PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Programa de Pós-Graduação em Informática

Wanderson Luiz Gomes Soares

O USO DE TÉCNICAS DE MACHINE LEARNING NA ANÁLISE DA MORTALIDADE INFANTIL: UM ESTUDO DE CASO DOS ESTADOS BRASILEIROS DE SANTA CATARINA E AMAPÁ

Belo Horizonte

2019

Wanderson Luiz Gomes Soares

O USO DE TÉCNICAS DE MACHINE LEARNING NA ANÁLISE DA MORTALIDADE INFANTIL: UM ESTUDO DE CASO DOS ESTADOS BRASILEIROS DE SANTA CATARINA E AMAPÁ

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Informática.

Orientadora: Profa. Dra. Cristiane Neri Nobre

Co-orientador: Prof. Dr. Mark Alan Junho Song

Belo Horizonte

2019

FICHA CATALOGRÁFICA
Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

Soares, Wanderson Luiz Gomes
S676u O uso de técnicas de *machine learning* na análise da mortalidade infantil: um estudo de caso dos estados brasileiros de Santa Catarina e Amapá / Wanderson Luiz Gomes Soares. Belo Horizonte, 2019.
84 f. : il.

Orientadora: Cristiane Neri Nobre
Coorientador: Mark Alan Junho Song
Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática

1. Mortalidade infantil - Banco de dados. 2. Aprendizado do computador. 3. Processamento eletrônico de dados. 4. Algoritmos. 5. Mineração de dados (Computação). 6. Mortalidade infantil – Estatísticas - Santa Catarina. 7. Mortalidade infantil – Estatísticas - Amapá. I. Nobre, Cristiane Neri. II. Song, Mark Alan Junho. III. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática. IV. Título.

CDU: 681.3.011
Ficha catalográfica elaborada por Fabiana Marques de Souza e Silva - CRB 6/2086

Wanderson Luiz Gomes Soares

USO DE TÉCNICAS DE MACHINE LEARNING NA ANÁLISE DA MORTALIDADE INFANTIL: UM ESTUDO DE CASO DOS ESTADOS BRASILEIROS DE SANTA CATARINA E AMAPÁ

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Informática.

Professora Dra. Cristiane Neri Nobre –
PUC Minas (Orientadora)

Professor Dr. Mark Alan Junho Song
PUC Minas (Co-orientador)

Professor Dr. Luis Enrique Zárate
PUC Minas (Banca Examinadora)

Professora. Dra. Déborah Ribeiro Carvalho
Pontifícia Universidade Católica do Paraná
(Banca Examinadora)

Belo Horizonte/MG, 09 de agosto de 2019.

Dedico esta dissertação:
A Deus e ao Senhor Jesus Cristo.
À professora Dra. Cristiane e ao professor Dr. Mark.

AGRADECIMENTOS

Agradeço...

A Deus e ao Senhor Jesus Cristo.

Agradeço à professora Dra. Cristiane Neri Nobre e ao professor Dr. Mark Alan Junho Song.

À secretária Giovana Cassia e ao funcionário Maurício Eustáquio Gomes.

Ao meu pai Soares (In memoriam).

À minha mãe Nedina.

À minha esposa Leiziane.

À minha família.

À amiga Virgínia Magalhães.

À amiga Thaiza.

Agradeço também a todos professores, alunos e funcionários da secretaria do programa de Pós-Graduação stricto sensu em Informática na PUC Minas São Gabriel.

O autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Muito obrigado!

"...A persistência é o menor caminho do êxito."

Charles Chaplin

RESUMO

A mortalidade infantil é caracterizada pela morte de crianças menores de um ano e é um problema que afeta todas as nações ao redor do mundo. Diante deste contexto, diferentes trabalhos utilizam técnicas de aprendizado de máquina visando identificar padrões que caracterizem a mortalidade infantil. O objetivo deste trabalho é empregar conceitos de descoberta de conhecimento em bancos de dados, especificamente de aprendizado de máquina na fase de mineração de dados, para caracterizar a mortalidade infantil nos estados de Santa Catarina e do Amapá que respectivamente apresentam a menor e a maior taxa de mortalidade infantil brasileira. Foram utilizados os algoritmos: C4.5, RIPPER, Random Forest, SVM e RNA, sendo feita uma comparação detalhada dos resultados obtidos nos dois estados. Os resultados mostram que as características que mais se destacaram no processo de caracterização da mortalidade infantil foram: peso, idade da mãe, gestação, quantidade de filho vivo, APGAR1, APGAR5 e escolaridade da mãe, cujos resultados corroboram com os outros trabalhos da literatura. Relata-se que foi constatada uma diferença bem significativa entre os bebês que sobreviveram após um ano de vida e os bebês que faleceram no período neonatal tendo como causa de óbito afecções originadas no períodos perinatal, cujas diferenças concentram basicamente nas seguintes características: PESO, APGAR e gestação.

Palavras-chave: Mortalidade Infantil, Classificação, Aprendizagem de Máquina.

ABSTRACT

Child mortality is characterized by the death of children under the age of one and is a problem that affects all nations around the world. In this context, different works use machine learning techniques to identify patterns that characterize infant mortality. The objective of this paper is to employ knowledge discovery concepts in databases, specifically machine learning in the data mining phase, to characterize child mortality in the states of Santa Catarina and Amapá, which respectively have the lowest and highest rates of Brazilian infant mortality. The algorithms C4.5, RIPPER, Random Forest, SVM and RNA were used, and a detailed comparison of the results obtained in both states was made. The results show that the characteristics that most stood out in the process of characterizing infant mortality were: weight, mother's age, pregnancy, number of live children, APGAR1, APGAR5 and mother's education, whose results corroborate with other studies in the literature. It is reported that a very significant difference was found between infants who survived after one year of life and infants who died in the neonatal period due to perinatal conditions, whose differences are basically concentrated in the following characteristics: weight, APGAR and pregnancy.

Keywords: Infant mortality, Classification, Machine Learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE GRÁFICOS

LISTA DE ABREVIATURAS E SIGLAS

AP Amapá

CGBP Casas da Gestante, Bebê e Puérpera

DATASUS Departamento de Informática do Sistema Único de Saúde

DF Distrito Federal

DM Mineração de Dados, do inglês *Data Mining*

EHG Eletrohisterograma

ESF Estratégia Saúde da Família

FJP *Fundação João Pinheiro*

IBGE Instituto Brasileiro de Geografia e Estatística

IDH Índice de Desenvolvimento Humano

IDHM Índice de Desenvolvimento Humano Municipal

IPEA *Instituto de pesquisa econômica aplicada*

IREP *Incremental Reduced Error Pruning*

KDD Descoberta do Conhecimento em Base de Dados, do inglês *Knowledge Discovery in Databases*

KNN Algoritmo Vizinhos Mais Próximo, do inglês *K-Nearest Neighbors Algorithm*

MC Malformações Congênitas

MI Mortalidade Infantil

ML Aprendizagem de Máquina, do inglês *Machine Learning*

MLP Perceptron Multicamadas, do inglês *MultiLayer Perceptron*

NB *Naive Bayes*

PBF Programa Bolsa Família

PNUD *Programa das nações unidas para o desenvolvimento*

RAS Rede de Atenção à Saúde

RIPPER *Repeated Incremental Pruning to Produce Error Reduction*

RL Regressão linear

RNA Redes Neurais Artificiais

SC Santa Catarina

SESMG Secretaria de Estado de Saúde de Minas Gerais

SGBD Sistema de Gerenciamento de Banco de Dados

SIM Sistema de Informações sobre Mortalidade

SIMI Sistema de Investigação da Mortalidade Infantil do Paraná

SINASC Sistema de Informações sobre Nascidos Vivos

SISPRENATAL Sistema de Monitoramento e Avaliação do Pré-Natal, Parto, Puerpério e Criança

SMO *Sequential Minimal Optimization*

SUS Sistema Único de Saúde

SVM *Support Vector Machine*

TMI Taxa de Mortalidade Infantil

TMN Taxa de Mortalidade Neonatal

UF Unidades da Federação

VTJ48 *Visually Tuned J48*

WEKA *Waikato Environment for Knowledge Analysis*

SUMÁRIO

1 INTRODUÇÃO

A Mortalidade Infantil (MI) é o termo que designa a morte de crianças no seu primeiro ano de vida sendo uma situação preocupante que atinge todo o globo e especificamente os países mais pobres. A falta de saneamento básico, por exemplo, provoca a contaminação da água e de alimentos e por consequência, aumenta a probabilidade da ocorrência de inúmeras doenças. Segundo Goldani et al. (2001), a Taxa de Mortalidade Infantil (TMI) e a renda são importantes indicadores sociais que estão relacionados entre si indicando que as regiões mais pobres apresentam maior taxa de mortalidade infantil.

A métrica utilizada para definir o número de óbitos se faz por meio da TMI, definida como o número de óbitos de menores de um ano de idade, por mil nascidos vivos, em uma determinada população e ano. Ou seja, é um indicador que estima o risco de uma criança nascida viva morrer antes de completar seu primeiro ano de vida (RIPSA, 2008). Valores altos de TMI refletem, de maneira geral, níveis precários de saúde, condições de vida e desenvolvimento sócio-econômico.

Existe, ainda, uma categorização da taxa de mortalidade infantil quanto à gravidade do problema. A Organização Mundial da Saúde (OMS) estabelece que a TMI é alta quando está acima de 50 mortes para cada mil nascidos vivos; média quanto está entre 20 e 49 mortes para cada mil nascidos vivos; baixa quando menor que 20 mortes para cada mil nascidos vivos (RIPSA, 2008).

A MI pode ser dividida nos períodos neonatal e pós-neonatal, a depender da idade de morte da criança. A mortalidade neonatal refere-se às mortes que ocorrem nas quatro primeiras semanas de vida da criança; divide-se em neonatal precoce (de 0 a 6 dias), e neonatal tardio (de 7 a 27 dias). Já a mortalidade infantil pós-neonatal trata das mortes que ocorrem de 28 dias a 364 dias de vida (RIPSA, 2008).

Essa divisão é importante, uma vez que é possível investigar as causas da MI em diferentes períodos de vida. Por exemplo, as causas da mortalidade neonatal mais frequentes estão relacionadas à gestação, ao parto e aos fatores genéticos; e as respectivas causas do período pós-neonatal são determinadas pelas condições de vida e características familiares (FERRARI; BERTOLOZZI, 2012). Segundo Leal et al. (2018), aproximadamente 50% dos óbitos infantis ocorreram no período neonatal precoce e 70% durante o período neonatal tardio.

A Agência Central de Inteligência dos Estados Unidos elaborou um *ranking*[*] mundial sobre a MI em 2018. Este *ranking* pode ser utilizado para realizar um estudo comparativo sobre a situação do Brasil em relação a outros países. No contexto mundial,

[*]Disponível em https://www.cia.gov/library/publications/resources/the-world-factbook/fields/354rank.html

os países com as maiores TMI estão localizados no sul da Ásia e na África (UNICEF et al., 2018), cujos países não possuem os melhores Índices de Desenvolvimento a nível mundial. Desta forma, os países com maior TMI são o Afeganistão com 108,50 mortes e a Somália com 93, ambas calculadas a cada mil nascimentos. Já os países com menor índice são Eslovênia e Mônaco, com respectivamente 1,60 e 1,80 mortes a cada mil nascimentos. Nesse *ranking*, o Brasil encontra-se na 91ª posição.

A redução da TMI no Brasil foi relatada no artigo de Martins, Pontes e Higa (2018), onde foi constatada uma redução nesta taxa no período de 2000 até 2010 em todos os estados brasileiros. Foi comprovado também que houve melhoria do Índice de Desenvolvimento Humano Municipal (IDHM) nas regiões brasileiras, o qual contempla os seguintes indicadores de três dimensões do desenvolvimento humano: longevidade, educação e renda. A redução na taxa de mortalidade infantil foi verificada também pelo relatório da UNICEF et al. (2018), indicando ainda uma redução na mortalidade neonatal na América Latina e no Brasil de 34% e de 43,5% respectivamente, conforme apresentado na Tabela 1.

Tabela 1 – Mortalidade infantil e neonatal na América Latina e no Brasil

Ano	TMI (Taxa de mortalidade infantil)		TMN (Taxa de mortalidade neonatal)	
	América Latina	Brasil	América Latina	Brasil
1990	44	53	23	25
2017	15	13	10	9

O estudo da taxa de mortalidade infantil pode revelar quais aspectos precisam ser aprimorados nos serviços de atenção à saúde. Além disso, pode apresentar as questões referentes às condições de saúde da população e mostrar uma relação, por exemplo, entre a desigualdade social e a mortalidade infantil (HERNANDEZ et al., 2011), considerando o acesso aos serviços prestados (BRASIL, 2009) e, por isso, trata-se de um fator decisivo para o desenvolvimento de políticas de saúde eficazes na redução da mortalidade infantil de um país. Desta forma, é importante que cada país possa realizar o levantamento adequado das informações referentes à mortalidade infantil para identificar as causas e consequentemente, nortear a adoção de políticas de saúde materna, neonatal e infantil visando a redução da MI (BLACK et al., 2010).

Visando obter informações sobre a MI, o governo Brasileiro implantou o Sistema de Informações sobre Mortalidade (SIM) em 1975 e o Sistema de Informações sobre Nascidos Vivos (SINASC) em 1990. Estes dois sistemas epidemiológicos ganharam relevância por disporem de dados essenciais para o cálculo de indicadores de monitoramento da situação de saúde e de avaliação de ações programáticas.

Para caracterizar a MI, foram escolhidos neste trabalho os estados de Santa Catarina (SC) e do Amapá (AP) devido ao fato de que estes possuem, respectivamente,

a menor e maior TMI do país na média dos anos de 2015 e 2016. Salienta-se que SC está situado ao sul do Brasil, e o AP, ao norte e que o Brasil possui 27 Unidades da Federação (UF) e mais de cinco mil municípios distribuídos em território de dimensões continentais, com desigualdades profundas entre as diferentes UF. Para realizar o respectivo levantamento visando caracterizar a MI em SC e no AP e considerando que os sistemas SIM e SINASC englobam um grande volume de dados, foram adotadas neste trabalho técnicas de mineração de dados, especificamente abordagens de Machine Learning (ML) e foi possível processar toda a base de dados com o objetivo de analisar a mortalidade infantil nos referidos estados. Para a mineração dos dados foram utilizados os seguintes algoritmos: C4.5[†], *Repeated Incremental Pruning to Produce Error Reduction* (RIPPER)[‡], Random Forest, *Support Vector Machine* (SVM) e Redes Neurais Artificiais (RNA). Os algoritmos C4.5 e RIPPER foram importantes na identificação das regras que caracterizam a mortalidade infantil nestes dois estados brasileiros.

1.1 Problema

Este trabalho visa identificar especificamente quais aspectos são determinantes para a ocorrência da Mortalidade Infantil nos estados de Santa Catarina e do Amapá.

A partir do contexto deste trabalho, o problema em questão pode ser formulado a partir das seguintes perguntas:

1) Quais aspectos inerentes aos estados de SC e do AP em 2016 que são determinantes para a caracterização da MI?
2) Quais são as causas básicas de mortalidade infantil no período neonatal nos estados de SC e do AP em 2016?
3) Quais são as causas básicas de mortalidade infantil no período pós-neonatal nos estados de SC e do AP em 2016?

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo deste trabalho é caracterizar por meio de regras de classificação a Mortalidade Infantil a partir das bases SIM e SINASC disponíveis. Para isso, são considerados os estados de SC e do AP.

[†]O algoritmo C4.5 foi implementado em Java e no WEKA foi estabelecida a seguinte nomenclatura: J48

[‡]O algoritmo RIPPER foi implementado em Java e no WEKA foi estabelecida a seguinte nomenclatura: JRIP

Foram utilizados algoritmos (que descobrem classificadores): C4.5, RIPPER, Random Forest, SVM e RNA. A justificativa pela escolha destes algoritmos é pelo fato de que o C4.5 e o RIPPER descrevem as regras de classificação, enquanto o Random Forest categoriza os atributos mais relevantes da base de dados. Foi possível avaliar as regras e os atributos indicados para se identificar os principais fatores que contribuem para a mortalidade infantil nos dois estados considerados. Os algoritmos SVM e RNA, apesar de serem métodos que não explicitam de forma interpretável o conhecimento adquirido pelos modelos, normalmente oferecem bons resultados de predição e foram utilizados a fim de compararmos os resultados destes métodos, considerados métodos caixa preta, com os métodos que descrevem o conhecimento adquirido de uma forma mais explícita.

1.2.2 Objetivos específicos

A fim de atender o objetivo geral, têm-se os seguintes objetivos específicos:

a) Modelar e criar uma base de dados unificada para armazenar os dados originários do SIM e do SINASC.
b) Desenvolver um software com o propósito específico de importar os dados do SIM e do SINASC e exportar para a base de dados que foi criada. Este software será disponibilizado para a comunidade acadêmica.
c) Empregar conceitos de descoberta de conhecimento em bancos de dados, especificamente de aprendizado de máquina na fase de mineração de dados, para caracterizar a mortalidade infantil.
d) Identificar os perfis de mortalidade infantil nos estados de SC e do AP.

1.3 Justificativa

Devido ao fato da MI ser um problema que afeta todas as nações do mundo, principalmente as nações em desenvolvimento, é necessária a adoção de ações governamentais e de políticas de saúde que possam tratar adequadamente os aspectos determinantes para a ocorrência da Mortalidade Infantil. Assim, a motivação para realizar este trabalho é identificar as causas da MI nos estados de Santa Catarina e do Amapá para subsidiar ações governamentais visando a redução da mortalidade infantil.

É importante relatar que uma das limitações encontradas neste trabalho está no fato que o dicionário de dados disponibilizado pelo Departamento de Informática do Sistema Único de Saúde (DATASUS) contempla somente parte dos atributos existentes no sistemas SIM e no SINASC, desta forma, somente 17 (Dezessete) atributos foram analisados. Por outro lado, constata-se a existência de um grande volume de dados (dados de nascimento e morte de bebês, mães e etc.) que podem ser utilizados para caracterizar

a MI. Desta forma, foram utilizados algoritmos de aprendizado de máquina para extrair conhecimento da MI nos estados de SC e do AP com relação ao ano de 2016.

