

Óbitos infantis entre os anos 2000 e 2017 em uma cidade do sul do Brasil: técnicas de mineração de dados

Infant mortality between the years 2000 and 2017 in a city in southern Brazil: data mining techniques

Muertes infantiles entre los años 2000 y 2017 en una ciudad del sur de Brasil: técnicas de minería de datos

Recebido: 12/08/2020 | Revisado: 21/08/2020 | Aceito: 26/08/2020 | Publicado: 29/08/2020

Márian Oleques Pires

ORCID: <https://orcid.org/0000-0002-3162-3337>

Universidade Franciscana, Brasil

E-mail: piresoleques@gmail.com

Sylvio Andre Garcia Vieira

ORCID: <https://orcid.org/0000-0002-1484-4728>

Universidade Franciscana, Brasil

E-mail: sylvio@ufn.edu.br

Dirce Stein Backes

ORCID: <https://orcid.org/0000-0001-9447-1126>

Universidade Franciscana, Brasil

E-mail: backesdirce@ufn.edu.br

Maclaine de Oliveira Roos

ORCID: <https://orcid.org/0000-0002-4747-4196>

Universidade Franciscana, Brasil

E-mail: pim4crs@gmail.com

Nathalia Adames

ORCID: <https://orcid.org/0000-0003-2273-3857>

Universidade Franciscana, Brasil

E-mail: natiadames@gmail.com

Resumo

Objetivo: Analisar padrões de óbitos infantis entre os anos 2000 e 2017 em uma cidade do sul do Brasil. Método: A partir da extração de dados da base de mortalidade infantil do

Departamento de Informática do Sistema Único de Saúde (DATASUS), aplicar técnicas de análise de dados baseados no método de Descoberta de informações, incorporando, na fase de mineração de dados, algoritmos de aprendizado de máquina provindos do software WEKA. Resultados: O estudo evidencia a associação entre o baixo peso ao nascer e reforça a sua manutenção como critério isolado para a identificação de crianças vulneráveis ao óbito em menores de um ano. Apesar das lacunas encontradas na base de dados do sistema de bases de mortalidade, evidencia que esta é uma potencial fonte de informação para a detecção de fatores associados à mortalidade infantil. Considerações Finais: Algoritmos computacionais são eficazes para realizar associações, agrupamentos e previsões de maneira eficaz em dados de saúde, no entanto, a base de dados necessita ser alimentada com dados confiáveis regularmente. **Palavras-chave:** Mortalidade Infantil; Mineração de dados; Sistemas de informação.

Abstract

Objective: To analyze infant death patterns between the years 2000 and 2017 in a city in southern Brazil. Method: From the data extraction from the infant mortality database of the Informatics Department of the Unified Health System (DATASUS), apply data analysis techniques based on the Knowledge Discovery in Databases, method incorporating, in the data mining phase, machine learning algorithms from WEKA software. Results: The study highlights the association between low birth weight and reinforces its maintenance as an isolated criterion for the identification of children vulnerable to death in children under one year old. Despite the gaps found in the database of the mortality base system, it shows that this is a potential source of information for detecting factors associated with infant mortality. Final Considerations: Computational algorithms are effective to carry out associations, clusters and predictions effectively on health data, however, the database needs to be fed with reliable data regularly.

Keywords: Infant mortality; Data Mining; Information systems.

Resumen

Objetivo: Analizar patrones de muerte infantil entre los años 2000 y 2017 en una ciudad del sur de Brasil. Método: A partir de la extracción de datos de la base de datos de mortalidad infantil del Departamento de Informática del Sistema Único de Salud (DATASUS), aplicar técnicas de análisis de datos basadas en el método Descubrimiento de Información, incorporando, en la fase de minería de datos, algoritmos de aprendizaje automático del software WEKA. Resultados: El estudio destaca la asociación entre bajo peso al nacer y refuerza su

mantenimiento como criterio aislado para la identificación de niños vulnerables a la muerte en menores de un año. A pesar de las brechas encontradas en la base de datos del sistema base de mortalidad, se muestra que esta es una fuente potencial de información para detectar factores asociados a la mortalidad infantil. Consideraciones finales: Los algoritmos computacionales son efectivos para realizar asociaciones, clusters y predicciones de manera eficiente en datos de salud, sin embargo, la base de datos necesita ser alimentada con datos confiables de manera regular.

