



UNIVERSIDADE FEDERAL DO CEARÁ – UFC
FACULDADE DE MEDICINA
DEPARTAMENTO DE SAÚDE COMUNITÁRIA – DSC
PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE COLETIVA
DOUTORADO ACADÊMICO EM SAÚDE COLETIVA

WELLINGTON SOUSA AGUIAR

**DESENVOLVIMENTO DE MODELOS PREDITIVOS DE MORTALIDADE
INFANTIL COM BASE EM INTELIGÊNCIA ARTIFICIAL NO ESTADO DO
CEARÁ**

FORTALEZA-CE
2019

WELLINGTON SOUSA AGUIAR

DESENVOLVIMENTO DE MODELOS PREDITIVOS DA MORTALIDADE
INFANTIL COM BASE EM INTELIGÊNCIA ARTIFICIAL NO ESTADO DO
CEARÁ

Tese apresentada ao Curso de Doutorado em Saúde Coletiva do Departamento de Saúde Comunitária da Universidade Federal do Ceará - UFC, como parte dos requisitos para obtenção do título de Doutor em Saúde Coletiva.

Área de concentração: Avaliação de Políticas e Tecnologias em Saúde.

Orientador: Prof. Dr. Álvaro Jorge Madeiro Leite

Coorientador: Prof. Dr. Francisco Herlânio Costa Carvalho

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A233d Aguiar, Wellington Sousa.

Desenvolvimento de modelos preditivos de mortalidade infantil com base em Inteligência Artificial no estado do Ceará : Aplicação de Inteligência Artificial na Saúde Pública / Wellington Sousa Aguiar. – 2019.

163 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Faculdade de Medicina, Programa de Pós-Graduação em Saúde Coletiva, Fortaleza, 2019.

Orientação: Prof. Dr. Álvaro Jorge Madeiro Leite.

Coorientação: Prof. Dr. Francisco Herlânio Costa Carvalho.

1. Saúde Pública. 2. Mortalidade Infantil. 3. Inteligência Artificial. 4. Algoritmos. 5. Redes Neurais.
I. Título.

CDD 362.1

WELLINGTON SOUSA AGUIAR

DESENVOLVIMENTO DE MODELOS PREDITIVOS DA MORTALIDADE
INFANTIL COM BASE EM INTELIGÊNCIA ARTIFICIAL NO ESTADO DO
CEARÁ

Tese de doutorado apresentada ao Curso de Doutorado em Saúde Coletiva do Departamento de Saúde Comunitária da Universidade Federal do Ceará - UFC, como parte dos requisitos para obtenção do título de Doutor em Saúde Coletiva. Área de concentração: Avaliação de Políticas e Tecnologias em Saúde.

Orientador: Prof. Dr. Álvaro Jorge Madeiro Leite

Coorientador: Prof. Dr. Francisco Herlânio Costa Carvalho

Aprovado: ____/____/____

BANCA EXAMINADORA

Prof. Dr. Álvaro Jorge Madeiro Leite (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Herlânio Costa Carvalho (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Antonio de Araújo e Souza
Centro Universitário Estácio do Ceará

Prof. Dr. Ricardo Soares Pontes
Universidade Federal do Ceará – UFC

Aos meus pais: Maurício Soares de Aguiar (*in memoriam*) e Maria Sônia Sousa Aguiar, pelo amor incondicional e exemplo de família que me guiarão por toda vida.

AGRADECIMENTOS

A Deus, pela vida, conhecimentos e oportunidades.

A minha maravilhosa família, a base de tudo, aos meus pais e irmãos, que me ensinaram e orientaram o caminho da retidão e da conquista pela luta e esforço. Dedico este trabalho especialmente ao meu pai Maurício Soares de Aguiar e meu irmão Carlos Felix Neto que já nos deixaram, mas estão sempre presentes com seus exemplos.

A minha esposa Regina Cláudia, pelo seu apoio incondicional e respeito a minha dedicação aos estudos. A sua dedicação à nossa família, aos seus pais e ao trabalho, nos ensina a cada dia como viver com dignidade e honrar a família.