Após a obtenção dos resultados e com a caracterização da Mortalidade Infantil nos estados de SC e do AP, acredita-se que as informações geradas neste trabalho poderão ser utilizadas para nortear as ações, os investimentos e as políticas públicas de saúde nestes estados visando reduzir a respectiva TMI. Inclusive, como proposta de trabalho futuro, pretende-se analisar as políticas de saúde adotadas nos estados de Santa Catarina e do Amapá, considerando as regiões de cada estado, visando identificar os principais problemas relacionados à mortalidade. Assim, espera-se que este trabalho possa contribuir para as melhorias das condições de saúde da população nos dois estados brasileiros avaliados.

1.4 Organização geral do trabalho

Este trabalho está dividido em 6 capítulos. O Capítulo 2 descreve o referencial teórico contemplando mortalidade infantil, políticas públicas de saúde, mineração de dados, descoberta de conhecimento e algoritmos de aprendizagem de máquina. O Capítulo 3 apresenta os trabalhos relacionados. Já o Capítulo 4 aborda as questões relativas aos materiais e métodos, contemplando base de dados, processamentos e métricas de avaliação. O Capítulo 5 traz os resultados e as discussões. Finalmente, a conclusão está inserida no Capítulo 6. Este trabalho contempla um apêndice.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta o referencial teórico conceituando melhor a mortalidade infantil. Além disso, traz as definições de mineração de dados e de descoberta de conhecimento em banco de dados, e dos algoritmos (que descobrem classificadores): *C4.5*, *RIPPER*, *Random Forest*, *SVM* e *RNA*.

2.1 Mortalidade Infantil

O estudo do índice de mortalidade infantil pode revelar detalhes sobre aspectos que precisam ser aprimorados na população e trata-se de um fator decisivo sobre o desenvolvimento do estado. A Mortalidade Infantil apresenta, sob o aspecto científico e social, uma forma de avaliar tanto a questão comunitária, quanto as medidas de saúde adotadas em uma determinada região e trata-se de um evento que aflige o mundo inteiro (BLACK et al., 2010).

No trabalho de Hernandez et al. (2011) foi relatado que a MI possui aspectos associados aos problemas de desigualdade social. Diante deste contexto, a taxa de MI é um parâmetro relevante que poderá revelar as condições de saúde de uma determinada população (VIANNA et al., 2010), e também com o acesso dessa população aos serviços de saúde prestados (BRASIL, 2009).

Visando obter informações sobre MI, o governo brasileiro implantou o Sistema de Informações sobre Mortalidade (SIM) em 1975 e o Sistema de Informações sobre Nascidos Vivos (SINASC) em 1990. O SIM é uma base de dados que inclui todos os registros sobre mortalidade e o SINASC reúne informações sobre os nascidos vivos em todo o território brasileiro.

A MI engloba os óbitos dos seguintes períodos: neonatal precoce (0-6 dias de vida), neonatal tardio (7-27 dias de vida) e pós-neonatal (28 e 364 dias de vida). É importante salientar que o período neonatal engloba os períodos neonatal precoce e tardio. As afecções que ocorrem no período perinatal, que se inicia nas 22 semanas completas de gestação e até os 7 dias completos após o nascimento, poderão eventualmente ser a causa de eventuais óbitos de bebês no período neonatal, principalmente no período precoce.

A partir dos dados da (UNICEF et al., 2018), foi elaborada a Tabela 2, cujo objetivo é comparar a Taxa de Mortalidade Infantil entre Brasil, Alemanha, Estados Unidos, Japão, França, América Latina (e Caribe) nos anos de 1990 e 2017. Nesta comparação, é possível verificar que o Brasil possui uma alta TMI com relação aos países desenvolvidos comparados, apesar de que em 2017 apresentou uma TMI inferior quando comparada com a América Latina e Caribe.

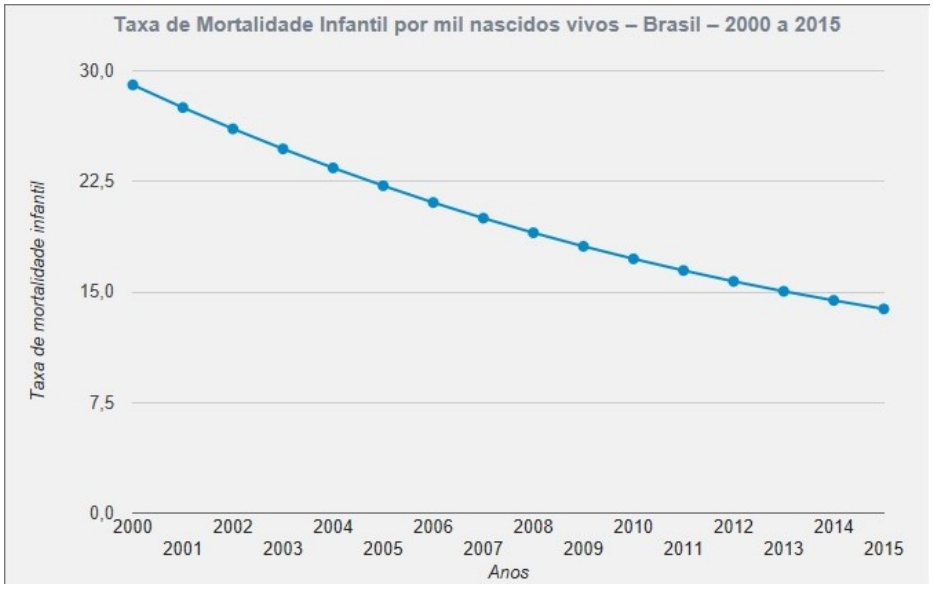
Tabela 2 – Taxa de mortalidade infantil em 1990 e em 2017.

Ano	TMI (Taxa de mortalidade infantil)					
	Brasil	Alemanha	Estados Unidos	Japão	França	América Latina e Caribe
1990	53	7	9	5	7	44
2017	13	3	6	2	4	15

Fonte: UNICEF - 2018.

Estes dados corroboram com o Instituto Brasileiro de Geografia e Estatística (IBGE), que indica que a taxa de MI no Brasil tem apresentado um declínio significativo e tem se mostrado associada com os fatores sociais e econômicos do nosso país (GOLDANI et al., 2001). No Gráfico 1*, observa-se que a taxa de *MI* foi de 29,02/1.000 no ano 2000, para 13,8/1.000 em 2015. Quando comparamos a TMI entre os anos de 1940 a 2017 no Brasil, a redução é de praticamente de 91%, com uma TMI caindo de 146,6 para 13 (IBGE, 2018).

Gráfico 1 – Taxa de mortalidade infantil no Brasil - 2000 a 2015 - IBGE



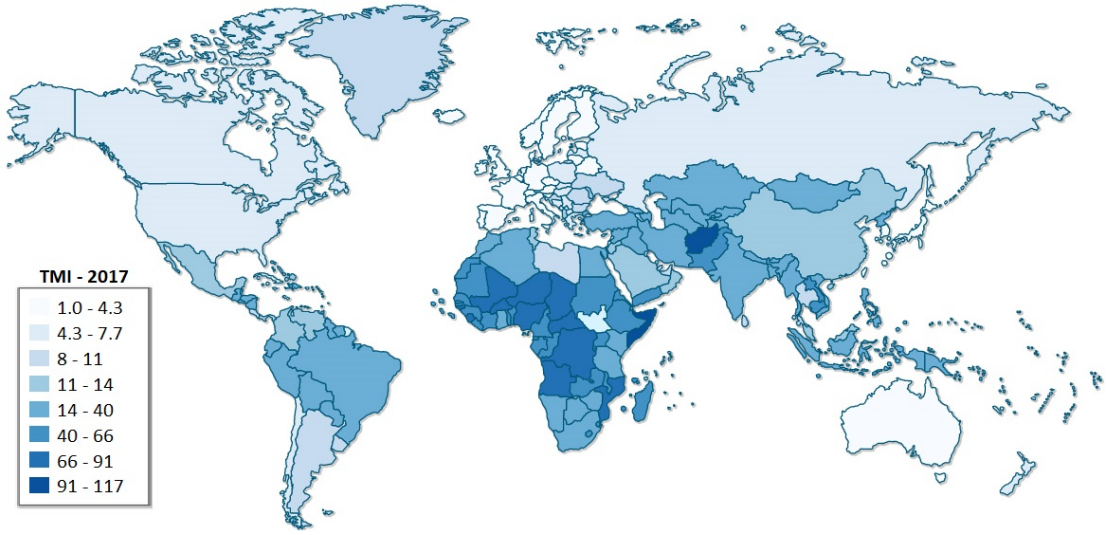
Fonte: IBGE, Projeção da População do Brasil - 2013.

Temos ainda um comparativo da TMI ao redor do mundo, no ano de 2017, veja Figura 1. Quanto mais clara a cor em que o país está colorido, menor é a TMI. É possível observar que, em geral, os países do hemisfério norte possuem taxas baixas de mortalidade infantil, região onde se encontra boa parte dos países desenvolvidos. Por outro lado, observa-se que o continente africano possui as taxas mais altas. Vê-se também que o

*Disponível em https://brasilemsintese.ibge.gov.br/populacao/taxas-de-mortalidade-infantil.html

Brasil tem TMI semelhante ao Marrocos, Argélia e Egito na África e Índia e Cazaquistão na Ásia.

Figura 1 – Comparativo da taxa de mortalidade infantil no mundo em 2017



Fonte: Disponível em https://www.indexmundi.com/map/

Além disso, de acordo com o *Programa das nações unidas para o desenvolvimento* (PNUD), *Instituto de pesquisa econômica aplicada* (IPEA) e *Fundação João Pinheiro* (FJP) (2013), quanto mais desenvolvido o estado/município em que o bebê nasceu/vive, maior será a probabilidade de que a MI esteja relacionada com causas endógenas, inerente ao período neonatal (que engloba o neonatal precoce e o tardio). No caso dos estados/municípios menos desenvolvidos, constata-se que a MI está relacionada, além das causas endógenas, também com as exógenas, por exemplo: respiratórias e doenças infecciosas.

Segundo o IBGE (2018), o governo brasileiro, visando a redução da Taxa de Mortalidade Infantil, tem adotado nos últimos anos as seguintes ações: atenção ao pré-natal, campanhas de vacinação e programas de nutrição infantil, dentre outras. Salienta-se que fatores tais como o aumento da renda, o aumento de domicílios com saneamento adequado e o aumento da escolaridade também contribuíram para a redução da TMI.

De acordo com Oliveira et al. (2012), uma parte das mortes contabilizadas no contexto de mortalidade infantil são causadas por fatores que poderiam ter sido evitados. De uma forma geral, a redução da TMI não é trivial por envolver de uma forma complexa os aspectos sociais e biológicos e também pela qualidade e pela abrangência dos serviços de saúde antes, durante e após o nascimento do bebê.

Assim, visando melhor prever a mortalidade, Apgar et al. (1958) propuseram uma forma de avaliar as condições do bebê recém-nascido com o objetivo de adotar as medidas necessárias para tentar evitar um possível óbito que ficou popularmente conhecido como teste de Apgar, cuja avaliação é realizada no primeiro e no quinto minuto de vida conhecido popularmente como APGAR1 e APGAR5.

O teste de Apgar consiste em avaliar cada um dos seguintes 5 (cinco) aspectos do bebê em uma escala de valores de 0 a 2:

1) *Cor da pele*:
   - 0: Roxo/escuro
   - 1: Com coloração azulada/arroxeada nas extremidades
   - 2: Sem coloração azulada/arroxeada
2) *Tônus muscular*:
   - 0: Flácido
   - 1: Flexão braços/pernas
   - 2: Muitas flexões
3) *Esforço respiratório*:
   - 0: Ausente
   - 1: Irregular
   - 2: Forte
4) *Irritabilidade reflexa*:
   - 0: Sem resposta a estímulo
   - 1: Algum movimento
   - 2: Choro vigoroso, tosse ou espirro
5) *Frequência cardíaca*:
   - 0: Sem pulso
   - 1: Menor que 100/minuto
   - 2: Maior que 100/minuto

Assim, a nota total do teste de Apgar consiste no somatório das notas de cada um dos cinco aspectos descritos. Um bebê que obtiver uma nota total entre 7 e 10 indica que está possivelmente saudável. Já um bebê que obtiver uma nota total entre 4 a 6 indica um quadro de saúde moderado. Nos casos dos bebês que tiverem notas totais entre 0 e 3 há uma gravidade na saúde do respectivo bebê (DUIM; NAMPO; SOUZA, 2017).

2.2 Políticas públicas de saúde

Políticas públicas são definidas como um conjunto de intervenções que resultam em programas e projetos de atuação governamental. A saúde é um exemplo e o cuidado com o recém nascido e a criança é uma das prioridades desta atuação. Diversos trabalhos discutem políticas públicas de saúde no Brasil.

O artigo de Moreira et al. (2012) cita as seguintes políticas públicas de saúde: o *Programa de Redução da Mortalidade Infantil e Materna em Minas Gerais (Viva Vida)* e o Projeto Rede Cegonha. O Programa Viva Vida foi criado como resultado do alto índice de óbitos de mãe e de crianças que poderiam ter sido evitados. O programa foi lançado em outubro de 2003 e consiste na sistematização de ações, cuja estratégia fundamental é a parceria entre governo e sociedade civil visando a redução da mortalidade materna e infantil. Em 2013, a Secretaria de Estado de Saúde de Minas Gerais (SESMG) voltou a dar continuidade ao processo de adoção de comitês de Prevenção de Mortalidade Materna, Infantil e Fetal, cujo objetivo é reduzir a mortalidade materna e infantil por meio de ações de intervenção.

Com relação ao Projeto Rede Cegonha, o mesmo foi lançado pelo governo federal em 2011 e tem a participação das três esferas: federal, estadual e municipal, cujo financiamento é compartilhado. O Projeto Rede Cegonha tem como objetivo reduzir a mortalidade infantil e, para alcançar este objetivo, busca oferecer às mulheres saúde e qualidade de vida no decorrer da gestação, do parto, do pós-parto, inclusive buscando garantir o desenvolvimento da criança até completar dois anos de vida. Assim, os quatro componentes de atuação desse projeto são:

- Pré-natal;
- Parto e nascimento;
- Puerpério e atenção integral à saúde da criança e sistema logístico;
- Sistema logístico (transporte sanitário e regulação);

Segundo informações do Ministério da Saúde[†], no atual cenário, o Projeto Rede Cegonha abrange 5.488 municípios, atendendo 2,6 milhões de gestantes. Desde 2011, os investimentos realizados ultrapassam R$ 3,1 bilhões de reais para a realização das ações desse projeto. Com relação ao ano de 2013, por exemplo, alcançou-se o número de 18,9 milhões de consultas pré-natais que foram realizadas pelo Sistema Único de Saúde (SUS). Assim, as ações desse projeto contribuíram para a redução da mortalidade materna e infantil. Um outro aspecto importante é que o Ministério da Saúde também incentiva os municípios a adotarem Casas da Gestante, Bebê e Puérpera (CGBP). As CGBP são locais que visam receber mulheres e bebês que necessitam de atenção especial e que estão

[†]Disponível em http://www.saude.gov.br/acoes-e-programas/rede-cegonha

perto do hospital, mas sem precisar de estarem internadas. Uma outra ação importante foi a ampliação e qualificação de leitos para gestantes de alto risco, UTI e UCI neonatal visando assegurar a saúde das mulheres e dos bebês no momento do parto.

O estado de Santa Catarina iniciou a implantação da Rede Cegonha em 2011 e passou a ter 16 planos de ação em 2013, sendo no Brasil a primeira Rede Cegonha com abrangência a nível estadual e a primeira Rede de Atenção à Saúde (RAS) com abrangência em todo estado. Implantou inclusive o Sistema de Monitoramento e Avaliação do Pré-Natal, Parto, Puerpério e Criança (SISPRENATAL), que faz parte do Projeto Rede Cegonha, cujo sistema possibilita cadastrar as gestantes visando monitorar e avaliar os atendimentos oferecidos pela rede pública de saúde a cada gestante e recém-nascido. Desta forma, este sistema visa assegurar a saúde da mãe e do bebê e visa ajudar na identificação das importantes causas da mortalidade e principais aspectos associados à gravidez de risco. Com relação ao estado do Amapá, a Rede Cegonha faz parte da política pública de saúde do referido estado e é uma das subdivisões da RAS.

Apesar de não ser uma política pública voltada somente para a área da saúde, Silva e Paes (2019) descrevem o Programa Bolsa Família (PBF), lançado pelo governo federal, como sendo uma das políticas públicas que visa combater a pobreza e a desigualdade social. O PBF consiste em um programa de transferência de renda desde que as famílias contempladas atinjam certas condicionantes a partir destes três eixos:

- Complemento da renda: as famílias atendidas por este programa recebem mensalmente um benefício;
- Acesso a direitos: para receber o benefício, as famílias precisam atender algumas condições;
- Articulação com outras ações: o PBF visa atuar de uma forma integrada com outras ações;

Em linhas gerais, o PBF engloba as seguintes classes de família:

- Famílias extremamente pobres: possui renda mensal de até R$89,00 por pessoa;
- Famílias consideradas pobres: possui renda mensal entre R$89,01 e R$178,00 por pessoa, mas tem que ser composta por: gestantes e crianças ou adolescentes entre 0 e 17 anos.;

Segundo Silva e Paes (2019), foram analisados os dados de 1.133 municípios no semiárido no período de 2004 a 2010 e constatou-se que o PBF e a Estratégia Saúde da Família (ESF) desempenharam um papel relevante no que se refere à redução da mortalidade infantil e do analfabetismo e ao aumento de consultas pré-natal, além de outros fatores. Inclusive, existem estudos que relataram que o PBF propiciou uma redução da pobreza e da desigualdade de renda, favorecendo também a redução do número de mortes de crianças. É importante salientar que mesmo com a redução da taxa de

mortalidade infantil nos municípios do Semiárido, esta ainda é considerada alta quando comparada com países desenvolvidos.

Por outro lado, os autores (SILVA; PAES, 2019) relatam que mantendo a cobertura do PBF e também da ESF será possível reduzir ainda mais a taxa de mortalidade infantil, podendo alcançar futuramente os níveis de países desenvolvidos. Salienta-se que as melhorias na área de educação e a diminuição dos níveis de analfabetismo também são fatores que tendem a reduzir a pobreza e a desigualdade de renda, mas é necessário que seja concedida uma atenção ainda maior à área da saúde visando principalmente que a infraestrutura pública de saúde possa ser de fato melhorada continuamente.