Palabras clave: Mortalidad infantil; Procesamiento de datos; Sistemas de información.

1. Introdução

No decorrer das últimas décadas, com o avanço de tecnologias computacionais, tornou-se simples a tarefa de coletar e armazenar dados. Em consequência, a maneira de realizar a análise de dados adequou-se com o objetivo de focar na extração de informações úteis em bases de dados, por meio de técnicas computacionais específicas para cada tipo de dados (Marquesone, 2016). Este processo de descoberta automática de informações, combinando métodos tradicionais de análise de dados com algoritmos específicos, em grandes repositórios de dados, pode ser chamada de mineração de dados (Tan, Steinbach & Kumar, 2009).

Diariamente, Unidades Básicas de Saúde (UBS), hospitais e clínicas de saúde produzem uma grande quantidade de dados relacionados a prontuários de pacientes, leitos, estoques e outros. A partir destes dados, é possível obter informações como perfis de pacientes, tipos de doenças, bem como a qualidade de administração desses estabelecimentos. No Brasil, desde 1991, o Departamento de Informática do Sistema Único de Saúde (DATASUS) é o órgão responsável pela captação, a integração operacional e a divulgação dos dados de estabelecimentos de saúde pública (Datusus, 2017), os dados são disponibilizados, publicamente, pela plataforma de informações de saúde TABNET.

Entre os anos de 1982 e 2005, a taxa de mortalidade infantil (TMI), no Brasil, apresentou uma queda de aproximadamente 80,4%, em decorrência de medidas públicas e diversas ações estratégicas relacionadas à melhoria das condições de vida e assistência à saúde (Ministério da Saúde 2016). Entretanto, taxas de mortalidade em crianças que vivem em áreas urbanas, crianças indígenas e regiões periféricas, ainda, permanecem com a TMI acima do recomendado pela ONU.

O fato da TMI ser maior em regiões vulneráveis, reflete a qualidade de saúde populacional das pessoas que vivem nessas regiões. A mortalidade infantil é um indicador que

está associado, além dos fatores biológicos, sociais e culturais, visto que com a tecnologia atual e acompanhamento devido no período perinatal, é possível minimizar este fenômeno (França e Lanksy, 2016). Devido a abrangência do tema, sua análise se torna um instrumento para: definir prioridades nas políticas públicas, planejar ações e serviços, monitorar a situação de saúde, bem como avaliar o desempenho do sistema de saúde de uma região (Paiz, 2018).

Dispor de ferramentas como o DATASUS, propicia a análise de indicadores pelos gestores de saúde e técnicas, considerando que a mineração de dados possibilita o desenvolvimento de estratégias prospectivas no âmbito da organização e gestão em saúde (Jothi, 2015). Compreendendo a extensão, a complexidade e a importância do tema, o presente estudo tem como foco utilizar técnicas de mineração de dados, no sentido de identificar padrões de óbitos infantis na cidade de Santa Maria, Rio Grande do Sul (RS), no período de 2000 a 2017 (último ano disponibilizado pela ferramenta DATASUS).

Com base no exposto, teve-se por objetivo geral do estudo explorar técnicas de mineração de dados para analisar padrões de óbitos infantis ocorridos entre os anos de 2000 a 2017 em uma cidade do sul do Brasil e, por objetivos específicos: Preparar a base de dados de mortalidade do DATASUS para que o conhecimento possa ser extraído; Definir técnicas computacionais adequadas com os dados utilizados; Utilizar a ferramenta Weka para identificar regras de associação que possam auxiliar na identificação dos perfis; Avaliar entre os algoritmos utilizados qual expõe melhores resultados na base trabalhada; Analisar os resultados e identificar os perfis das crianças que faleceram com até um ano de idade.

