m-nascido>. Acesso em: 14 jan. 2021.

UNICEF *et al.* Levels and trends in child mortality 2019. **Unicef for every child**, 2019. Disponível em: <https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2019>. Acesso em: 11 jan. 2021.

VASCONCELOS, L. M. R.; CARVALHO, C. L. Aplicação de regras de associação para mineração de dados na web. **Revista Telfract**, v. 1, n. 1, 2018.

VIANNA, R. C. X. F. *et al.* Mineração de dados e características da mortalidade infantil. **Cadernos de Saúde Pública**, v. 26, p. 535-542, 2010.

WITTEN, I. *et al.* **Data Mining Practical Machine Learning Tools and Techniques**. Amsterdam. Boston. Heidelberg. London. New York. Oxford. 4^a.ed. Paris. San Diego. San Francisco. Singapore. Sydney. Tokyo: Elsevier, 2016.