2.3 Mineração de dados e a descoberta de conhecimento em bancos de dados

A mineração de dados pode ser descrita como a atividade de extrair conhecimento e/ou padrões a partir de uma grande quantidade de dados (QUILICI-GONZALEZ; ZAMPIROLLI, 2014). A maior parte da literatura trata a mineração de dados como Descoberta do Conhecimento em Base de Dados, do inglês *Knowledge Discovery in Databases* (KDD). No entanto, alguns autores consideram a mineração como uma etapa do processo KDD (FACELI et al., 2011), que pode ser assim dividido (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996):

1) *Seleção*: na primeira fase do processo são escolhidos os dados contendo todas as possíveis variáveis e registros que farão parte da análise, ou seja, os mais importantes para se alcançar o objetivo pretendido;
2) *Limpeza e integração de dados*: os dados são filtrados e os que estiverem errôneos, redundantes, ausentes ou inconsistentes são retirados;
3) *Transformação dos dados*: os dados são formatados e armazenados para que o software de mineração consiga processá-los;
4) *Mineração de dados*: é a etapa mais importante do processo de KDD. Nela são utilizados os algoritmos de descoberta de padrões. As técnicas de data mining podem ser aplicadas a tarefas como associação, classificação, regressão e agrupamento; e
5) *Avaliação e apresentação dos resultados*: nessa fase as informações são direcionadas para a realização de análise dos dados com o objetivo de verificar se possuem validade para o problema proposto.

2.4 Aprendizagem de Máquina

O termo Aprendizagem de Máquina, do inglês *Machine Learning* (ML), consiste basicamente na adoção de algoritmos visando obter algum tipo de informação que seja útil e não trivial. Assim, a ML pode ser utilizada no processo do KDD, especificamente

na etapa de mineração de dados.

2.4.1 Algoritmos

Como já salientado, foram utilizados algoritmos (que descobrem classificadores), os quais representam o conhecimento por meio de regras (RIPPER), regras e árvores (*C4.5*) e que categoriza atributos mais relevantes (Random Forest). Além disso, foram também utilizados dois métodos de classificação reconhecidamente eficientes: o SVM e a RNA. As subseções seguintes descrevem resumidamente cada um destes algoritmos.

2.4.1.1 C4.5

O algoritmo C4.5 é uma implementação em Java que visa a geração de árvores de decisão permitindo o tratamento de atributos numéricos e/ou nominais (QUINLAN, 1993) e no *Waikato Environment for Knowledge Analysis* (WEKA) recebe a nomenclatura: J48. O objetivo principal é selecionar para cada nó um atributo que melhor subdivide o conjunto das amostras. Consideram-se os atributos mais significativos aqueles que mais agregam na classificação.

Para a geração da árvore de decisão, utiliza-se a estratégia de dividir para conquistar. Desta forma, divide o problema em subproblemas menores e os trata de forma recursiva, assemelhando-se com a estratégia gulosa. Assim, a seleção dos atributos é realizada a partir da maximização do critério de divisão e, desta forma, os principais atributos selecionados são alocados nos primeiros nós visando proporcionar uma melhor subdivisão da árvore gerada.

Esse algoritmo utiliza a razão de ganho para escolher o atributo que particiona o conjunto de dados em cada iteração. O método utiliza também o conceito de entropia (Equação 2.1), que mede o grau de pureza de um conjunto. Sendo assim, ao construir a árvore de decisão o objetivo é reduzir a entropia, ou seja, reduzir a dificuldade para predizer a variável alvo (FACELI et al., 2011).

$$E(A) = -\sum_{i} P_i * log_2 p_i \tag{2.1}$$

onde A é uma variável aleatória e p_i é a probabilidade para cada valor em A.

2.4.1.2 RIPPER

O algoritmo RIPPER foi implementado na ferramenta WEKA na qual utiliza a nomenclatura: JRip. O RIPPER, cujo algoritmo foi descrito em (COHEN, 1995), refere-se

à versão otimizada do algoritmo *Incremental Reduced Error Pruning* (IREP) e adota uma abordagem de executar várias vezes o IREP até obter um conjunto de regras com baixa taxa de erro de classificação.

O algoritmo inicia com um conjunto vazio de regras e contempla a fase de crescimento e de poda. A fase de crescimento do conjunto de regras visa otimizar a precisão, realizando combinações buscando a condição de maior ganho, cuja abordagem de parada consiste em encontrar a representação completa das regras ou quando a incorporação de novas regras não propiciar um aumento da taxa de acerto da regra produzida (LIBRALON, 2007). Já a fase de poda evita a elaboração de regras específicas e também realiza a poda das regras que não agregam no aumento da taxa de acerto. Assim, todo o conjunto de dados e a poda ocorrem quando o conjunto crescente apresenta uma quantidade de regras excessivas e assim é selecionada uma poda que acarrete uma redução de erro (LIBRALON, 2007).

2.4.1.3 Random Forest

O algoritmo Random Forest adota a abordagem *bagging* (BREIMAN, 1996) para a tomada de decisão final, a qual visa reduzir a variância do conjunto de dados; o que normalmente afeta algoritmos de árvore, como o C4.5, por exemplo (FRIEDMAN; HASTIE; TIBSHIRANI, 2009).

Este algoritmo, que resolve problemas de regressão e de classificação, consiste em um grande número de árvores de decisão, as quais são caracterizadas por terem seus atributos e instâncias definidos de forma aleatória (BREIMAN, 2001). Assim, cada árvore de decisão classifica cada conjunto de dados e a classificação final dar-se-á pela maioria dos votos de classificação, utilizando-se o conceito de voto majoritário (MARINS, 2016).

De maneira geral, o modelo tende a apresentar uma maior precisão quanto maior for o número de árvores de decisão. No entanto, é importante considerar que o aumento do número de árvores poderá chegar a um ponto que acréscimos de novas árvores poderão em determinado ponto não agregar positivamente nos resultados (MARINS, 2016). Segundo Khoshgoftaar, Golawala e Hulse (2007), este algoritmo apresenta resultado satisfatório mesmo com a presença de ruído/outliers.

2.4.1.4 Suport Vector Machine

O SVM é um algoritmo de aprendizado de máquina proposto por Boser, Guyon e Vapnik (1992) e por Cortes e Vapnik (1995). O objetivo deste algoritmo é resolver problemas de classificação e de regressão, tanto lineares quanto não-lineares.

O *Suport Vector Machine* é caracterizado por possuir duas fases: treinamento e

classificação. Com relação à fase de treinamento, as instâncias das classes estão dispostas em um plano linear e a separação das instâncias, de acordo com a classe, ocorre a partir do hiperplano. O hiperplano pode ser definido como sendo a separação entre classes, sendo ideal encontrar um hiperplano cuja distância entre as classes no plano seja o maior possível. Para realizar o treinamento do SVM, Platt (1998) propôs o algoritmo *Sequential Minimal Optimization* (SMO) que foi implementado em Java.

Assim, o SVM busca encontrar o hiperplano ótimo, apresentando a maior distância que separa as classes. Já na fase de classificação, as instâncias restantes são classificadas a partir do hiperplano obtido na fase de treinamento. Em problemas de classificação não-linear disposto em um plano de alta dimensão, é necessária uma função chamada *kernel*. De acordo com Costa (2016), o algoritmo SVM pode utilizar funções *kernel*, como por exemplo: 1) Linear; 2) Polinomial; 3) Base Radial (RBF); 4) Base Radial Exponencial; 5) Tangente Hiperbólica, dentre outras.

Segundo Bonesso (2013), é importante a parametrização adequada do *kernel* do SVM visando obter resultados satisfatórios. Esta parametrização pode ser realizada da forma manual ou automatizada. No entanto, a definição manual da parametrização pode exigir um conhecimento prévio para que se possa obter um resultado adequado. Com relação à definição automatizada, pode-se adotar, dentre outras coisas, a abordagem Busca em Grade, que consiste em um algoritmo que tem como objetivo encontrar uma melhor parametrização visando obter resultados mais significativos (BONESSO, 2013).

As máquinas de SVM são tolerantes a ruídos e, na maioria das vezes, são superiores quando se compara a outros algoritmos.

2.4.1.5 Rede Neural Artificial

A RNA foi inspirada no funcionamento do cérebro humano e consiste no aprendizado a partir de exemplos e na característica de generalizar o conhecimento adquirido no aprendizado. Cada neurônio biológico possui: corpo celular, dendritos e axônio, salientando que cada parte tem funções específicas. Basicamente, o corpo celular tem a função de processar a informação, os dendritos tem a função de recebê-las e o axônio de enviá-las a outros neurônios. Assim, a informação é enviada a partir do ponto de encontro entre o axônio e o dendrito e é chamada de sinapse.

Uma rede neural do tipo Multi-Layer Perceptron (MLP ou Perceptron Multi camadas) é composta por diversos neurônios artificiais, representando a camada de entrada, uma ou mais camadas intermediárias e uma camada de saída. Estes neurônios são interconectados e passam por um processo de aprendizado. Um algoritmo importante de aprendizado supervisionado do tipo MLP é o Backpropagation (HAYKIN, 2001).

Da mesma forma que o SVM, a RNA necessita de ajustes de hiperparâmetros para o seu perfeito funcionamento, além de ser considerado um método caixa preta, ou seja, apresenta dificuldade em explicar o conhecimento adquirido.

2.4.2 Características dos algoritmos

Na maioria dos casos, tende a ser importante que sejam utilizados algoritmos de aprendizado de máquina que possibilitem a interpretação de como previsões são realizadas. Dentro deste contexto, pode-se dividir os algoritmos de ML em duas classes: caixa branca e caixa preta.

Os algoritmos classificados como caixa branca são normalmente mais simples e eles são mais fáceis de interpretar. Neste trabalho, por exemplo, foram utilizados os seguintes algoritmos caixa branca: C4.5 e RIPPER. Já os algoritmos classificados como caixa preta normalmente oferecem uma maior precisão, mas são mais difíceis de serem interpretados. Neste trabalho, por exemplo, foram utilizados os seguintes algoritmos caixa preta: SVM e a RNA.

3 TRABALHOS RELACIONADOS

Todos os trabalhos descritos nesta seção estão envolvidos com a aplicação de técnicas de KDD na área da saúde, incluindo o estudo sobre a MI e a descoberta de padrões que possam resultar em alguma medida de intervenção para diminuir a TMI.

Para Oliveira (2001), os sistemas SIM e SINASC são importantes fontes de informação no contexto de saúde, especialmente com relação à mortalidade infantil, e esse trabalho relata que a mineração de dados é aplicável para descobrir conhecimento a partir de conjunto de dados relacionados à área da saúde. Esse trabalho utilizou técnicas estatísticas e técnicas de classificação associadas ao processo de KDD para traçar o perfil de recém-nascidos e identificar as variáveis associadas à mortalidade infantil. Os resultados estatísticos apresentam uma forte correlação ao peso do bebê, ao nível de APGAR (exame que avalia o nível de adaptação do bebê à vida fora do útero) do primeiro e quinto minuto de vida, à duração da gestação (em semanas) e ao tipo de gravidez (única, dupla, etc.). Esse trabalho limitou-se a analisar os dados do município de Florianópolis localizado no estado de Santa Catarina e somente considerou o ano de 1996.

No trabalho de Kitsantas, Hollander e Li (2006) foram identificados subgrupos de mulheres com alto risco de desenvolver uma gestação em que o bebê nasça com baixo peso. Esse trabalho limitou-se a utilizar os dados de sete regiões da Flórida nos Estados Unidos e somente considerou o ano de 1998, onde, aplicando técnicas de mineração de dados, foi possível identificar vários subgrupos de alto risco. Como exemplo, os autores citam o seguinte grupo de risco: mãe brancas, hispânicas ou outras mães não brancas que eram saudáveis e fumavam, mas com um ganho de peso da mãe inferior a 9 kg aproximadamente. Esse grupo tem um risco maior de dar à luz bebês com baixo peso em comparação com aquelas com as mesmas características, mas com um ganho de peso da mãe superior a 9 kg aproximadamente. Além das características já citadas, fatores como quantidade de parto anterior e estado civil foram preditores importantes para os resultados da gravidez entre as não brancas, hispânicas ou outras mães não brancas. Ficou demonstrado que árvores de classificação podem ser usadas para identificar subgrupos de risco de ter baixo peso ao nascer.

Vianna et al. (2010) relata que a taxa de mortalidade infantil é um indicador relevante para mensurar, por exemplo, as condição de saúde de uma determinada população. Devido à importância do tema, o estudo visa, dentre outras coisas, realizar a caracterização da mortalidade infantil no Paraná. Desta forma, esse trabalho limitou-se a englobar as bases de dados do SINASC, do SIM e do Sistema de Investigação da Mortalidade Infantil do Paraná (SIMI)* no período de 2000 a 2004 com relação ao estado

*O SIMI foi descontinuado e não está mais ativo.

do Paraná e somente considerou o período de 2000 a 2004, salientando que o SIMI é um sistema que foi desenvolvido em 2000 no estado do Paraná. Neste sentido, foi utilizada a técnica de KDD, e uma das etapas do processo é a Mineração de Dados, do inglês *Data Mining* (DM), que usou as referidas bases de dados do SINASC, do SIM e do SIMI; também foi escolhido o J48 disponível na ferramenta WEKA, visando a descoberta de padrões e geração de regras. Foram selecionadas 4.230 regras, por exemplo: mãe adolescente e peso ao nascer menor que 2.500 gramas parto pós-termo e mãe adolescente com outro filho, ou com afecções maternas, que aumentam o risco para óbito neonatal. Diante deste contexto, foi verificado que é importante que ocorram, dentre outras, ações especialmente voltadas para mães adolescentes, principalmente as que já têm outro filho, mãe com problemas na gestação, mães com filhos que possuem baixo peso ao nascer e com pós-datismo. O pós-datismo é um termo que caracteriza uma gestação prolongada, cujo tempo da gestação está entre 40 e 42 semanas.

As Malformações Congênitas (MC) são anomalias que podem surgir no feto durante a gestação e podem comprometer a vida do bebê após o nascimento. Choudhry, Qamar e Chaudhry (2015) associam as MC como um dos fatores que contribuem para a mortalidade infantil. Diante dessa questão, o objetivo desses autores foi encontrar padrões de ocorrência das malformações congênitas limitando a analisar os dados da população de duas cidades no Paquistão: Rawalpindi e Islamabad, visando nortear ações e políticas de saúde no referido país. Para isto, foi realizada a mineração de dados de uma grande base de informações de mães grávidas utilizando a técnica de aprendizagem de máquina não supervisionada. Os resultados desse estudos constataram regras como as seguintes, por exemplo: 1) mulheres com peso maior que 65 kg e hemoglobina menor que 10 têm 92% de possibilidades de ter um filho com MC; 2) mulheres com peso superior a 65 kg e níveis de açúcar fora dos padrões normais terão uma probabilidade maior de ter um bebê com MC.

No trabalho de Chen, Oster e Williams (2016) foi avaliada a taxa de mortalidade infantil nos EUA e comparada com quatro países europeus: Áustria, Finlândia, Reino Unido e Bélgica, já que a mortalidade nos EUA é considerada a mais alta se comparada a outros países desenvolvidos (LORENZ et al., 2016). Os autores identificaram que os EUA têm uma taxa de mortalidade no período neonatal (0 a 28 dias) semelhante aos países avaliados, mas tem uma taxa maior de mortalidade no período pós-neonatal (29 dias a 1 ano). Os autores associaram a mortalidade com um menor nível socioeconômico das pessoas, e sugerem que sejam tomadas iniciativas, como acontece na Finlândia e Áustria, em que há políticas públicas que levam enfermeiros ou outros profissionais de saúde para visitar os pais e as crianças em casa. Essas visitas combinam exames do bebê com aconselhamento e apoio do cuidador e poderiam colaborar para a redução da mortalidade no período crítico, pós-neonatal, identificado. Ainda nos EUA, identificaram

que a mortalidade entre os negros é duas vezes maior que entre os brancos, e além disso, há uma mortalidade maior entre bebês prematuros. Nesse trabalho, menciona-se uma política de saúde que consiste na visita de enfermeiras às casas dos pais de recém-nascidos visando realizar exames nos bebês e além de aconselhamentos que podem contribuir eventualmente para a redução da mortalidade infantil.

O estudo conduzido por Ramos et al. (2017) apresenta um sistema na área de saúde cujo objetivo é aperfeiçoar a atenção à saúde de gestantes e recém-nascidos. Assim, foi desenvolvido um framework, chamado GISSA, que possui um conjunto de serviços de saúde, além de um sistema de alerta para que os gestores possam adotar as medidas necessárias para cada recém-nascido visando reduzir a mortalidade infantil. Esse sistema utiliza técnicas de DM para gerar alertas quando há risco da ocorrência de óbito de recém-nascidos a partir das bases de dados dos sistemas SINASC e SIM. Os melhores resultados foram obtidos com o algoritmo *Naive Bayes* (NB). O primeiro estudo de caso desse trabalho limitou-se a contemplar somente a região nordeste do Brasil, já que é uma região que possui estados brasileiros mais vulneráveis com relação à mortalidade infantil.

O artigo Sartorelli et al. (2017) adotou técnicas de KDD para a análise dos resultados e foram utilizados os dados do SINASC e do SIM, limitando-se a caracterizar a mortalidade infantil de um município do estado do Paraná, compreendendo o período de 2010 a 2014. A mineração de dados foi realizada por meio do J48 que gera um conjunto de regras. Na fase de análise dos resultados, utilizou-se o algoritmo proposto por Teixeira, Colmanetti e Carvalho (2015) para a identificação de um conjunto de variáveis que sejam vigorosamente relacionadas. Assim, os resultados indicaram que o baixo peso ao nascer, idade gestacional e presença de anomalias são características referentes à mortalidade infantil neste município analisado. Foi constatado também que a escolaridade da mãe, por si só, não se apresentou relevante; talvez possa ter um papel significativo se não for analisada isoladamente. Constatou-se ainda também que os antecedentes obstétricos referentes à quantidade de nascidos mortos é uma característica significativa na ocorrência de prematuridade nas subsequentes gestações de uma mãe. Uma das reveladoras constatações está no fato que de 56,8% dos óbitos eram evitáveis, sendo uma grande parte reduzíveis quando prestada uma atenção devida à mulher na gestação, salientando que uma parcela significativa das respectivas mulheres havia realizado sete ou mais consultas no pré-natal. Desta forma, os resultados do estudo são muito importantes para que possam ser adotadas políticas de saúde eficazes, indicando a necessidade de enfatizar a importância de focar adequadamente na questão da saúde da mãe, do pré-natal, do parto, do puerpério e dos nascimentos prematuros.