2. Metodologia

Trata-se de uma pesquisa descritiva, para a qual se utilizou técnicas de mineração de dados. Considerou-se, para tanto, um conjunto de dados locais, possibilitando sua análise e reflexão sobre o tema abordado (Prodanov & de Freitas, 2013).

Utilizou-se, para a coleta de dados, documentos de uma plataforma online de publicação de dados e indicadores básicos de saúde brasileiros (explanado no tópico Extração de Dados). Aplicou-se técnicas computacionais, com a finalidade de obter conhecimento sobre uma base de dados e interpretou-se os dados de acordo com a realidade do local estudado (Pereira, Shitsuka, Parreira & Shitsuka, 2018). O processo metodológico da pesquisa *knowledge discovery in databases - KDD*, propõe cinco etapas sequenciais e dinâmicas na extração de informações úteis em um conjunto de dados: extração de dados, pré-processamento de dados,

transformação, mineração de dados e interpretação dos dados, detalhados nos seguintes tópicos (França, Carvalho & Tsunoda, 2016).

Extração de Dados

O processo de extração de dados consiste em coletar os dados necessários para cumprir o objetivo determinado (Tan, Steinbach & Kumar 2009). As bases utilizadas nesta pesquisa foram extraídas do portal de serviços do DATASUS, referente aos arquivos disseminados para tabulação do Sistema de Mortalidade (SIM). A partir da página de arquivos, foram extraídas bases de dados de declarações de óbitos infantis no período de 2000 até o ano mais recente publicado em domínio público (2017), do estado do Rio Grande do Sul, constituindo-se, assim, uma amostra de 1560 óbitos infantis. Estes dados foram convertidos para um arquivo e incluíse, apenas, os óbitos da Cidade de Santa Maria.

Pré-Processamento e Transformação de Dados

Segundo Tan, Steinbach & Kumar (2009) a etapa referente ao pré-processamento e transformação de dados inclui procedimentos de fusão de dados de múltiplas fontes, limpeza de dados para remoção de ruídos, observações duplicadas, além da seleção de registros e de características relevantes para o estudo.

A base contendo o conjunto de dados de mortalidade do ano 2000 possui, inicialmente, 40 classes. No decorrer dos anos, ocorreram atualizações no SIM, com classes acrescentadas e modificadas até seu último ano de publicação (2017), contendo 92 classes na base. Entre as classes incluídas na base, ao longo dos anos, pode-se citar: Escolaridade da Mãe a partir de 2010, CRM do médico que atendeu o paciente no estabelecimento, atestado do óbito, necropsia (caso tenha sido realizada ou não), informações do pai da criança, entre outros.

Por meio da linguagem de programação *python* e auxílio da biblioteca *pandas*, as bases de todos os anos de mortalidade infantil de Santa Maria foram integralizados e, por meio desta, realizou-se a primeira análise utilizando a ferramenta WEKA, para normalização dos dados.