Aos meus filhos: Thiago, Daniel, Mariana, Raquel e Rafael, às Minhas noras: Diana e Alice, às minhas netinhas: Cecília e Celina e a todos os netos e netas que ainda virão, que a minha dedicação e esforço em busca desse sonho sirvam de exemplos para suas vidas.

Ao meu orientador Prof. Dr. Álvaro Jorge Madeiro Leite, por acreditar e vislumbrar oportunidades nas tecnologias propostas, orientando e participando de cada fase do projeto possibilitando grandes momentos de aprendizado e troca de experiências.

Ao Prof. Dr. Francisco Herlânio Costa Carvalho, Coorientador, amigo e mais um irmão que a vida me proporcionou, meu maior incentivador para desenvolver e aplicar a tecnologia na saúde pública.

Ao Prof. Dr. Ricardo Soares, Coorientador e mais um amigo que ganhei nessa caminhada, nos mostrou a filosofia para a saúde e para a vida.

Aos amigos e amigas do doutorado que conquistei nessa caminhada, cada um contribuindo com suas experiências e conhecimentos, recebendo um profissional da Computação dentro da Saúde Coletiva com muito carinho e atenção, possibilitando muitas trocas de conhecimentos e experiências.

Aos funcionários da secretaria do Doutorado na UFC pela dedicação e presteza, e aos nossos professores Dra. Larissa Fortunato, Dr. Alberto Novaes, Dra. Maria Lúcia Bosi, Dr. Henrique Alencar, entre outros, pela dedicação e parceira com todos os discentes.

Aos colegas professores: Péricles, Henrique Mota, Francisco Alves, Francisco Araujo, Antônio Carlos, Rejane Cunha, Otávio Frota, Graça Carlo, aos coordenadores Aminadabe Barbosa e Janete Amaral, a Reitora Profa. Ana Flávia e a Pró-reitora Profa. Letícia Adriana, todos do Centro Universitário Estácio do Ceará que sempre apoiaram e incentivaram com suas valiosas orientações.

Aos amigos e a toda diretoria da Odonto System que incentivaram e apoiaram este projeto. Essa conquista é nossa: Prof. Mário Fontenelle, Prof. Felipe Carneiro, Egnaldo Carvalho, Fábio Lima, Pedro Felix, Valclemir Rodrigues, Ramiro dos Santos e toda Equipe de TI (Fortaleza e Salvador), obrigado pelos incentivos e apoio incondicional.

Aos amigos do MADE (Mestrado em Administração) da UNESA-RJ que sempre me incentivaram e tornaram-se grandes amigos: Leandro Herculano, Fred, Paulo Duavy e Eduardo Carpegianni e os professores que nos receberam com tanto carinho e atenção do Rio de Janeiro: Prof. Antônio Augusto Gonçalves (Orientador), Prof. Cláudio Pitassi (Coorientador), Prof. José Geraldo (Coordenador) e demais professores.

A todos os componentes do Grupo de Pesquisa DEG (Data Explorers Group) que pesquisam Bigdata e Inteligência Artificial porque acreditaram no meu sonho de que é possível fazer pesquisa aplicada no Ceará com poucos recursos e boa vontade. Obrigado pela confiança.

E a todos os meus alunos e ex-alunos, a minha missão é tocá-los com algum conhecimento útil que possa transformá-los em profissionais de qualidade e cidadãos dignos e éticos.

“Meditai se só as nações fortes podem fazer ciência ou se é a ciência que as torna fortes”.