Visto que o nascimento prematuro é uma das causas significativas que impactam na taxa de mortalidade infantil, Despotovic et al. (2018) adotam a abordagem de aprendizado máquina visando predizer a ocorrência de nascimento prematuro, entre 22 e 25 semanas da

gestação, a partir dos dados Eletrohisterograma (EHG) utilizando os seguintes algoritmos: Algoritmo Vizinhos Mais Próximo, do inglês *K-Nearest Neighbors Algorithm* (KNN), Random Forest e SVM. O EHG consiste no registro da atividade elétrica a partir da contração do útero. Segundo os resultados apresentados, o algoritmo *Random Forest* obteve os melhores resultados, com uma precisão de 99,23% e uma sensibilidade de 98,40%. Os resultados também apontaram que o monitoramento durante a gravidez poderá auxiliar na identificação de gestações com risco de parto prematuro. Esse trabalho limitou-se a utilizar os dados sobre EHG disponíveis na Physionet[†].

O artigo de Toscano e Hossain (2018) aborda a relação entre a taxa de mortalidade infantil e os seguintes principais fatores: renda per capita, nível de escolaridade do pai e da mãe, indicadores dos serviços de saúde. Esse estudo utiliza regressão polinomial visando analisar a relação entre o nível de renda per capita e o nível da escolaridade do pai e da mãe, considerando 136 países, de acordo com a disponibilidade de dados. Os dados do Catar não foram considerados neste trabalho devido ao fato de que a renda per capita apresenta uma alta discrepância quando comparada com os demais países analisados. Por exemplo, para cada ano de escolaridade da mãe, a TMI pode reduzir entre 11,937% e 2,588% e para cada ano de escolaridade do pai, a TMI pode reduzir entre 9,789% e 3,273%. Diante deste contexto, constatou-se que o aumento da escolaridade do pai e da mãe tende diminuir a taxa de mortalidade infantil, pois um pai e/ou mãe com nível educacional alto tende a ser mais cuidadoso com a saúde dos seus filhos.

[†]Para maiores informações sobre a Physionet, consulte o artigo Goldberger et al. (2000)

4 MATERIAIS E MÉTODOS

Para a caracterização da mortalidade infantil nos estados de SC e do AP, foram seguidas as seguintes etapas: 1) Modelagem e criação da base de dados; 2) Desenvolvimento de um software para a importação dos dados; 3) Pré-processamento da base de dados; 4) Utilização de algoritmos (que descobrem classificadores): C4.5 com o plugin VTJ48*, RIPPER, Random Forest, SVM e RNA, e 5) análise dos resultados de cada algoritmo e comparação de resultados.

4.1 Descrição da base de dados

A base de dados sobre mortalidade infantil foi obtida no site do DATASUS† considerando o período de 2006 a 2016 de todos os estados e também do Distrito Federal (DF) a partir dos seguintes sistemas:

A) *SINASC* (Nascidos vivos)
B) *SIM Óbito Fetal* (Óbitos fetais)
C) *SIM Óbito Infantis* (Óbitos infantis)

Uma vez que a mortalidade infantil engloba o número total de óbitos nos períodos neonatal precoce (0-6 dias de vida), neonatal tardio (7-27 dias de vida) e pós-neonatal (28 e 364 dias de vida), os dados do SIM óbito fetal poderão ser utilizados em trabalhos futuros, mas não foram utilizados neste trabalho, pois o foco é analisar o perfil da mortalidade infantil.

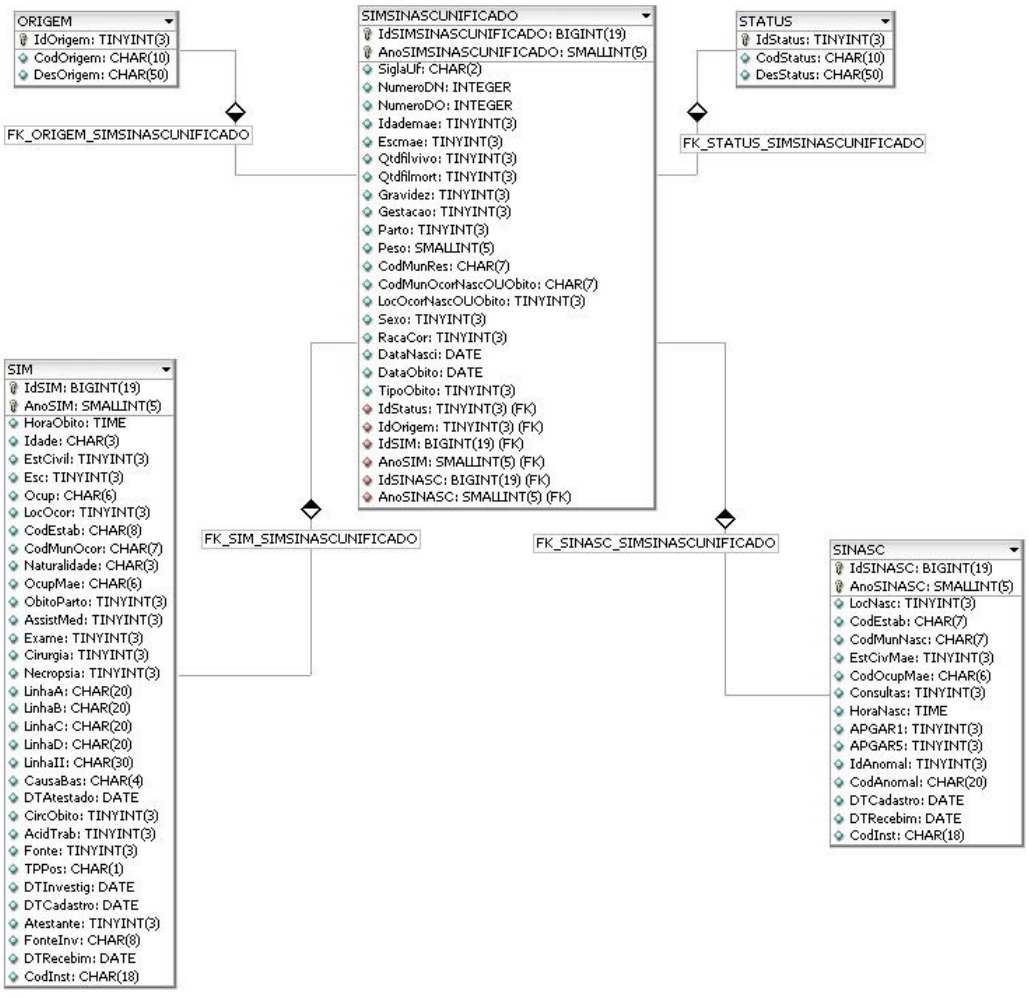
O DATASUS disponibiliza os dados no formato DBC que precisa ser convertido para o formato DBF‡. Após a conversão, o tamanho total de todos os arquivos do SINASC e do SIM Óbito Infantil é de 6GBs (Seis Gigabytes), aproximadamente. Nesta primeira etapa, foram realizados *downloads* de cada arquivo de cada estado para cada ano, contemplando especialmente os nascidos vivos e os óbitos infantis que ocorreram dentro do período de um ano após o nascimento. Por se tratar de uma pesquisa que envolve base de dados de uma forma pública, não envolvendo direta ou indiretamente seres humanos, a pesquisa não foi apreciada por um comitê de ética em pesquisa.

Na segunda etapa, foi modelado um banco de dados conforme apresentado no Quadro 1. Após a conclusão da modelagem do banco de dados, foram criadas as tabelas no Sistema de Gerenciamento de Banco de Dados (SGBD) Microsoft SQL Server 2017. Desta forma, a tabela SIMSINASCUNIFICADO armazena de uma forma integrada os

*Plugin disponível em: http://www.ri.fzv.um.si/vtj48/.
†Disponível em: http://www.datasus.gov.br/DATASUS/index.php
‡Foi utilizada a ferramenta TabWin disponível em: http://www.datasus.gov.br/DATASUS/index.php

dados comuns ao SIM e ao SINASC e esta integração foi possível a partir do cruzamento das seguintes informações: número da declaração de nascidos vivos, estado aonde nasceu e a data de nascimento do bebê. Também existem as tabelas para armazenar os dados não comuns dos respectivos sistemas, conforme Quadro 1. A tabela STATUS indica se o registro refere-se ao nascimento (VIVO) ou mortalidade (MORTO) e a tabela ORIGEM relata de qual arquivo é originado o registro: SINASC e SIM óbito infantil. O objetivo de ter modelado este banco de dados foi unificar todos os dados originários do DATASUS em uma única base de dados, visando propiciar melhores condições para realizar a Mineração dos Dados com relação ao período anteriormente informado.

Quadro 1 – Modelagem da base de dados



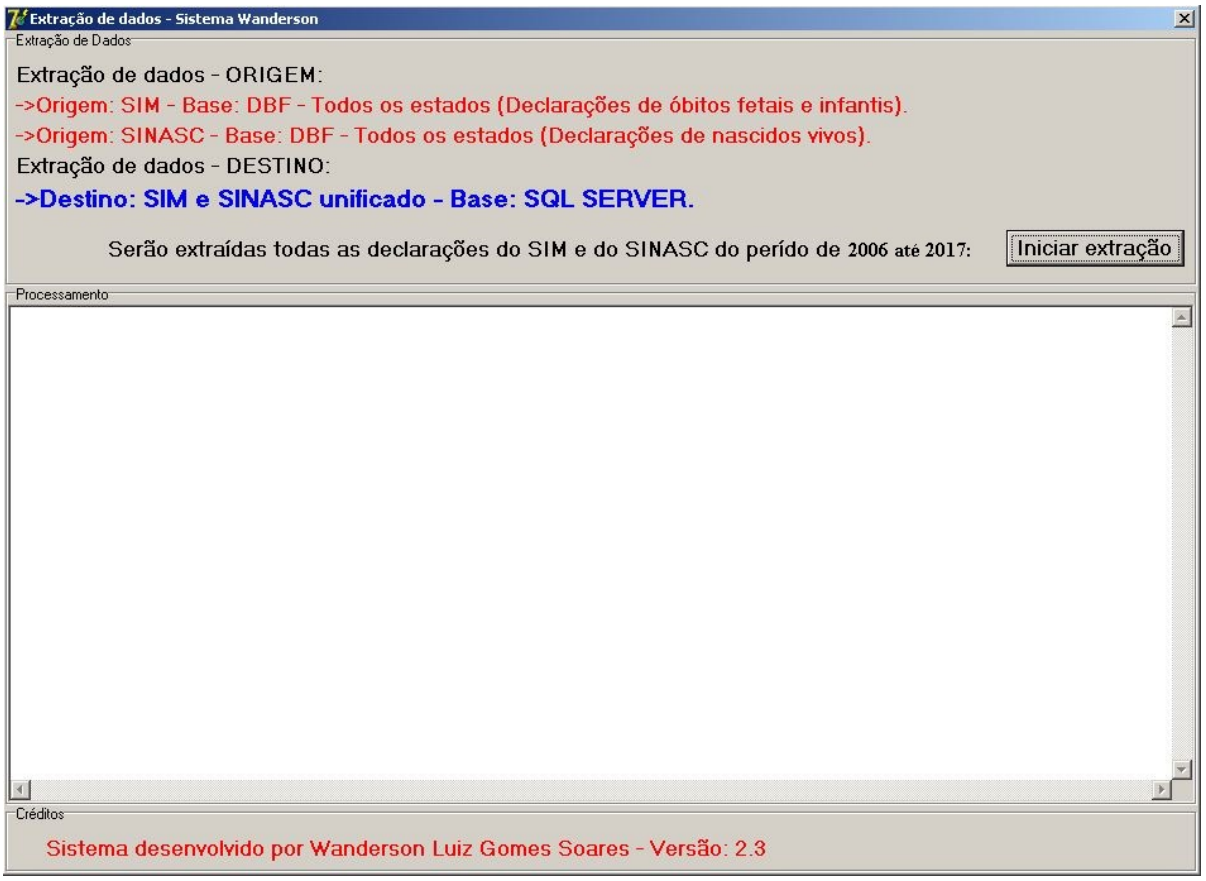
Fonte: Elaborado pelo autor.

Na terceira etapa, foi desenvolvido um software[§] cuja interface gráfica está apresentada na Figura 2. O objetivo é extrair os dados originários de cada arquivo DBF do DATASUS e importar para o banco de dados que foi apresentado anteriormente no

[§]Este software foi desenvolvido na linguagem Delphi/Pascal.

Quadro 1. Desta forma, depois de executar o programa, todos os dados originários do DATASUS que estavam separados em cada arquivo por estado e por ano (referentes ao período de 2006 até 2017) do SIM e do SINASC no formato em DBF foram importados para o banco de dados que foi modelado.

Figura 2 – Interface gráfica do software para extração dos dados



Fonte: Elaborado pelo autor.

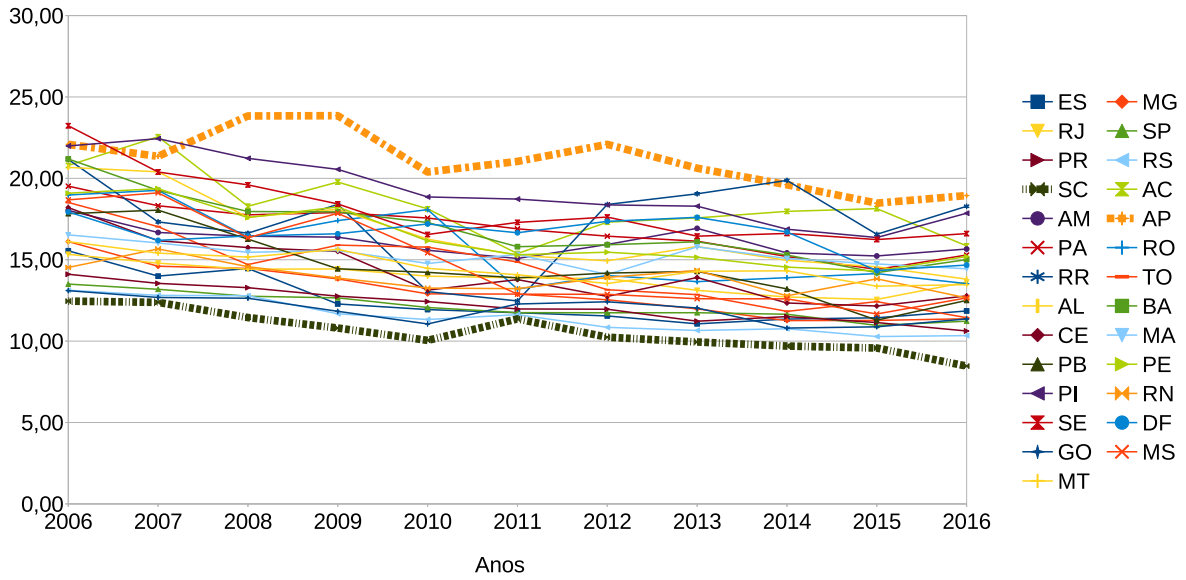
4.2 Processamento da base de dados

O DATASUS disponibilizou os dados do SINASC e do SIM e assim, foram obtidos todas as informações dos bebês que nasceram e morreram em todos os estados brasileiros no período de 2006 a 2017. O Gráfico 2 apresenta a TMI para todos os estados brasileiros. A taxa reflete o número de óbitos de menores de um ano de idade, por mil nascidos vivos, para cada estado e ano considerado.

No Gráfico 2 é possível visualizar que os estados de SC e do AP apresentam, respectivamente, a menor (9,58) e maior (18,48) taxa de mortalidade infantil quando comparada com os demais estados brasileiros. Esta é inclusive a razão de termos

considerado estes 02 (dois) estados como estudo de caso para análise da mortalidade infantil no Brasil.

Gráfico 2 – Taxa de mortalidade infantil nos estados do Brasil - 2006 a 2016



Fonte: Elaborado pelo autor.

No pré-processamento da base de dados, os registros foram assim classificados: os do SIM, que também estavam presentes no SINASC, foram rotulados como 'Óbito infantil'. Os demais registros do SINASC mantiveram sua classificação como 'Vivo'. A descrição dos 17 atributos é apresentada na Tabela 3, sendo que existem os seguintes atributos comuns na base de dados do SINASC e do SIM: idade da mãe, escolaridade da mãe, quantidade de filho vivos, quantidade de filhos mortos, gravidez, gestação, parto, peso, sexo e raça.

Após a criação da base de dados[¶] referente às informações de crianças vivas e mortas, foram realizados diferentes processamentos com o objetivo de:

1) *Eliminar instâncias inconsistentes*: foram eliminados os registros do SIM e SINASC que eram iguais, mas que continham classificações diferentes.
2) *Eliminar instâncias redundantes*: Para evitar que instâncias redundantes participassem mais de uma vez do processo de ajuste de parâmetros de um modelo, apenas uma ocorrência do registro foi mantida na base de dados.
3) *Transformar atributos numéricos*: Uma vez que os limites inferior e superior dos valores dos atributos eram discrepantes, torna-se necessário realizar a normalização

[¶]Neste trabalho foram considerados os dados do ano de 2016, pois o dados de 2017 não possuíam os lançamentos de possíveis óbitos que eventualmente ocorreu em 2018.

Tabela 3 – Descrição dos atributos da base SINASC

Atributos	**Descrição**
Idade da Mãe	Em anos
Escolaridade da Mãe	Em anos
Quantidade de filhos vivos	Numérico contínuo
Quantidade de filhos mortos	Numérico contínuo
Gravidez	Única, dupla, tripla ou mais
Gestação	Em semanas
Parto	Normal, cesáreo
Peso	Em gramas ao nascer
Sexo	Masculino, feminino
Raça	Branca, Preta, Amarela, Parda, Indígena
Local de ocorrência	Hospital, domicílio, outros, etc
Estado Civil	Solteira, Casada, Separada, Viúva, etc
Grupo da ocupação da Mãe	Grande grupo de ocupação definidos na CBO
Número de consultas pré-natal	De 1 a 3, de 4 a 6, etc
APGAR1	Numérico contínuo
APGAR5	Numérico contínuo
Anomalia congênita	Sem anomalia, com anomalia
Classificação	Vivo, óbito infantil

dos atributos numéricos para a aplicação dos algoritmos de SVM e RNA. Para os algoritmos de árvore e regras, os valores numéricos não sofreram transformação.