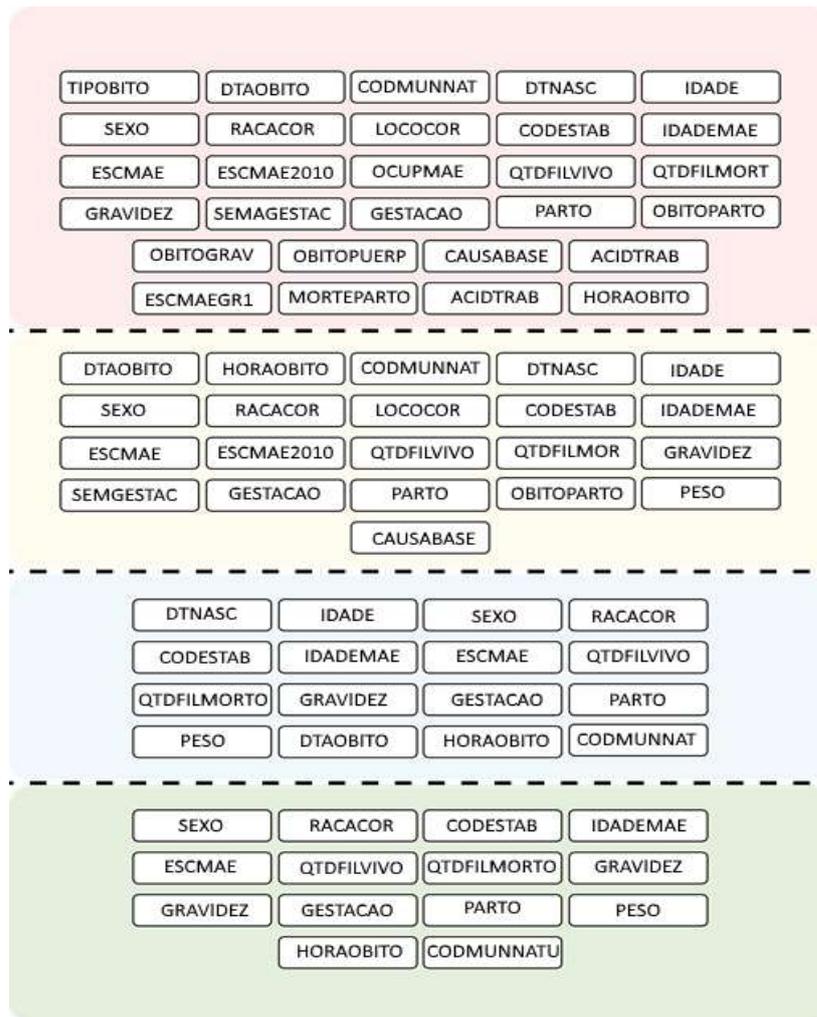
Ao denotar a inconstância nas classes das bases de dados, a partir da revisão de literaturas de (Sartorelli et. al. 2017) e (Ferreira et. al.2012), selecionou-se as classes com maior relevância, que permaneceram fixas no processo de análise do conjunto de dados, tais como: idade da mãe, escolaridade materna, filhos nascidos vivos e mortos, tipo de parto, tipo de gravidez, peso ao nascer, idade gestacional e tempo de vida do recém-nascido que fora a óbito.

Em seguida, utilizando a ferramenta WEKA, organizou-se os dados para detectar outras classes relevantes. Com base em (Ferreira et. al. 2012), foram estipuladas as seguintes regras para o pré-processamento de dados:

- **Eliminação:** Classes nas quais possuíam dados em branco ou incompletos e identificadores foram removidos da base.
- **Integração:** Classes que possuíam dados repetidos, devido a atualizações do SIM, foram unificadas, determinada por a que possuir mais campos completos. Para a melhor análise, foi criado grupos de dados para classes que possuíam uma grande variedade de dados, como idades, datas e horários.
- **Transformação:** A fim de obter uma análise mais precisa do algoritmo aplicado, dados foram convertidos para tipo numérico para inteiro de acordo com a necessidade.

Das 92 classes originais da base de dados importada, por meio das regras aplicadas e literatura, resultaram 28 classes mostradas no primeiro processo da Figura 1, contendo as classes: Tipo Óbito, Data Óbito, Hora Óbito, Município Residência, Data Nascimento, Idade, Sexo, Raça, Local Ocorrência, Estabelecimento, Idade da Mãe, Escolaridade da Mãe, Escolaridade da Mãe em 2010, Ocupação Mãe, Quantidade de filhos mortos e vivos, Tipo de Gravidez, Semana de Gestaçã, Gestaçã, Tipo de Parto, Óbitos durante a gravidez, Causa Base das mortes, Acidente de trabalho, Escolaridade 2010 agregada e Morte no Parto.