4) *Efetuar conversão simbólico-numérica*: Técnicas tais como RNA e SVM lidam apenas com dados numéricos. Uma vez que a base de dados, descrita no Quadro 1, continha atributos discretos, foi necessário realizar uma conversão destes atributos. Desta forma, os atributos nominais não ordinais, tais como 'Parto' que continham as opções 'normal', 'cesáreo', foram binarizados. Ou seja, este atributo foi codificado como presença ou ausência de parte 'normal' e 'cesáreo'. Isso foi realizado para todos os outros atributos nominais não ordinais presentes na base de dados.
5) *Eliminar instâncias com ruído*: Instâncias com ruído contêm objetos que, aparentemente, não pertencem à distribuição que gera os dados analisados. Neste trabalho foram encontrados os seguintes ruídos: idade da mãe igual ou acima de 75 anos, quantidade de filhos vivos igual ou acima de 75, quantidade de filhos mortos igual ou acima de 75 e APGAR menor que 0 (zero) / maior que 10 (dez). Desta forma, por não se conhecer os valores reais para os atributos, os registros referentes aos respectivos ruídos foram desconsiderados da base de dados. Todas estas etapas do pré-processamento foram realizadas com o objetivo de gerar uma base de dados única que englobasse as classes 'Vivo' e 'Óbito Infantil'. Após unificadas, verificou-se que as bases continham instâncias desproporcionais quanto às classes, sendo necessária a aplicação de técnicas de balanceamento de dados; sendo esta, portanto, a última etapa do pré-processamento da base de dados.

6) *Balanceamento de classes*: Um dos problemas que frequentemente prejudica o desempenho dos algoritmos é a quantidade desbalanceada de classes que acarreta uma sobreposição estatística entre a classe majoritária e a minoritária (PRATI, 2006). Ou seja, em alguns contextos, o número de instâncias de uma classe é muito maior do que de outra, o que acaba influenciando o aprendizado da classe majoritária e prejudicando a classe minoritária (FACELI et al., 2011). Diante de dados desbalanceados, pode-se realizar o balanceamento a partir das seguintes abordagens:

A) *Oversampling*: consiste na replicação de instâncias da classe minoritária. Nesta situação, o acréscimo de instâncias poderá incorporar situações que nunca ocorrerão na prática.

B) *Undersampling*: consiste na eliminação de instâncias da classe majoritária que, entretanto, poderá levar à eliminação de dados relevantes que poderão comprometer a indução do modelo.

Observando-se a Tabela 4, nota-se que existe uma desproporção entre as classes 'Vivo' e 'Óbito Infantil' nos estados de Santa Catarina e do Amapá.

Tabela 4 – Dimensões das Bases de Dados Unificadas com relação ao ano de 2016, antes e após o balanceamento

	Antes do balanceamento		Após o balanceamento	
	Santa Catarina	Amapá	Santa Catarina	Amapá
Vivo	92558	13079	501	152
Óbito Infantil	501	152	501	152

Diante deste fato, neste trabalho, foi utilizada a abordagem *undersampling* no ambiente WEKA[‖], um software gratuito e de código aberto dedicado à área de aprendizagem de máquina. O método utilizado para o balanceamento de classes foi o *SpreadSubsample* do WEKA. O *SpreadSubsample* é caracterizado pelo fato de eliminar instâncias da(s) classe(s) majoritárias produzindo um subconjunto aleatório a partir de um conjunto de dados visando que a classe majoritária fique com o mesmo número de instâncias da classe minoritária, configurando o *spread* igual a 1 (distribuição uniforme).

4.3 Descrição dos métodos

Para a construção da árvore, utilizou-se o algoritmo C4.5, uma implementação de código aberto em Java na ferramenta WEKA. Para utilizá-lo é preciso o ajuste de alguns de seus parâmetros, tais como o limite de confiança para a poda da árvore (**C**) e o número mínimo de instâncias por folha da árvore (**M**).

[‖]Disponível em http://www.cs.waikato.ac.nz/ml/weka

Para este ajuste, foi utilizado o pacote *Visually Tuned J48* (VTJ48), desenvolvido por Stiglic et al. (2012), que ajusta de forma automática estes parâmetros visando construir árvores menores e mais fáceis de serem visualizadas. Assim, os valores dos parâmetros C e M, ajustados pelo VTJ48, estão descritos na Tabela 5.

Tabela 5 – Valores dos parâmetros C e M com relação ao C4.5

Santa Catarina	
Parâmetros	**2016**
C	0.078125
M	2
Amapá	
Parâmetros	**2016**
C	0.03125
M	2

Quanto ao algoritmo RIPPER, este foi executado utilizando-se de parâmetros *default* fornecidos pela ferramenta. Com relação ao Random Forest, foi realizado o ajuste dos seguintes parâmetros: *numIterations*, *numFeatures* e *maxDepth*. Para ajustar os hiperparâmetros do Random Forest, foi utilizado o algoritmo *MultiSearch*[**], similar ao algoritmo *GridSearch* (CHANG; LIN, 2011), o qual realiza uma busca exaustiva sobre os valores dos parâmetros especificados e encontra as melhores combinações possíveis para a base de dados considerada. A Tabela 6 apresenta os parâmetros estabelecidos para a busca e a Tabela 7 apresenta os valores adotados para cada hiperparâmetro após o ajuste.

Tabela 6 – Faixa de ajuste dos valores dos hiperparâmetros com relação ao Random Forest

Santa Catarina / Amapá - Ano: 2016			
Parâmetros	**Num Iterations**	**Num Features**	**maxDepth**
Valor mínimo	1	2.0	2.0
Valor máximo	200	16.0	12.0

Tabela 7 – Valores dos hiperparâmetros com relação ao Random Forest

Santa Catarina / Amapá - Ano: 2016			
Parâmetros	**Num Iterations**	**Num Features**	**maxDepth**
Santa Catarina	60	2	5
Amapá	180	2	12

Para o treinamento do algoritmo SVM, foram ajustados os seguintes parâmetros: *degree*, *cost*, *eps*, *gamma* e *coef0*. Estes parâmetros são importantes para que sejam encontrados bons resultados em sua execução, pois estão diretamente ligados à etapa de

[**]Disponível em http://weka.sourceforge.net/packageMetaData/multisearch/index.html

treinamento e classificação do algoritmo. Foi utilizada a função polinomial como kernel do SVM. O ajuste dos hiperparâmetros do SVM foi realizado utilizando-se o algoritmo *MultiSearch*. A Tabela 8 apresenta os parâmetros estabelecidos para a busca e a Tabela 9 exibe os valores encontrados para cada hiperparâmetro após o ajuste.

Tabela 8 – Faixa de ajuste dos valores dos hiperparâmetros com relação ao SVM

Santa Catarina / Amapá - Ano: 2016					
Parâmetros	**Degree**	**Cost**	**Eps**	**Gamma**	**Coef0**
Valor mínimo	1	0,1	0,0001	1	0,1
Valor máximo	4	100	1	0,2	100

Tabela 9 – Valores dos hiperparâmetros com relação ao SVM

Santa Catarina / Amapá - Ano: 2016					
Parâmetros	**Degree**	**Cost**	**Eps**	**Gamma**	**Coef0**
Santa Catarina	1	10	1	0.25	1
Amapá	1	1	1	0.125	1

Para o treinamento da RNA, utilizou-se a estrutura *Multilayer Perceptron* com o algoritmo *Backpropagation* e foi realizado o ajuste dos seguintes parâmetros: *hiddenLayers*, *learningRate* e *momentum*. Foi utilizada uma camada com $2n+1$ neurônios (LI; VITáNYI, 1990), sendo n=54 número de entradas para a rede, totalizando 109 neurônios na camada intermediária. Quanto à camada de saída foram utilizados 2 neurônios, 1 para cada classe. Os parâmetros estabelecidos para a busca e definidos após o ajuste estão descritos nas Tabelas 10 e 11.

Tabela 10 – Faixa de ajuste dos valores dos hiperparâmetros com relação à RNA

Santa Catarina / Amapá - Ano: 2016			
Parâmetros	**Hidden Layers**	**Learning Rate**	**Momentum**
Valor mínimo	1	0.0	0.0
Valor máximo	3	1.0	1.0

Tabela 11 – Valores dos hiperparâmetros com relação à RNA

Santa Catarina / Amapá - Ano: 2016			
Parâmetros	**Hidden Layers**	**Learning Rate**	**Momentum**
Santa Catarina	1	0,6	0,5
Amapá	2	0,6	0,5

4.4 Métricas de avaliação

Para avaliação da qualidade dos modelos obtidos, foram utilizadas as métricas de precisão, sensibilidade e F-measure.

A Precisão (Equação 4.1) mede a proporção de instâncias classificadas em determinada classe que são realmente da classe:

$$Pr = \frac{VP}{VP + FP} \tag{4.1}$$

A Sensibilidade (Equação 4.2) estabelece a proporção de instâncias corretamente classificadas dentre todas as instâncias de uma classe:

$$Sen = \frac{VP}{VP + FN} \tag{4.2}$$

Por sua vez, o F-measure (Equação 4.3) representa a média harmônica entre precisão e sensibilidade:

$$F - measure = \frac{2 \times Sen \times Pr}{Sen + Pr} \tag{4.3}$$

em que VP= Verdadeiros Positivos, FP=Falsos Positivos e FN = Falsos Negativos.

Para definir os conjuntos de treinamento e teste, utilizou-se o método *cross-validation* de 10 (dez) dobras que tem como objetivo avaliar a capacidade de generalização do modelo (KOHAVI, 1995).

5 RESULTADOS E DISCUSSÕES

Este capítulo exibe os resultados quantitativos com relação à mortalidade infantil nas cinco regiões do Brasil: Sul, Sudeste, Centro-Oeste, Nordeste e Norte. Também são apresentados os resultados dos algoritmos com relação à caracterização da mortalidade infantil nos dois estados brasileiros avaliados: Santa Catarina e Amapá.

5.1 Quantitativo da taxa de mortalidade infantil nas cinco regiões brasileiras

O Brasil é dividido em 05 (cinco) regiões que englobam os seguintes estados:

- Sul: *PR (Paraná), RS (Rio Grande do Sul) e SC (Santa Catarina).*
- Sudeste: *ES (Espírito Santo), MG (Minas Gerais), RJ (Rio de Janeiro) e SP (São Paulo).*
- Centro-Oeste: *DF (Distrito Federal), GO (Goiás), MS (Mato Grosso do Sul) e MT (Mato Grosso).*
- Nordeste: *AL (Alagoas), BA (Bahia), CE (Ceará), MA (Maranhão), PB (Paraíba), PE (Pernambuco), PI (Piauí), RN (Rio Grande do Norte) e SE (Sergipe).*
- Norte: *AC (Acre), AM (Amazonas), AP (Amapá), PA (Pará), RO (Rondônia), RR (Roraima) e TO (Tocantins).*

As regiões do Brasil apresentam diferentes taxas de mortalidade infantil com relação aos seus estados, conforme mostra a Figura 3. Ela apresenta as informações sobre a quantidade de nascidos e a quantidade de óbito infantil de todos estados das regiões do Brasil. É possível constatar que os estados de SC e do AP possuem, respectivamente, a menor e maior TMI do país quando analisados os últimos 02 (dois) anos (2015 e 2016), com valores médios de 9,02 e 18,71, conforme Tabela 12. Observa-se uma TMI muito alta para o estado do Amapá.

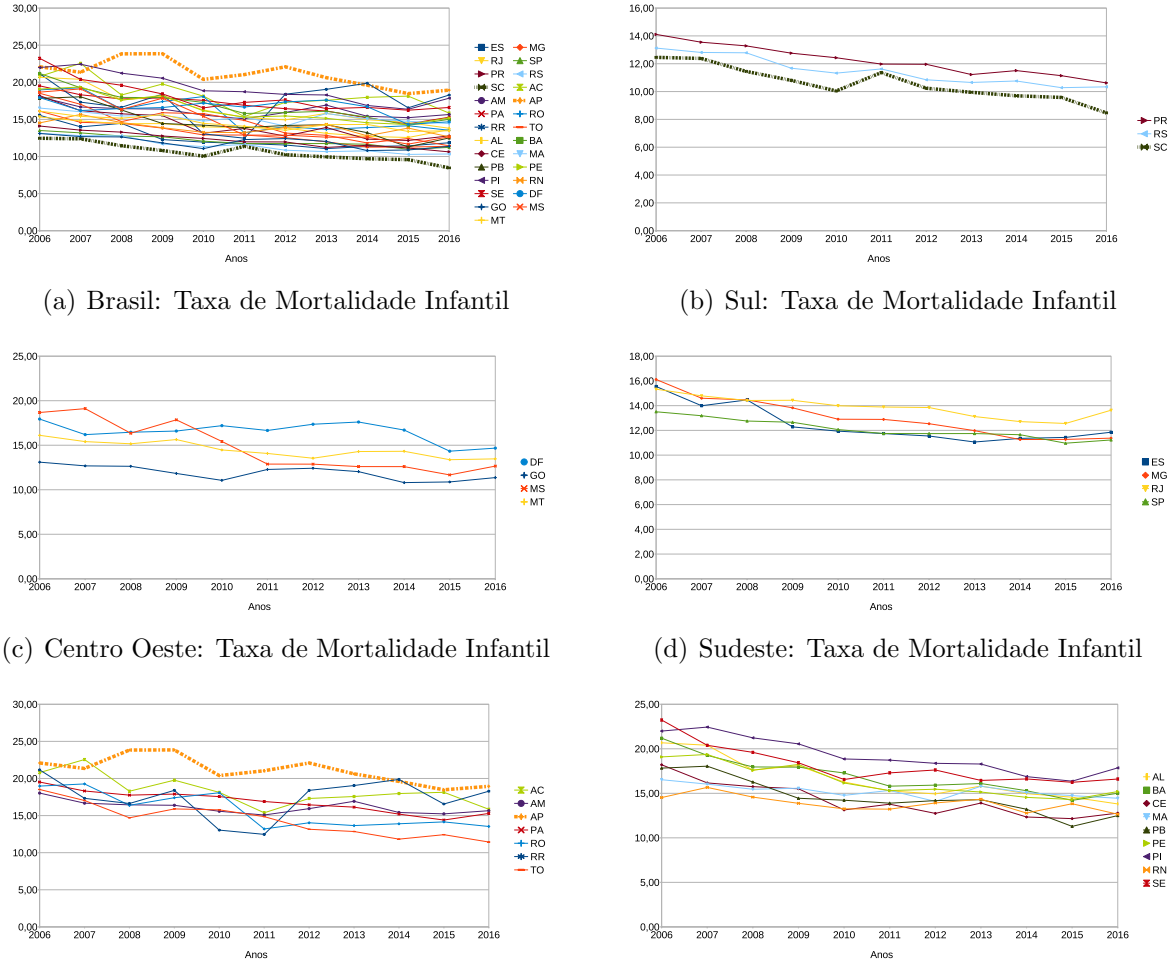
Tabela 12 – Taxa de mortalidade infantil em Santa Catarina e no Amapá

Taxa de Mortalidade Infantil			
	2015	**2016**	**Média (anos: 2015 e 2016)**
Santa Catarina	9,58	8,47	9,02
Amapá	18,48	18,94	18,71

Segundo o IBGE (2018), a TMI e as expectativas de vida são indicadores que estão relacionados com os aspectos sanitários, de segurança e de saúde de uma determinada

população. Assim, estes indicadores são instrumentos importantes para avaliação e para tomada de decisão na adoção de políticas públicas de acordo com a necessidade constatada.

Figura 3 – Taxa de mortalidade infantil no Brasil.



(a) Brasil: Taxa de Mortalidade Infantil

(b) Sul: Taxa de Mortalidade Infantil

(c) Centro Oeste: Taxa de Mortalidade Infantil

(d) Sudeste: Taxa de Mortalidade Infantil

(e) Norte: Taxa de Mortalidade Infantil

(f) Nordeste: Taxa de Mortalidade Infantil

Fonte: Elaborado pelo autor.

Diante deste contexto e analisando a Figura 3, percebe-se que os estados mais desenvolvidos e que apresentam, em linhas gerais, melhores condições sanitárias, de saúde e de educação tendem a apresentar as menores taxas de mortalidade infantil quando comparados com os estados menos desenvolvidos. Por exemplo, os estados que possuem as menores TMI em 2016 são Santa Catarina, Rio Grande do Sul e Paraná, com TMI de 8,47, 10,34 e 10,62, respectivamente. Já os estados do Amapá, Roraima e Piauí, localizados em regiões mais vulneráveis, possuem as maiores taxas de mortalidade infantil: 18,94, 18,28 e 17,86, respectivamente.

É importante salientar ainda que muito provavelmente há diferenças significativas

de TMI entre as microrregiões de cada estado, visto o tamanho dos estados brasileiros e as diferenças entre as microrregiões.

5.2 Caracterização da mortalidade infantil nos estados de Santa Catarina e do Amapá

Esta seção apresenta os resultados obtidos após a aplicação dos algoritmos C4.5, RIPPER, Random Forest, SVM e RNA no contexto de mortalidade infantil para os estados brasileiros de SC e do AP no ano de 2016. Desta forma, a Tabela 13 apresenta os resultados das métricas de avaliação de qualidade dos modelos para os dois estados analisados.

Tabela 13 – Resultados dos algoritmos C4.5, RIPPER, Random Forest, SVM e RNA, em porcentagem, com relação ao ano de 2016.

	Santa Catarina			**Amapá**		
	Precisão	**Sensibilidade**	**F-Measure**	**Precisão**	**Sensibilidade**	**F-Measure**
C4.5						
Vivo	81,4	93,6	87,1	77,3	92,1	84,1
Óbito Infantil	92,5	78,6	85,0	90,2	73,0	80,7
Média	87,0	86,1	86,0	83,8	82,6	82,4
RIPPER						
Vivo	78,9	91,0	84,5	81,1	87,5	84,2
Óbito Infantil	89,4	75,6	81,9	86,4	79,6	82,9
Média	84,1	83,3	83,2	83,8	83,6	83,5
Random Forest						
Vivo	83,7	93,2	88,2	84,5	93,4	88,8
Óbito Infantil	92,3	81,8	86,8	92,6	82,9	87,5
Média	88,0	87,5	87,5	88,6	88,2	88,1
SVM						
Vivo	83,5	93,8	88,3	82,0	92,8	87,0
Óbito Infantil	92,9	81,4	86,8	91,7	79,6	85,2
Média	88,2	87,6	87,6	86,8	86,2	86,1
RNA						
Vivo	82,4	93,2	87,5	82,8	88,8	85,7
Óbito Infantil	92,2	80,0	85,7	87,9	81,6	84,6
Média	87,3	86,6	86,6	85,4	85,2	85,2

Observando-se a tabela, vemos que os algoritmos Random Forest, SVM e Rede Neural obtiveram os melhores resultados, com valores muito semelhantes. É importante salientar ainda que para muitos problemas complexos a SVM e a RNA podem eventualmente classificar melhor o problema quando comparado com o C4.5/RIPPER. No entanto, existe uma desvantagem de não representar o conhecimento de uma forma mais fácil de se interpretar. Neste caso, então, os algoritmos C4.5 e RIPPER, que explicitam por meio de regras o conhecimento adquirido pelos métodos de aprendizado, apresentaram um comportamento promissor na caracterização da mortalidade infantil, já que os valores

são também semelhantes.