Figura 1 – Progresso Pré-Processamento dos Dados.



Fonte: Autores.

Esta base inicial foi levada para a ferramenta WEKA, a fim de detectar associações possíveis entre as classes. Não tendo encontrando grandes relações, repetiu-se as regras do pré-processamento sucessivamente, até o último processo mostrado na Figura 1. Essa análise é necessária para entender quais classes influenciam mais no resultado dos algoritmos e, assim, aumentar sua acurácia e precisão ao mostrar os resultados.

Mineração de Dados

A mineração de dados corresponde ao processo de extração de conhecimento dos dados anteriormente pré-processados (Amaral, 2016). Os dados pré-processados foram submetidos a testes nos algoritmos de aprendizado de máquina do software WEKA. As árvores de decisão, J48 e REPTree, comumente aplicadas devido sua facilidade na interpretação de dados,

juntamente com o modelo preditivo cross validation (10 *folds*) utilizado para medir o desempenho e os progressos na aprendizagem do modelo, obtiveram os resultados mais satisfatórios dentre os demais algoritmos testados no software WEKA (Medeiros et al, 2016).

3. Resultados e Discussão

Com respectivamente 60% e 61% de acertos, os modelos J48 e REPTree, produziram 89 regras relacionadas a classe causa base das doenças. Das associações geradas a partir da amostra de dados de óbitos infantis entre 2000 e 2017 das classes selecionadas, podemos destacar:

- Dos óbitos ocorridos, 48% foram causadas por problemas relacionados a complicações no período perinatal e 15% decorrentes de malformações congênicas, deformidades e anomalias cromossômicas.
- Aproximadamente 44% dos partos foram do tipo cesáreo;
- Aproximadamente 37% das mães possuíam idade entre 21 e 27 anos e 25% entre 33 e 39 anos;

As principais regras constatadas pelo algoritmo foram relacionadas as classes idade (tempo de vida), parto (tipo de parto), local e peso do óbito das crianças. Das 89 regras geradas pelas árvores de decisão J48 e REPTree destacam-se:

- Óbitos ocorridos em menos de um dia, tendem a ser decorrentes a complicações no período perinatal;
- Óbitos ocorridos em menos de uma hora foram associados a peso menos de um quilo e a causa base decorrente a malformações congênicas, deformidades e anomalias cromossômicas.
- Partos do tipo vaginal foram associados a óbitos devido a doenças infecciosas e parasitárias e partos do tipo cesáreo foram associados aos óbitos a doenças no período perinatal.
- Óbitos ocorridos em menos de uma hora relacionaram-se a classe peso quando menor de 1 quilo.
- Crianças com mais de seis meses de vida que foram a óbito relacionados às Unidades de Tratamento Intensivo, isto é, crianças com mortes ocorridas por Malformações congênicas, deformidades e anomalias cromossômicas e causas perinatais.
- Em hospitais com UTI neonatal, em óbitos ocorridos até seis meses, foram encontradas relações com doenças respiratórias ou pulmonares quando a gestação da mãe foi maior de 28 semanas e para partos do tipo cesáreo.

Embora o Brasil tenha reduzido as taxas de mortalidade infantil, na última década, estudo demonstra corrobora com os achados deste estudo, ao argumentar que as diferenças estatísticas em relação às taxas de mortalidade nas crianças são reflexo das iniquidades sociais que ainda dificultam o acesso e o uso dos meios de promoção, proteção e recuperação da saúde, configurando-se, portanto, um dos melhores indicadores do nível de vida e bem-estar social de uma população (Batista & Cruz, 2015).

Na mesma direção, outro estudo evidencia que a associação entre o baixo peso ao nascer e mortalidade infantil demonstra a interação entre fatores biológicos e sociais e reforça a sua manutenção como critério isolado para a identificação de crianças vulneráveis ao óbito em menores de um ano. Logo, o baixo peso ao nascer está diretamente relacionado aos fatores socioeconômicos, além da situação materna que pode se apresentar desfavorável ao desenvolvimento saudável do conceito (Lansky et al., 2014).