Analisando-se a média harmônica entre a precisão e sensibilidade, representada pela métrica F-measure, o valor médio dos algoritmos foi de aproximadamente de 90% para os dois estados. Nota-se ainda que a sensibilidade da classe 'Óbito Infantil' é ligeiramente menor do que a classe 'Vivo' para todos os algoritmos. No entanto, a precisão é ligeiramente maior para esta mesma classe. Isso sinaliza que os modelos estão errando mais instâncias da classe 'Óbito Infantil' do que da classe 'vivo', mas o número de falsos positivos para a classe de 'Óbito Infantil' é menor. Ou seja, pelos resultados obtidos, há bebês que morreram, mas os modelos de classificados utilizados os classificaram como vivos. Isso pode indicar que há bebês que têm perfil de estarem vivos, mas, por alguma razão, morreram. Uma vez analisadas as métricas de avaliação dos modelos, avaliamos também as regras que caracterizam as classes 'Vivo' e 'Óbito infantil'. Estas regras estão apresentadas nas Tabelas 14 e 15.

Tabela 14 – Principais regras geradas pelos algoritmos C4.5 e RIPPER referente ao estado do Amapá no ano de 2016

Regras obtidos pelo C4.5		
Número	Descrição da regra	Instâncias (porcentagem)
1	Se PESO <= 2.300 gramas e ENTÃO MORTO	63%
2	Se PESO > 2.300 gramas e APGAR5 <= 8 ENTÃO MORTO	13%
3	Se PESO > 2.300 gramas e APGAR5 > 8 ENTÃO VIVO	92%

Regras obtidos pelo RIPPER		
Número	Descrição da regra	Instâncias (porcentagem)
1	Se PESO <= 2.300 gramas ENTÃO MORTO	61%
2	Se APGAR1 <= 7 ENTÃO MORTO	17%
3	Se GESTAÇÃO = ? ENTÃO MORTO	3%
4	SE NÃO se enquadra em nenhuma das regras anteriores ENTÃO VIVO	89%

Tabela 15 – Principais regras geradas pelos algoritmos C4.5 e RIPPER referente ao estado de Santa Catarina no ano de 2016

Regras obtidos pelo C4.5		
Número	Descrição da regra	Instâncias (porcentagem)
1	Se APGAR1 <= 6 ENTÃO MORTO	65%
2	Se APGAR1 > 6 e SEM ANOMALIA CONGÊNITA e PESO <= 1.535 gramas ENTÃO MORTO	7%
3	Se APGAR1 > 6 e SEM ANOMALIA CONGÊNITA e PESO > 1.535 gramas ENTÃO VIVO	95%

Regras obtidos pelo RIPPER		
Número	Descrição da regra	Instâncias (porcentagem)
1	Se APGAR5 >= 9 e GESTAÇÃO = 37 a 41 semanas e QUANTIDADE FILHO VIVO <= 0 ENTÃO VIVO	29%
2	Se APGAR5 >= 9 e PESO >= 3.520 gramas ENTÃO VIVO	27%
3	Se PESO >= 2.775 gramas e APGAR5 >= 10 ENTÃO VIVO	14%
4	Se APGAR1 >= 8 e GESTAÇÃO = 37 a 41 semanas e SEM ANOMALIA CONGÊNITA ENTÃO VIVO	11%
5	Se APGAR5 >= 9 e PESO >= 3.250 gramas ENTÃO VIVO	6%
6	Se PESO >= 1.575 gramas e APGAR1 >= 7 e SEM ANOMALIA CONGÊNITA QUANTIDADE FILHO VIVO <= 2 ESCOLARIDADE DA MÃE = 8 a 11 anos ENTÃO VIVO	5%
7	SE NÃO se enquadra em nenhuma das regras anteriores ENTÃO MORTO	83%

A Tabela 14 apresenta as duas principais regras de classificação obtidas a partir do C4.5 para o estado de Amapá no ano de 2016: em linhas gerais, se o bebê tiver um peso menor ou igual a 2.300 gramas então é classificado como MORTO, Assim, esta única regra classifica 63% das instâncias da classe 'Óbito Infantil'. Quanto à classe 'Vivo', a maioria das instâncias, 92%, é classificada como VIVO se o peso for maior que 2.300 gramas e se o APGAR 5 for maior que 8. Com relação ao RIPPER, foi encontrada uma regra que se o peso do bebê for menor ou igual a 2.300 gramas, classifica como MORTO 61% das instâncias, similar à regra encontrada pelo C4.5.

De acordo com a Tabela 15, as duas principais regras de classificação obtidas para cada classe a partir do algoritmo C4.5 para o estado de Santa Catarina são as seguintes: se o bebê tiver APGAR 1 maior que 6 e não tiver anomalia congênita e o peso for maior que 1.535 gramas é classificado como VIVO. Desta forma, esta regra sozinha classifica 95% das instâncias desta classe. Por outro lado, o bebê é classificado como MORTO se o APGAR 1 for menor ou igual a 6, assim esta regra sozinha classifica 65% das instâncias da classe 'Óbito Infantil'.

Analisando a Tabela 14, especificamente a regra 4 do RIPPER, constata-se que a mesma classifica 89% das instâncias como MORTO, se não se enquadrar em nenhuma das regras 1, 2 e 3 anteriormente descritas. Situação semelhante é constatada na Tabela 15, especificamente a regra número 7 do RIPPER, constata-se que a mesma classifica 83% instâncias como MORTO, se não se enquadrar em nenhuma das regras anteriormente descritas.

Assim, é possível observar que os atributos que compõem as regras no estado do Amapá são: peso, APGAR 5, APGAR 1 e gestação. Quanto ao estado de Santa Catarina, estes atributos são: APGAR 1, APGAR 5, peso, anomalia congênita, quantidade de filho vivo, escolaridade da mãe e gestação. Desta forma, comparando as regras geradas pelo C4.5 e pelo RIPPER, constata-se que os atributos comuns em ambos os estados foram: peso, APGAR 1, APGAR 5 e gestação.

Baseando nas regras principais descritas nas Tabelas 14 e 15, verifica-se que peso e o APGAR 1 se apresentaram como sendo um dos importantes atributos na classificação do bebê como VIVO para os dois estados avaliados. Constatou-se também que o peso mínimo, identificado pelo C4.5, no estado do Amapá foi maior que no estado de Santa Catarina, cujos pesos mínimos foram de 2.300 gramas e de 1.535 gramas para classificar o bebê como VIVO, dentre outros atributos que foram considerados. Ou seja, os bebês, para sobreviverem no estado do Amapá, precisam ter um peso maior do que os bebês de Santa Catarina.

Além dos atributos identificados pelos algoritmos C4.5 e RIPPER, foram analisados também os principais atributos indicados pelo algoritmo Random Forest, conforme

apresentado nas Tabelas 16 e 17. Assim, constatou-se que os atributos comuns no ranking dos dois estados foram: peso, idade da mãe, gestação, quantidade de filho vivo, APGAR 1, APGAR 5 e escolaridade da mãe.

Tabela 16 – Ranking dos 8 principais atributos gerados pelo Random Forest referente ao estado do Amapá em 2016

Amapá	
Atributos	**Fator de importância**
Peso	0.5
Idade da mãe	0.46
Quantidade de filho vivo	0.43
Gestação	0.42
Escolaridade da mãe	0.37
Número de consultas pré-natal	0.37
APGAR 5	0.36
APGAR 1	0.35

Tabela 17 – Ranking dos 8 principais atributos gerados pelo Random Forest referente ao estado de Santa Catarina em 2016

Santa Catarina	
Atributos	**Fator de importância**
Peso	0.37
Gestação	0.31
Idade da mãe	0.31
APGAR 1	0.3
APGAR 5	0.28
Quantidade de filho vivo	0.28
Estado civil da mãe	0.25
Escolaridade da mãe	0.24

A partir dos respectivos atributos comuns, foram realizados os cálculos das médias e dos desvios padrão dos atributos numéricos, conforme apresentado na Tabela 18. Com relação aos atributos nominais: gestação e escolaridade da mãe, foi realizado um levantamento estatístico, apresentado nas Tabelas 19 e 20.

A Tabela 18 destaca as médias e os desvios padrão da idade da mãe, da quantidade de filho vivo, da quantidade de filho morto, do peso, do APGAR1 e do APGAR5 com relação aos estados de SC e do AP no ano de 2016. Salienta-se que as médias e os desvios padrão foram calculados considerando os seguintes grupos: bebês que sobreviveram no primeiro ano de vida (Vivo), bebês que morreram no período neonatal (Neonatal) e os bebês que morreram no período pós-neonatal (Pós-Neonatal).

Tabela 18 – Resultados das médias e dos desvios padrão com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Média	**Desvio padrão**	**Média**	**Desvio padrão**
Vivo				
Idade da mãe	27,28	6,49	24,83	6,66
Quantidade de filho vivo	0,85	1,10	1,40	1,72
Quantidade de filho morto	0,21	0,52	0,25	0,58
Peso	3.231	536	3.212	534
APGAR1	8,34	1,12	8,50	1,08
APGAR5	9,30	0,77	9,60	0,83
Óbito Infantil - Neonatal				
Idade da mãe	27,26	7,19	24,81	6,74
Quantidade de filho vivo	1,21	1,24	2,16	1,87
Quantidade de filho morto	0,31	0,69	0,39	0,93
Peso	1.659	1.123	1.642	1.013
APGAR1	4,17	2,93	4,79	3,03
APGAR5	5,65	3,13	6,44	2,86
Óbito Infantil - Pós-Neonatal				
Idade da mãe	26,27	6,56	25,63	6,39
Quantidade de filho vivo	1,48	1,37	2,50	1,67
Quantidade de filho morto	0,28	0,63	0,52	1,2
Peso	2.430	985	2.546	900
APGAR1	6,55	2,46	8,11	1,55
APGAR5	8,18	1,77	9,19	1,04

Analisando esta tabela, verifica-se que os atributos APGAR 1 e o peso têm apresentado uma significativa diferenciação entre a classe de bebês que sobreviveram após um ano (VIVO) com aqueles que vieram a óbito no período neonatal (ÓBITO INFANTIL - NEONATAL). No caso do estado de SC, o bebê que sobreviveu após um ano (VIVO) nasceu em média com um peso de 3.231 gramas (desvio padrão de 536 gramas) e com uma média do APGAR 1 de 8,34 (desvio padrão de 1,12). Já o bebê que faleceu no período neonatal (ÓBITO INFANTIL - NEONATAL) nasceu com um peso médio de 1.639 gramas (1.123 gramas) e com uma média do APGAR 1 de 4,17 (desvio padrão de 2,93).

Esta diferença também é constatada no estado de AP, pois o bebê que sobreviveu após um ano (VIVO) nasceu com um peso médio de 3.212 gramas (desvio padrão de 534 gramas), e com um APGAR 1 médio de 8,50 (desvio padrão de 1,08). Já o bebê que faleceu no período neonatal (ÓBITO INFANTIL - NEONATAL) nasceu em média com um peso de 1.642 gramas (1.013 gramas) e com uma média do APGAR 1 de 4,79 (desvio padrão de 3,03).

Na Tabela 19 é apresentada uma análise do atributo 'escolaridade da mãe'. A partir desta tabela, verifica-se que as mães no estado de Santa Catarina apresentam uma escolaridade um pouco mais elevada, em termos percentuais, nas faixas de 8 a 11 anos e

de 12 ou mais anos quando comparadas com as mãe no estado do Amapá. Se analisarmos a escolaridade das mães dos bebês que morreram no período pós-neonatal, por exemplo, apenas 8,33% tinham 12 anos ou mais de estudo.

Tabela 19 – Levantamento estatístico com relação à escolaridade da mãe nos estados de Santa Catarina e do Amapá no ano de 2016.

	Santa Catarina		**Amapá**	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Vivo				
Aspecto: Escolaridade da mãe				
Não informado	254	0,27%	36	0,27%
Nenhuma	112	0,12%	88	0,67%
De 1 a 3 anos	1.198	1,29%	461	3,52%
De 4 a 7 anos	12.035	13,00%	3.067	23,44%
De 8 a 11 anos	54.761	59,16%	7.554	57,75%
De 12 ou mais anos	24.198	26,14%	1.873	14,32%
Óbito Infantil - Neonatal				
Aspecto: Escolaridade da mãe				
Não informado	9	2,35%	13	11,20%
Nenhuma	6	1,55%	2	1,72%
De 1 a 3 anos	18	4,65%	5	4,31%
De 4 a 7 anos	84	21,70%	35	30,17%
De 8 a 11 anos	178	45,99%	47	40,51%
De 12 ou mais anos	92	23,77%	14	12,06%
Óbito Infantil - Pós-Neonatal				
Aspecto: Escolaridade da mãe				
Não informado	1	0,87%	0	0,0%
Nenhuma	4	3,50%	0	0,0%
De 1 a 3 anos	1	0,87%	7	19,44%
De 4 a 7 anos	24	21,05%	13	36,11%
De 8 a 11 anos	65	57,01%	13	36,11%
De 12 ou mais anos	19	16,66%	3	8,33%

A Tabela 20 apresenta uma análise do atributo 'gestação'. Verifica-se que a maioria dos bebês classificados como VIVO tiveram uma gestação de 37 a 41 semanas, com um comportamento similar em ambos os estados. Em relação aos bebês que faleceram no período neonatal, foi constatado que uma parcela significativa destes bebês nasceram de uma forma prematura (entre menos de 22 semanas até 32 a 36 semanas), representando 71,56% em SC e 68,09% no AP. Já os percentuais de bebês prematuros que faleceram no período pós-neonatal foram de 42,09% em SC e 38,88% no AP.

Uma outra questão observada é que nos dois estados o número de óbitos infantis é maior no período neonatal quando comparado ao período pós-neonatal, conforme Tabela 21.

De fato, um número expressivo de bebês falecem no período neonatal em ambos

Tabela 20 – Levantamento estatístico com relação à gestação nos estados de Santa Catarina e do Amapá no ano de 2016.

	Santa Catarina		Amapá	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Vivo				
Aspecto: Gestação				
Não informado	149	0,16%	580	4,43%
Menos de 22 semanas	16	0,01%	3	0,02%
De 22 a 27 semanas	234	0,25%	27	0,20%
De 28 a 31 semanas	790	0,85%	117	0,89%
De 32 a 36 semanas	9.023	9,74%	1.465	11,20%
De 37 a 41 semanas	80.378	86,84%	10.347	79,11%
De 42 a mais semanas	1.968	2,12%	540	4,12%
Óbito Infantil - Neonatal				
Aspecto: Gestação				
Não informado	6	1,55%	10	8,62%
Menos de 22 semanas	10	2,58%	15	12,93%
De 22 a 27 semanas	148	38,24%	27	23,27%
De 28 a 31 semanas	58	14,98%	36	22,41%
De 32 a 36 semanas	61	15,76%	11	9,48%
De 37 a 41 semanas	102	26,35%	27	23,27%
De 42 a mais semanas	2	0,51%	0	0,0%
Óbito Infantil - Pós-Neonatal				
Aspecto: Gestação				
Não informado	6	5,26%	4	11,11%
Menos de 22 semanas	2	1,75%	4	11,11%
De 22 a 27 semanas	13	11,40%	2	5,55%
De 28 a 31 semanas	12	10,52%	4	11,11%
De 32 a 36 semanas	21	18,42%	4	11,11%
De 37 a 41 semanas	59	51,75%	18	50,00%
De 42 a mais semanas	1	0,87%	0	0,0%

Tabela 21 – Quantitativo de óbitos infantis no período neonatal e pós-neonatal nos estados de Santa Catarina e do Amapá com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito infantil				
Período NeoNatal	387	77,25%	116	76,32%
Período Pós-NeoNatal	114	22,75%	36	23,68%
Total	501	100,00%	152	100,00%

os estados. Em 2016 atingiu os percentuais de 77,25% em SC e de 76,32% no AP. Diante disso, verificou-se oportuno buscar encontrar possíveis causas que expliquem este comportamento, e assim, foram realizados outros levantamentos conforme as Tabelas 22 a 28.

A Tabela 22 exibe as causas da mortalidade infantil nos períodos neonatal e pós-neonatal. Somando as causas básicas de óbito que ocorreram nestes períodos, constata-se que os 02 (dois) principais motivos de mortalidade infantil foram afecções originadas no período perinatal e malformações congênitas.

Tabela 22 – Resultados estatísticos sobre as causas básicas de óbito com relação ao ano de 2016.

	Santa Catarina		**Amapá**	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito Infantil - Neonatal				
Afecções originadas no período perinatal	290	74,93%	99	85,34%
Doenças endócrinas	1	0,25%	0	0,0%
Doenças infecciosas e parasitárias	2	0,51%	0	0,0%
Malformações congênitas	90	23,25%	17	14,65%
Achados anormais de exames	4	1,03%	0	0,0%
Óbito Infantil - Pós-Neonatal				
Afecções originadas no período perinatal	34	29,82%	12	33,33%
Causas externas de mortalidade	4	3,50%	0	0,0%
Doenças da pele e do tecido subcutâneo	1	0,87%	2	5,55%
Doenças do aparelho circulatório	6	5,26%	0	0,0%
Doenças do aparelho digestivo	0	0,0%	2	5,55%
Doenças do aparelho geniturinário	0	0,0%	2	5,55%
Doenças do aparelho respiratório	11	9,64%	5	13,88%
Doenças do ouvido e da apófise mastóide	0	0,0%	1	2,77%
Doenças imunitárias/hematopoéticas	1	0,87%	1	2,77%
Doenças do sistema nervoso	3	2,63%	0	0,0%
Doenças endócrinas	1	0,87%	1	2,77%
Doenças infecciosas e parasitárias	3	2,63%	3	8,33%
Malformações congênitas	37	32,45%	5	13,88%
Neoplasias [tumores]	2	1,75%	0	0,0%
Achados anormais de exames	11	9,64%	2	5,55%

Os principais sub-grupos das afecções originadas no período perinatal relacionadas no Capítulo XVI* da CID-10 são os seguintes: feto/recém-nascido afetado por fatores maternos/complicações da gravidez/parto, transtornos relacionados com a duração da gestação/crescimento fetal, traumatismo de parto, transtornos respiratórios/cardiovasculares específicos do período perinatal, infecções específicas do período perinatal, transtornos hemorrágicos/hematológicos do feto/recém-nascido, transtornos endócrinos/metabólicos transitórios específicos do feto/recém-nascido, transtornos do aparelho digestivo do feto/recém-nascido, afecções comprometendo o tegumento/regulação térmica do feto/recém-nascido, outros transtornos originados no período perinatal.

Diante das 02 (duas) principais causas de mortalidade infantil nos dois estados,

*Disponível em http://www.datasus.gov.br/cid10/V2008/WebHelp/cap16_3d.htm

foram calculadas as médias e os desvios padrão da idade da mãe, da quantidade de filho vivo, da quantidade de filho morto, do peso, do APGAR1 e do APGAR5 visando apresentar as características da mortalidade infantil com relação às duas principais causas básicas de mortalidade infantil no período neonatal, cujos resultados estão descritos na Tabela 23.

Tabela 23 – Características da mortalidade infantil com relação às duas principais causas de óbitos infantis no período neonatal com relação ao ano de 2016.

	Santa Catarina		**Amapá**	
	Média	**Desvio padrão**	**Média**	**Desvio padrão**
Óbito Infantil - Neonatal Afecções originadas no período perinatal				
Idade da mãe	27,13	7,23	24,70	6,84
Quantidade de filho vivo	1,17	1,22	2,15	1,92
Quantidade de filho morto	0,32	0,69	0,41	0,99
Peso	1.421	1.077	1.471	947
APGAR1	3,98	2,89	4,87	2,98
APGAR5	5,49	3,08	6,51	2,77
Óbito Infantil - Neonatal Malformações congênitas				
Idade da mãe	28,15	7,06	25,41	6,04
Quantidade de filho vivo	1,33	1,29	2,23	1,55
Quantidade de filho morto	0,3	0,69	0,29	1,45
Peso	2.329	938	2.635	788
APGAR1	4,58	2,91	4,47	3,29
APGAR5	5,96	3,24	6,05	3,31

Segundo Rego et al. (2018), as ocorrências de falecimentos de bebês no período perinatal são situações que, muitas vezes, poderiam ter sido evitadas e que de certa forma evidenciam o nível de qualidade dos serviços de saúde que são eventualmente oferecidos às mulheres, seja no pré-natal ou no parto. Analisando as informações com relação às afecções originárias no período perinatal descritas na Tabela 23, constata-se que o campo peso apresenta, de forma mais significativa, uma diferenciação entre a classe de VIVO descrita na Tabela 18 em comparação com a classe ÓBITO INFANTIL - Neonatal descrita na referida Tabela 23. Comparando as duas classes destas Tabelas, pode-se constatar que os pesos dos bebês na classe VIVO nos dois estados foram, respectivamente, de 3.231 gramas e 3.212 gramas, ambos com o mesmo desvio padrão de 546 e 534 gramas, respectivamente. Ao considerar os dados dos bebês que vieram a óbito tendo como causa básica afecções originadas no período perinatal, verificou-se que os pesos dos bebês nos estados de SC e do AP foram respectivamente de 1.421 gramas (desvio padrão de 1.077 gramas) e 1.471 gramas (desvio padrão de 947 gramas). Desta forma, pode-se considerar que os bebês que nascerem e forem acometidos de alguma forma por afecções originárias no período

perinatal e tiverem o peso igual ou inferior aos valores de 2.498 gramas em Santa Catarina e de 2.418 gramas no Amapá, fatalmente terão um grande risco de falecerem antes de completar um ano de vida.

A Tabela 24 descreve informações relativas ao aspecto da escolaridade da mãe com relação às duas principais causas de óbitos infantis no período neonatal para os dois estados avaliados.

Tabela 24 – Escolaridade da mãe com relação às duas principais causas de óbitos infantis no período neonatal com relação ao ano de 2016.

	Santa Catarina		**Amapá**	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito Infantil - Neonatal Afecções originadas no período perinatal Aspecto: Escolaridade da mãe				
Não informado	8	2,75%	10	10,10%
Nenhuma	5	1,72%	1	1,01%
De 1 a 3 anos	11	3,79%	4	4,04%
De 4 a 7 anos	60	20,68%	32	32,32%
De 8 a 11 anos	137	47,24%	40	40,40%
De 12 ou mais anos	69	23,79%	12	12,12%
Óbito Infantil - Neonatal Malformações congênitas Aspecto: Escolaridade da mãe				
Não informado	1	1,11%	3	17,64%
Nenhuma	1	1,11%	1	5,88%
De 1 a 3 anos	6	6,66%	1	5,88%
De 4 a 7 anos	22	24,44%	3	17,64%
De 8 a 11 anos	38	42,22%	7	41,17%
De 12 ou mais anos	22	24,44%	2	11,76%

Analisando a Tabela 24, foi possível constatar em SC que os percentuais de mães com escolaridade de 8 a 11 anos e de 12 ou mais anos é maior na classe de bebês VIVO conforme descrito na Tabela 19, correspondendo à 85,30%, quando comparado com os bebês que faleceram no período neonatal tendo como causas básicas: afecções originadas no período perinatal corresponde à 71,03% e malformações congênitas, com 66,66%.

Com relação à gestação, foi realizado também um levantamento estatístico, descrito na Tabela 25. Assim, esta tabela apresenta informações relativas ao aspecto da gestação com relação às duas principais causas de óbitos infantis no período neonatal nos dois estados. Verifica-se que há diferenças significativas entre os bebês que faleceram no período neonatal devido às afecções originadas no período perinatal em comparação com os bebês que faleceram no mesmo período devido às malformações congênitas.

Em linhas gerais, uma das principais diferenças está no fato de que o percentual de bebês que tiveram uma gestação de 37 a 41 semanas e que faleceram no período neonatal

Tabela 25 – Gestação com relação às duas principais causas de óbitos infantis no período neonatal com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito Infantil - Neonatal Afecções originadas no período perinatal Aspecto: Gestação				
Não informado	4	1,37%	10	10,10%
Menos de 22 semanas	10	3,44%	14	14,14%
De 22 a 27 semanas	139	47,93%	27	27,27%
De 28 a 31 semanas	48	16,55%	23	23,23%
De 32 a 36 semanas	38	13,10%	7	7,07%
De 37 a 41 semanas	49	16,89%	18	18,18%
De 42 a mais semanas	2	0,68%	0	0,0%
Óbito Infantil - Neonatal Malformações congênitas Aspecto: Gestação				
Não informado	2	2,22%	0	0,0%
Menos de 22 semanas	0	0,0%	1	5,88%
De 22 a 27 semanas	8	8,88%	0	0,0%
De 28 a 31 semanas	10	11,11%	3	17,64%
De 32 a 36 semanas	23	25,55%	4	23,52%
De 37 a 41 semanas	47	52,22%	9	52,94%
De 42 a mais semanas	0	0,0%	0	0,0%

devido às afecções que foram originárias no período perinatal foram de 16,89% e de 18,18%, respectivamente. Já o percentual de bebês que tiveram uma gestação de 37 a 41 semanas e que faleceram no período neonatal devido às malformações congênitas foram de 52,22% e de 52,94%, respectivamente. Além disto, o percentual de bebês que tiveram uma gestação de 37 a 41 semanas e que sobreviveram após um ano de vida (VIVO) foram de 86,84% e de 79,11%, respectivamente, conforme apresentado na Tabela 20.

Uma outra diferença é que o percentual de bebês que tiveram uma gestação com menos de 22 semanas até 36 semanas e que faleceram no período neonatal devido às afecções originadas no período perinatal foram de 81,02% e de 71,71%, respectivamente. Já o percentual de bebês que tiveram uma gestação com menos de 22 semanas até 36 semanas e que faleceram no período neonatal devido às malformações congênitas foram de 45,54% e de 47,04%, respectivamente. Além disto, o percentual de bebês que tiveram uma gestação com menos de 22 semanas até 36 semanas e que sobreviveram após um ano de vida (VIVO) foram de 10,85% e de 12,31%, respectivamente, conforme apresentado na Tabela 20. Assim, constata-se que uma boa parte dos bebês que faleceram no período neonatal acometidos por afecções no período perinatal nasceram de uma forma prematura.

Agora, enfatizando o período pós-neonatal e diante das 02 (duas) principais causas

de mortalidade infantil nestes dois estados, foram calculadas as médias e os desvios padrão da idade da mãe, da quantidade de filho vivo, da quantidade de filho morto, do peso, do APGAR1 e do APGAR5 (Veja Tabela 26).

Tabela 26 – Características da mortalidade infantil com relação às principais causas de óbitos infantis no período pós-neonatal com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Média	**Desvio padrão**	**Média**	**Desvio padrão**
Óbito Infantil - Pós-Neonatal Afecções originadas no período perinatal				
Idade da mãe	25,41	6,04	24,50	5,46
Quantidade de filho vivo	2,23	1,55	1,83	1,21
Quantidade de filho morto	0,29	0,45	0,25	0,43
Peso	2.635	788	2.013	887
APGAR1	4,47	3,29	7,25	2,34
APGAR5	6,05	3,31	8,58	1,44
Óbito Infantil - Pós-Neonatal Malformações congênitas				
Idade da mãe	26,18	6,09	28,20	5,70
Quantidade de filho vivo	1,43	1,19	3	2,09
Quantidade de filho morto	0,40	0,78	1,8	2,71
Peso	2.495	798	2.240	720
APGAR1	6,62	2,20	8,60	0,48
APGAR5	8,16	1,42	9,6	0,48
Óbito Infantil - Pós-Neonatal Demais causas de óbito infantil				
Idade da mãe	26,76	6,15	25,68	6,89
Quantidade de filho vivo	1,60	1,49	2,78	1,67
Quantidade de filho morto	0,20	0,50	0,36	0,66
Peso	2.768	961	3.042	655
APGAR1	7,34	2,01	8,52	0,59
APGAR5	8,76	1,30	9,47	0,59

Com relação às informações sobre os bebês que faleceram no período pós-neonatal pelas demais causas de óbito infantil descritas na Tabela 26, constata-se que os campos peso, APGAR 1 e APGAR 5 apresentam valores satisfatórios em boa parte dos casos, inclusive quando são considerados os desvios padrão dos respectivos campos, mas a condição da saúde do bebê possivelmente foi afetada por algum tipo de fator externo da própria vida, que acabou levando o mesmo a óbito.

A Tabela 27 descreve informações relativas ao aspecto da escolaridade da mãe com relação às duas principais causas de óbitos infantis no período pós-neonatal.

Analisando os resultados relativos aos períodos pós-neonatal, descritos na Tabela 27, e neonatal, descritos na Tabela 24, verifica-se que a maior quantidade de óbitos infantis causados por afecções no período perinatal e por malformações congênitas ocorreram no

Tabela 27 – Escolaridade da mãe com relação às principais causas de óbitos infantis no período pós-neonatal com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito Infantil - Pós-Neonatal Afecções originadas no período perinatal Aspecto: Escolaridade da mãe				
Não informado	1	2,94%	0	0,0%
Nenhuma	0	0,0%	0	0,0%
De 1 a 3 anos	0	0,0%	1	8,33%
De 4 a 7 anos	7	20,58%	3	25%
De 8 a 11 anos	22	64,70%	5	41,66%
De 12 ou mais anos	4	11,76%	3	25%
Óbito Infantil - Pós-Neonatal Malformações congênitas Aspecto: Escolaridade da mãe				
Não informado	0	0,0%	0	0,0%
Nenhuma	1	2,70%	0	0,0%
De 1 a 3 anos	0	0,0%	1	20%
De 4 a 7 anos	7	18,91%	0	0,0%
De 8 a 11 anos	20	54,05%	4	80%
De 12 ou mais anos	9	24,32%	0	0,0%
Óbito Infantil - Pós-Neonatal Demais causas de óbito infantil Aspecto: Escolaridade da mãe				
Não informado	0	0,0%	0	0,0%
Nenhuma	3	6,97%	0	0,0%
De 1 a 3 anos	1	2,32%	5	26,31%
De 4 a 7 anos	10	23,25%	10	52,63%
De 8 a 11 anos	23	53,48%	4	21,05%
De 12 ou mais anos	6	13,95%	0	0,0%

período neonatal, nos dois estados avaliados.

A Tabela 28 descreve informações relativas ao aspecto da gestação com relação às principais causas de óbitos infantis no período pós-neonatal nos dois estados.

Os resultados obtidos neste trabalho corroboram com o trabalho de Oliveira (2001), que aponta que o peso do bebê, o nível de APGAR (exame que avalia o nível de adaptação do bebê à vida fora do útero) do primeiro e quinto minuto de vida são fatores importantes para a predição de mortalidade infantil no estado de Santa Catarina.

Em linhas gerais, um dos aspectos descritos na Tabela 28 se assemelha com a Tabela 25, devido ao fato de que um percentual expressivo de bebês que nasceram de uma forma prematura acabaram falecendo no período neonatal e no período pós-neonatal devido a afecções no período perinatal.

Tabela 28 – Gestação com relação às principais causas de óbitos infantis no período pós-neonatal com relação ao ano de 2016.

	Santa Catarina		Amapá	
	Quantidade	Porcentagem	Quantidade	Porcentagem
Óbito Infantil - Pós-Neonatal Afecções originadas no período perinatal Aspecto: Gestação				
Não informado	2	5,88%	0	0,0%
Menos de 22 semanas	1	2,94%	1	8,33%
De 22 a 27 semanas	10	29,41%	2	16,66%
De 28 a 31 semanas	5	14,70%	3	25%
De 32 a 36 semanas	8	23,52%	3	25%
De 37 a 41 semanas	8	23,52%	3	25%
De 42 a mais semanas	0	0,0%	0	0,0%
Óbito Infantil - Pós-Neonatal Malformações congênitas Aspecto: Gestação				
Não informado	1	2,70%	1	20%
Menos de 22 semanas	0	0,0%	1	20%
De 22 a 27 semanas	2	5,40%	0	0,0%
De 28 a 31 semanas	2	5,40%	0	0,0%
De 32 a 36 semanas	8	21,62%	1	20%
De 37 a 41 semanas	23	62,16%	2	40%
De 42 a mais semanas	1	2,70%	0	0,0%
Óbito Infantil - Pós-Neonatal Demais causas de óbito infantil Aspecto: Gestação				
Não informado	3	6,97%	3	15,78%
Menos de 22 semanas	1	2,32%	2	10,52%
De 22 a 27 semanas	1	2,32%	0	0,0%
De 28 a 31 semanas	5	11,62%	1	5,26%
De 32 a 36 semanas	5	11,62%	0	0,0%
De 37 a 41 semanas	28	65,11%	13	68,42%
De 42 a mais semanas	0	0,0%	0	0,0%

5.2.1 Avaliação da qualidade do teste

Finalmente, é importante salientar que, após a análise da qualidade do modelo gerado, foram avaliadas também as sequências que não foram utilizadas na construção do modelo, ou seja, aquelas que foram descartadas no processo de balanceamento.

Notou-se que a maioria das instâncias foram corretamente classificadas e, sendo assim, os modelos gerados pelos algoritmos C4.5, RIPPER, Random Forest, SVM e RNA são satisfatórios, já que todos conseguiram obter, em média, acertos de 89,07% para o estado do Amapá e de 91,24% para o estado de Santa Catarina, e conseguem fazer com uma boa predição relacionada à caracterização da mortalidade infantil nestes dois estados.

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, buscou-se caracterizar, por meio de regras, o problema da mortalidade infantil em 02 (dois) estados brasileiros: Santa Catarina e Amapá com relação ao ano de 2016. Para isso, foram utilizados os algoritmos C4.5 com o plugin VTJ48 e RIPPER, que representa o conhecimento adquirido pelo modelo explicitamente por meio de árvore e regras.

Além destes dois métodos, foi investigado também o comportamento dos algoritmos Random Forest, SVM e RNA no mesmo contexto. O objetivo foi verificar como estes métodos, que não explicitam diretamente as regras encontradas, mas que normalmente oferecem bons resultados, classificariam as classes 'Vivo' e 'Óbito Infantil'. Verificou-se que todos os métodos investigados tiveram um comportamento similar para o contexto analisado.

Quanto aos atributos mais relevantes na classificação de instâncias das duas classes, baseados nos rankings gerados pelo algoritmo Random Forest, foram os seguintes: peso, idade da mãe, gestação, quantidade de filho vivo, APGAR 1, APGAR 5 e escolaridade da mãe. A partir destes atributos, foram realizados levantamentos estatísticos em ambos os estados.

A mortalidade infantil nos estados de Santa Catarina e do Amapá em 2016 é caracterizada no período neonatal por 02 (duas) principais causas de óbito: afecções originadas no período perinatal e malformações congênitas que, somadas, representam 98,19% em SC e 100,00% no AP do total das causas básicas de óbito no período neonatal. Com base na análise dos dados, os bebês com afecções originárias no período perinatal em SC com peso médio de 1.541 gramas (desvio padrão de 1.077 gramas) e no AP com peso médio de 1.471 gramas (desvio padrão de 947 gramas) correm um sério risco de não completarem um ano de vida. Constatou-se que existe uma diferença significativa de peso e APGAR do bebê que completa um ano de vida para o bebê acometido por malformação congênita, mas a maior diferença de peso e do APGAR está entre os bebês com alguma afecção perinatal quando comparados com bebês que completaram um ano de vida.

Com relação ao período pós-neonatal, constatou-se que existe uma variedade de causas básicas de óbito em ambos os estados. Analisando os dados, foi possível constatar que somados os casos de afecções originadas no período perinatal e de malformações congênitas representaram 62,31% em SC e 47,21% no AP do total das causas básicas de óbito no período pós-neonatal. Desta forma, 37,69% em SC e 52,79% no AP são outras causas de óbito que podem eventualmente afetar bebês saudáveis, mesmo que os respectivos bebês tenham boas características para completar um ano de vida, mas foram acometidos por outras enfermidades no curso inicial da vida que, nestes casos, levaram ao

óbito do bebê. Isto, em teoria, pode ser uma possível explicação, pelo fato de que os óbitos no período pós-neonatal apresentam valores de peso e de APGAR mais elevados quando comparados com os perfis dos bebês mortos no período neonatal. É importante salientar novamente que uma parcela significativa de bebês que faleceram antes de completar um ano de vida tiveram em comum o fato de terem nascido de forma prematura, representando os percentuais de óbito no período neonatal de 71,56% em SC e de 68,09% no AP e no período pós-neonatal de 42,09% em SC e de 38,88% no AP.