Apesar das lacunas encontradas no SIM, essa base de dados se constitui em potencial fonte de informações para a detecção de fatores associados à mortalidade infantil. O pacto pela redução da mortalidade infantil, no Brasil, deve ser, portanto, um compromisso coletivo, que deve ser assumido pelo governo federal, mas também pelas instâncias locais, no sentido de prospectar estratégias prospectivas voltadas à promoção e proteção da saúde.

4. Considerações Finais

Teve-se como objetivo esta pesquisa, entender padrões de mortalidades ocorridas nos anos de 2000 a 2017 na cidade de Santa Maria, RS. Os dados utilizados foram extraídos do Sistemas de Informação de Mortalidade do DATASUS referentes a declarações de óbitos infantis.

Ainda que o Ministério da Saúde empenhe-se na coleta de informações mais precisas em relação aos dados públicos, visto pela atualização das bases no decorrer dos anos, são observadas inconsistências nas bases como dados incompletos, que poderiam propiciar uma boa análise social do local, referentes a códigos de estabelecimentos, código de ocupação da mãe e informações do pai da criança.

O uso das classificações do CID como critério de categorização, acabou encobrando as reais condições dos óbitos infantis, dificultando a interpretação provinda das associações dos algoritmos. Para a interpretação mais profundas das regras como: “partos do tipo cesáreo foram associados a óbitos a doenças no período perinatal”, “partos do tipo vaginal foram associados

a óbitos devido a doenças infecciosas e parasitárias” em próximas pesquisas, serão necessárias a criação de subcategorias para doenças.

Apesar de não haver uma distinção mais detalhada das classes e falta de dados precisos nas bases, podemos concluir que o algoritmo teve precisão em relacionar certos padrões de mortalidade infantil da cidade estudada, pois, há um consenso de que a maioria dos óbitos infantis ocorrem decorrente a complicações em períodos perinatais seguidos por malformações congênitas, deformidades e anomalias cromossômicas.

As regras de associação geradas pelas árvores J48 e REPTree foram relativas principalmente às classes: idade (tempo de vida de criança), peso, tipo de parto e estabelecimento. Dos padrões de acordo com a realidade da cidade estudada e literatura, podemos destacar a relação encontrada entre estabelecimentos com UTI neonatal e ligadas a óbitos ocorridos em mais de seis meses e a associação entre as classes com mortes relacionadas a Malformações congênitas, deformidades e anomalias cromossômicas e complicações no período perinatal. Tais regras geradas, foram de acordo com o padrão da literatura, o que reforça a ideia de que algoritmos computacionais podem ser utilizados para análise de dados da saúde, auxiliando na tomada de decisões relacionadas a políticas públicas.

Contudo, para que os algoritmos computacionais possam fazer suas associações, agrupamentos e previsões de maneira eficaz, é ressaltado possuir uma base de dados confiável e preenchida corretamente, o que destaca o desafio em pesquisas relacionadas a mineração de dados em saúde.

Verificou-se, nessa pesquisa, a influência do peso ao nascer com as causas de mortalidade infantil no local estudado. Este fator demonstra o perfil socioeconômico da realidade materna na região. Busca-se, em trabalhos futuros, o detalhamento das causas de mortes por meio da criação de uma subcategoria específica, a fim de gerar resultados mais precisos a partir de algoritmos e possibilitar a ampliação da faixa etária.

Referências

Amaral, F. (2016). *Introdução à Ciência de Dados: mineração de dados e big data*. Alta Books Editora.

Batista Filho, M., & Cruz, R. D. S. B. L. C. (2015). *A saúde das crianças no mundo e no Brasil*. *Revista Brasileira de Saúde Materno Infantil*, 15(4), 451-4.

Datasus 2017. *Departamento de Informática do Sistema Único de Saúde*, Disponível em <https://datasus.saude.gov.br/sobre-o-datasus/>

Ferreira, D., Oliveira, A., & Freitas, A. (2012). Applying data mining techniques to improve diagnosis in neonatal jaundice. *BMC medical informatics and decision making*, 12(1), 143.

França, G. E. F., Carvalho, D. R., & Tsunoda, D. F. (2016). Descoberta de padrões em ordens de serviço de tecnologia da informação em hospital. *Revista de Gestão em Sistemas de Saúde*, 5(1), 41-51.

Jothi, N., & Husain, W. (2015). Data mining in healthcare—a review. *Procedia computer science*, 72, 306-313..

Lansky, S., Friche, A. A. D. L., Silva, A. A. M. D., Campos, D., Bittencourt, S. D. D. A., Carvalho, M. L. D., ... & Cunha, A. J. L. A. D. (2014). Pesquisa Nascido no Brasil: perfil da mortalidade neonatal e avaliação da assistência à gestante e ao recém-nascido. *Cadernos de Saúde Pública*, 30, S192-S207.

Marquesone, R. (2016). *Big Data: Técnicas e tecnologias para extração de valor dos dados*. Editora Casa do Código.

Medeiros, L. B. D., Trigueiro, D. R. S. G., Silva, D. M. D., Nascimento, J. A. D., Monroe, A. A., Nogueira, J. D. A., & Leadebal, O. D. C. P. (2016). Integração entre serviços de saúde no cuidado às pessoas vivendo com aids: uma abordagem utilizando árvore de decisão. *Ciência & Saúde Coletiva*, 21, 543-552.

Menezes, A. M., Barros, F. C., Horta, B. L., Matijasevich, A., Bertoldi, A. D., Oliveira, P. D., ... & Pelotas Cohorts Study Group Barros Aluisio JD Bassani Diego G Wehrmeister Fernando C Gonçalves Helen Santos Iná S Murray Joseph Tovo-Rodrigues Luciana Assunção Maria Cecilia F Silveira Mariangela F Domingues Marlos Rodrigues Hallal Pedro RC. (2019). Stillbirth, newborn and infant mortality: trends and inequalities in four population-based birth cohorts in Pelotas, Brazil, 1982–2015. *International journal of epidemiology*, 48(Supplement_1), i54-i62.

Ministério da Saúde (2016). *Evolução da mortalidade na infância nos últimos 10 anos*, Disponível em: <http://www.saude.gov.br/images/pdf/2018/setembro/13/Oficina-mortalidade-materna-e-infantil-CIT-MESA-Ana-Nogales.pdf>

Paiz, J. C., Bigolin, M., Rosa, R. D. S., & Bordin, R. (2018). Mortalidade infantil e serviços de atenção primária à saúde em Porto Alegre (RS), Brasil. *Revista brasileira de medicina de família e comunidade*. Vol. 13, no. 40 (jan./dez. 2018), p. 1-13.

Pang-Ning, T., Steinbach, M., & Kumar, V. (2009). Introdução ao “data mining”. *Rio de Janeiro: Ciência Moderna*.

Pereira A.S. et al. (2018). *Metodologia da pesquisa científica*. [e-book]. Santa Maria/RS. Ed. UAB/NTE/UFSM. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.

Prodanov, C. C., & de Freitas, E. C. (2013). *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale.

Sartorelli, A. P., Gomes, D. C., Cubas, M. R., & Carvalho, D. R. (2017). Fatores que contribuem para a mortalidade infantil utilizando a mineração de dados. *Saúde e Pesquisa ISSN 2176-9206*, 10(1), 33-41.

Porcentagem de contribuição de cada autor no manuscrito

Márian Oleques Pires – 45%

Dirce Stein Backes – 20%

Sylvio Andre Garcia Vieira – 25%

Maclaine de Oliveira Roos – 5%

Nathalia Adames – 5%