A partir das políticas públicas de saúde e diante dos resultados descritos neste trabalho, sugere-se que o Projeto Rede Cegonha e inclusive o PBF possam incluir ações/condicionantes de saúde com ênfase no acompanhamento da mãe com alguma afecção no período perinatal e também o acompanhamento ainda mais intensivo da saúde dos bebês nos seguintes casos:

- De bebês prematuros;
- De bebês acometidos por algum tipo de afecção originária do período perinatal.;
- De bebês acometidos por algum tipo de malformação congênita;
- De bebês que obtiverem uma nota baixa no APGAR;
- De bebês que tiverem um baixo peso ao nascer;

Com relação ao Projeto Rede Cegonha, é importante salientar que ele abrange a participação das três esferas (federal, estadual e municipal). Inicialmente, pode-se constatar que Santa Catarina foi pioneira em implantar a Rede Cegonha com abrangência em todo o estado. Assim, possivelmente, este fator, dentre outros, pode ter colaborado para que a taxa de mortalidade infantil em SC fosse a menor do Brasil em comparação com os demais estados. Com relação ao estado do Amapá (que possuiu a maior TMI em 2016), sugere-se que o governo do AP possa verificar se o Projeto Rede Cegonha foi implantado adequadamente em todos os municípios e se cada município está de fato participando e colaborando efetivamente, além da participação do governo estadual e do governo federal, para melhoria das condições de saúde das mães e dos bebês, visando a redução da mortalidade infantil.

Como proposta de trabalhos futuros, temos as seguintes estratégias: 1) a aplicação dos mesmos procedimentos adotados neste trabalho em todos os estados do território brasileiro e em todos os anos disponibilizados pelo DATASUS visando caracterizar a MI em todos os demais estados e nortear os investimentos públicos para a diminuição da TMI no país; 2) analisar o perfil de mortalidade em cada mesorregião e levantar também outras políticas públicas de saúde adotadas nos dois estados avaliados visando propor melhorias nas ações de saúde das gestantes e dos bebês; 3) Finalmente, uma possibilidade viável de melhorar os resultados deste trabalho é ampliar o número de campos/atributos analisados, pois neste trabalho foi possível analisar somente 17 (dezessete) campos/atributos, devido

ao fato de que os dicionários de dados disponibilizados pelo DATASUS não englobam todos os campos/atributos existentes no SINASC/SIM. Para ampliar, uma das alternativas é obter do DATASUS novos dicionários de dados completos que contenham todos os campos/atributos que existem no SINASC/SIM. Assim, ao se analisar uma quantidade maior de atributos, existe a possibilidade de aprimorar os resultados obtidos, melhorando assim a caracterização da mortalidade infantil brasileira.

REFERÊNCIAS

PNUD; IPEA; FJP. **O índice de desenvolvimento humano municipal brasileiro**: Série atlas do desenvolvimento humano no brasil 2013. Brasília, Brasil: Bookman, 2013. 96 p. ISBN 9788578111717.

APGAR, V. et al. **Evaluation of the newborn infant-second report**. Journal of the American Medical Association, American Medical Association, v. 168, n. 15, p. 1985–1988, Dec. 1958. Disponível em: <https://jamanetwork.com/journals/jama/article-abstract/325068>.

BLACK, R. E. et al. **Global, regional, and national causes of child mortality in 2008: a systematic analysis**. The Lancet, Elsevier, v. 375, n. 9730, p. 1969–1987, 2010. ISSN 0140-6736. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140673610605491>.

BONESSO, D. **Estimação dos parâmetros do kernel em um classificador SVM na classificação de imagens hiperespectrais em uma abordagem multiclasse**. 108 f. Dissertação (Mestrado em Sensoriamento Remoto) — Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Porto Alegre, RS, Brasil.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. **A Training Algorithm for Optimal Margin Classifiers**. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. New York, NY, USA: ACM, 1992. (COLT '92), p. 144–152. ISBN 0-89791-497-X. Disponível em: <http://doi.acm.org/10.1145/130385.130401>.

BRASIL, M. d. S. **Manual de Vigilância do Óbito Infantil e Fetal e do Comitê de Prevenção do Óbito Infantil e Fetal**: Série a. normas e manuais técnicos. Brasília, DF, Brasil, 2009. 96 p. Disponível em: <http://bvsms.saude.gov.br/bvs/publicacoes/manual_obito_infantil_fetal_2ed.pdf>.

BREIMAN, L. **Bagging predictors**. Machine Learning, Kluwer Academic Publishers, v. 24, n. 2, p. 123–140, Ago. 1996. ISSN 1573-0565. Disponível em: <https://doi.org/10.1007/BF00058655>.

BREIMAN, L. **Random Forests**. Machine Learning, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, Out. 2001. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1010933404324>.

CHANG, C.-C.; LIN, C.-J. **LIBSVM: A Library for Support Vector Machines**. ACM Trans. Intell. Syst. Technol., ACM, New York, NY, USA, v. 2, n. 3, p. 27:1–27:27, May. 2011. ISSN 2157-6904. Disponível em: <http://doi.acm.org/10.1145/1961189.1961199>.

CHEN, A.; OSTER, E.; WILLIAMS, H. **Why Is Infant Mortality Higher in the United States Than in Europe?** American Economic Journal: Economic Policy, American Economic Association, v. 8, n. 2, p. 89–124, Mai. 2016. ISSN 1945-774X. Disponível em: <http://www.aeaweb.org/articles?id=10.1257/pol.20140224>.

CHOUDHRY, F.; QAMAR, U.; CHAUDHRY, M. **Rule based inference engine to forecast the prevalence of congenital malformations in live births**. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Takamatsu, Japan: IEEE, 2015. p. 1–7. ISBN 978-1-4799-8676-7. Disponível em: <https://ieeexplore.ieee.org/document/7176279>.

COHEN, W. W. **Fast Effective Rule Induction**. In: Proceedings of the Twelfth International Conference on International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (ICML'95), p. 115–123. ISBN 1-55860-377-8.

CORTES, C.; VAPNIK, V. **Support-Vector Networks**. Machine Learning, Kluwer Academic Publishers-Plenum Publishers, v. 20, n. 3, p. 273–297, Set. 1995. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1022627411411>.

COSTA, D. M. M. d. **Ensemble baseado em métodos de Kernel para reconhecimento biométrico multimodal**. 145 f. Dissertação (Mestrado em Ciências) — Universidade de São Paulo, Escola de Artes, Ciências e Humanidades, São Paulo, SP, Brasil.

DESPOTOVIC, D. et al. **A Machine Learning Approach for an Early Prediction of Preterm Delivery**. In: 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY). Subotica, Serbia: IEEE, 2018. p. 265–270. ISSN 1949-0488. Disponível em: <https://ieeexplore.ieee.org/document/8524818>.

DUIM, E.; NAMPO, F. K.; SOUZA, S. **Determinantes do escore de apgar e mortalidade neonatal em Foz do Iguaçu-PR - resultados preliminares**. Anais do VI Encontro de Iniciação Científica e II Encontro Anual de Iniciação ao Desenvolvimento Tecnológico e Inovação – EICTI, Universidade Federal da Integração Latino-Americana (Unila), Foz do Iguaçu - Parána - Brasil, Out. 2017. Disponível em: <http://dspace.unila.edu.br/123456789/3381>.

FACELI, K. et al. **Inteligência artificial : uma abordagem de aprendizado de máquina**. Rio de Janeiro, RJ, Brasil: LTC, 2011. 394 p. ISBN 9788521618805.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **The KDD Process for Extracting Useful Knowledge from Volumes of Data**. Commun. ACM, ACM, New York, NY, USA, v. 39, n. 11, p. 27–34, Nov. 1996. ISSN 0001-0782. Disponível em: <http://doi.acm.org/10.1145/240455.240464>.

FERRARI, R. A. P.; BERTOLOZZI, M. R. **Mortalidade pós-neonatal no território brasileiro: uma revisão da literatura**. Revista da Escola de Enfermagem da USP, Universidade de São Paulo, São Paulo, SP, Brasil, v. 46, n. 5, 2012. Disponível em: <https://www.revistas.usp.br/reeusp/article/view/48145/51974>.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The Elements of Statistical Learning**: Data mining, inference and prediction. New York, NY, USA: Springer, 2009. ISBN 978-0-387-84857-0.

GOLDANI, M. Z. et al. **Infant mortality rates according to socioeconomic status in a Brazilian city**. Revista de Saúde Pública, Scielo, São Paulo, SP, Brasil, v. 35, n. 3, p. 256–261, Jun. 2001. ISSN 0034-8910. Disponível em: <http://www.scielo.br/pdf/rsp/v35n3/5010.pdf>.

GOLDBERGER, A. L. et al. Physiobank, physiotoolkit, and physionet. Circulation, v. 101, n. 23, p. e215–20, 2000. Disponível em: <https://app.dimensions.ai/details/publication/pub.1032570273 and http://cps-www.bu.edu/hes/articles/gaghimmmps00.pdf>.

HAYKIN, S. **Redes Neurais - 2ed.** Bookman, 2001. Disponível em: <https://books.google.com.br/books?id=lBp0X5qfyjUC>. ISBN 9788573077186.

HERNANDEZ, A. R. et al. **Análise de tendências das taxas de mortalidade infantil e de seus fatores de risco na cidade de Porto Alegre, Rio Grande do Sul, Brasil, no período de 1996 a 2008**. Cadernos de Saúde Pública, Scielo, Rio de Janeiro, RJ, Brasil, v. 27, n. 11, p. 2188–2196, Nov. 2011. ISSN 0102-311X. Disponível em: <http://www.scielo.br/pdf/csp/v27n11/12.pdf>.

IBGE, I. B. de Geografia e E. **Tábua completa de mortalidade para o Brasil – 2017: Breve análise da evolução da mortalidade no Brasil**. In: . Rio de Janeiro, RJ, Brasil: Ministério do Planejamento, Desenvolvimento e Gestão, 2018. Disponível em: <ftp://ftp.ibge.gov.br/Tabuas$_C ompletas_d e_M ortalidade/Tabuas_C ompletas_d e_M ortalidade_2$017/*tabua*

KHOSHGOFTAAR, X. M.; GOLAWALA, M.; HULSE, J. V. **An Empirical Study of Learning from Imbalanced Data Using Random Forest**. In: International Conference on Tools with Artificial Intelligence (ICTAI 2007). IEEE, 2007. p. 310–317. ISBN 978-0-7695-3015-4. Disponível em: <https://ieeexplore.ieee.org/abstract/document/4410397>.

KITSANTAS, P.; HOLLANDER, M.; LI, L. **Using Classification Trees to Assess Low Birth Weight Outcomes**. Artificial Intelligence in Medicine, Elsevier Science Publishers Ltd., Essex, UK, v. 38, n. 3, p. 275–289, Nov. 2006. ISSN 0933-3657. Disponível em: <http://dx.doi.org/10.1016/j.artmed.2006.03.008>.

KOHAVI, R. **A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection**. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1-55860-363-8. Disponível em: <http://dl.acm.org/citation.cfm?id=1643031.1643047>.

LEAL, M. d. C. et al. **Saúde reprodutiva, materna, neonatal e infantil nos 30 anos do Sistema Único de Saúde (SUS)**. Ciência e Saúde Coletiva, Scielo, Rio de Janeiro, RJ, Brasil, v. 23, p. 1915–1928, Jun. 2018. ISSN 1413-8123. Disponível em: <http://www.scielo.br/pdf/csc/v23n6/1413-8123-csc-23-06-1915.pdf>.

LI, M.; VITáNYI, P. M. Chapter 4 - kolmogorov complexity and its applications. In: LEEUWEN, J. V. (Ed.). Algorithms and Complexity. Amsterdam: Elsevier, 1990, (Handbook of Theoretical Computer Science). p. 187 – 254. ISBN 978-0-444-88071-0.

LIBRALON, G. L. **Investigação de combinações de técnicas de detecção de ruído para dados de expressão gênica**. 99 p. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brasil.

LORENZ, J. et al. **Infant mortality in the United States**. Journal of Perinatology, Springer Nature Publishing Group, v. 36, n. 10, p. 797–801, Abr. 2016. ISSN 1476-5543. Disponível em: <https://www.nature.com/articles/jp201663>.

MARINS, M. A. **Classificação de Falhas em Máquinas Rotativas Utilizando Métodos de Similaridade e Random Forest**. Rio de Janeiro, RJ, Brasil: Escola Politécnica, Universidade Federal do Rio de Janeiro, Set. 2016.

MARTINS, P. C. R.; PONTES, E. R. J. C.; HIGA, L. T. **Convergência entre as Taxas de Mortalidade Infantil e os Índices de Desenvolvimento Humano no Brasil no período de 2000 a 2010.** Interações (Campo Grande), Scielo, Campo Grande, MS, Brasil, v. 19, n. 2, p. 291–303, Jun. 2018. ISSN 1518-7012. Disponível em: <http://www.scielo.br/pdf/inter/v19n2/1518-7012-inter-19-02-0291.pdf>.

MOREIRA, L. M. d. C. et al. Políticas públicas voltadas para a redução da mortalidade infantil: uma história de desafios. Rev Med Minas Gerais, v. 22, n. Supl 7, p. S48–S55, 2012.

OLIVEIRA, I. T. C. d. **Aplicação de data mining na busca de um modelo de prevenção da mortalidade infantil**. 95 f. Dissertação (Mestrado em Engenharia de Produção) — Universidade Federal de Santa Catarina - Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, SC, Brasil.

OLIVEIRA, T. G. d. et al. **Escore de Apgar e mortalidade neonatal em um hospital localizado na zona sul do município de São Paulo**. Einstein (São Paulo), Scielo, São Paulo, Brasil, v. 10, n. 1, p. 22–28, Mar. 2012. ISSN 1679-4508. Disponível em: <http://www.scielo.br/scielo.php?script=sci$_a rttextpid = S1679 - 45082012000100006 nrm = iso$>.

PLATT, J. **Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines**. Technical Report MSR-TR-98-14, Microsoft Research, Abr. 1998.

PRATI, R. C. **Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos**. 191 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação São Carlos, São Carlos, SP, Brasil. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-01092006-155445/pt-br.php>.

QUILICI-GONZALEZ, J. A.; ZAMPIROLLI, F. d. A. **Sistemas Inteligentes e Mineração de Dados**. Santo André, SP, Brasil: Triunfal Gráfica e Editora, 2014. 148 p. ISBN 978-85-61175-38-2.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0.

RAMOS, R. et al. **Using predictive classifiers to prevent infant mortality in the Brazilian northeast**. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). Dalian, China: IEEE, 2017. p. 1–6. ISBN 978-1-5090-6704-6. Disponível em: <https://ieeexplore.ieee.org/document/8210811>.

REGO, M. G. d. S. et al. **Óbitos perinatais evitáveis por intervenções do Sistema Único de Saúde do Brasil**. Revista Gaúcha de Enfermagem, Scielo, Porto Alegre, Rio Grande do Sul, Brasil, v. 39, Jul. 2018. ISSN 1983-1447. Disponível em: <http://www.scielo.br/scielo.php?script=sci$_a$rttextpid = S1983 − 14472018000100414nrm = iso>.

RIPSA, R. I. de Informação para a S. **Indicadores básicos para a saúde no Brasil**: conceitos e aplicações. Brasília, DF, Brasil: Organização Pan-Americana da Saúde, 2008. 349 p. Disponível em: <http://tabnet.datasus.gov.br/tabdata/livroidb/2ed/indicadores.pdf>. ISBN 978-85–87943–65-1.

SARTORELLI, A. P. et al. **Fatores que contribuem para a mortalidade infantil utilizando a mineração de dados**. Revista Saúde e Pesquisa, Unicesumar, Maringá, PR, Brasil, v. 10, n. 1, p. 33–41, Abr. 2017. ISSN 2176-9206. Disponível em: <http://periodicos.unicesumar.edu.br/index.php/saudpesq/article/view/5879>.

SILVA, E. S. d. A. d.; PAES, N. A. **Programa Bolsa Família e a redução da mortalidade infantil nos municípios do Semiárido brasileiro**. Ciência e Saúde Coletiva, scielo, v. 24, p. 623 – 630, 02 2019. ISSN 1413-8123. Disponível em: <http://www.scielo.br/scielo.php?script=sci$_a$rttextpid = S1413 − 81232019000200623nrm = iso>.

STIGLIC, G. et al. **Comprehensive decision tree models in bioinformatics**. PloS one, Public Library of Science, Rockville Pike, Bethesda MD, USA, v. 7, n. 3, p. e33812, Mar. 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/22479449>.

TEIXEIRA, R. d. S.; COLMANETTI, J. B. D.; CARVALHO, D. R. **Post-processing of classifiers - KDD**. Iberoamerican Journal of Applied Computing, Universidade Estadual de Ponta Grossa - UEPG, Ponta Grossa, PR, Brazil, v. 5, n. 1, Apr. 2015. ISSN 2237-4523. Disponível em: <http://www.revistas2.uepg.br/index.php/ijac/article/view/8430>.

TOSCANO, G.; HOSSAIN, G. **Predicting the Effect of Parental Education and Income on Infant Mortality Through Statistical Learning**. In: 2018 1st International Conference on Data Intelligence and Security (ICDIS). South Padre Island, TX, USA: IEEE, 2018. p. 99–102. ISBN 978-1-5386-5762-1. Disponível em: <https://ieeexplore.ieee.org/document/8367746>.

UNICEF et al. **Levels and Trends in Child Mortality Report 2018**. UNICEF, New York, USA, Set. 2018. Disponível em: <https://data.unicef.org/wp-content/uploads/2018/09/UN-IGME-Child-Mortality-Report-2018.pdf

VIANNA, R. C. X. F. et al. **Mineração de dados e características da mortalidade infantil**. Cadernos de Saúde Pública, Scielo, Rio de Janeiro, RJ, Brasil, v. 26, n. 3, p. 535–542, Mar. 2010. ISSN 0102-311X. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2010000300011&nrm=iso>.