Walter Oswaldo Cruz

RESUMO

O Estado do Ceará registrou nos últimos vinte anos muitos avanços no planejamento de políticas públicas de cuidados com a saúde que propiciaram a redução dos índices de mortalidade infantil, mas essa redução parou nos últimos anos ? A tecnologia também evoluiu em vários sentidos, na redução de custos de desenvolvimento de software, de armazenamento e de processamento de dados, na utilização de aplicativos móveis que facilitam o contato rápido com a população, tudo isso permitindo a utilização em larga escala de recursos como: Mineração de dados, Data Mart, BI (*Business Intelligence*), Inteligência Artificial entre outros, capazes de gerar inteligência e conhecimento que apoiem à tomada de decisão estratégica dos responsáveis pelas políticas e planejamento da saúde pública. Essa pesquisa multidisciplinar, envolvendo a Ciência da Computação, a Administração, a Estatística e a Saúde Pública, tem como objetivo desenvolver modelos preditivos de mortalidade infantil usando algoritmos da inteligência artificial e mineração de dados, capazes de gerar conhecimentos até então não percebidos ou não utilizados pela gestão da saúde pública para apoiar suas ações, políticas e planejamentos, visando a redução da mortalidade infantil no estado do Ceará. A metodologia utilizada foi a quantitativa que extrai informações estatísticas relevantes dos algoritmos aplicados, mas apoiada por pesquisas bibliográficas que sustentam as teorias. Essa também é uma pesquisa aplicada, pois seus resultados serão aplicados no mundo real em benefício de nossa comunidade. No Ceará nascem mais de 130 mil crianças por ano e morrem mais de 1.600 crianças antes de completar um ano de vida, TMI médio de 12,3. As principais causas da mortalidade infantil são: baixo peso ao nascer, prematuridade, poucas ou nenhuma consultas pré-natal, pouca idade e baixa escolaridade das mães, entre outras questões. Quando aplicamos milhares de vezes os algoritmos de Redes Neurais em 20 variáveis das características de 8.159 óbitos infantis e de 10.235 nascidos vivos que sobreviveram ao primeiro ano de vida, alcançamos acurácia média de 97,41% (+/- 0,40%) na predição do futuro. A partir desses resultados podemos inferir que é possível prever a mortalidade infantil usando algoritmos de Inteligência Artificial com margem de erro satisfatória, possibilitando o uso dessa tecnologia de baixo custo no apoio as estratégias da saúde pública no Brasil.

Palavras-Chaves: Saúde Pública; Mortalidade Infantil; Inteligência Artificial; Algoritmos.

ABSTRACT

The State of Ceará has registered, in the last twenty years, great progress in the planning of health care public policies that propagated the reduction in infant mortality rates, but has this reduction stopped in recent years ? Technology has also evolved in many ways recently, reducing the costs for software development, data storage and processing, the use of mobile applications that facilitate contact with population, all this allowing the large scale usage of resources such as: Data Mining, Data Mart, BI (Business Intelligence), Artificial Intelligence and others, capable of generating intelligence and knowledge to support strategic decision making for those responsible for public health planning and policies. This multidisciplinary research, involving Computer Science, Administration, Statistics and Public Health, has as its objective to develop child mortality predictive models using data mining and artificial intelligence algorithms, capable of generating data that was not perceived or not utilized until now by public health management to support its decisions, policies and planning, aiming for the infant mortality reduction in the State of Ceará. The methodology used was the quantitative, that extracts relevant statistic information from the applied algorithms, but with bibliographical research in support of theories. This is also an applied research, for its results will be applied in the real world to benefit our community. In Ceará more than 130 thousand children are born each year and more than 1.600 die before reaching one year old, average TMI of 12,3. The main causes of infant mortality are: low weight at birth, prematurity, few or none prenatal consultation, mother's young age and low scolarity, among others. When we apply the Neural Network algorithms thousands of times in 20 characteristic variables from 8.159 infant deaths and 10.235 living births who survived the first year of life, we reach the accuracy of 97,41% (with a margin of error of +/- 0,41%) at predicting the future. From these results we can infer that it's possible to predict infant mortality using Artificial Intelligence algorithms with a satisfactory margin of error, allowing the utilization of this low cost technology to support public health strategies in Brazil.

Key-words: Public health; Child mortality; Artificial intelligence; Algorithms.

RESUMÉ

Au cours des vingt dernières années, l'État de Ceará a réalisé de nombreux progrès dans la planification des politiques de santé publique qui ont permis de réduire les taux de mortalité infantile, mais cette réduction a-t-elle cessé ces dernières années? La technologie a également évolué de plusieurs manières au cours des dernières années, réduisant les coûts de développement de logiciels, de stockage et de traitement de données, ainsi que l'utilisation d'applications mobiles facilitant un contact rapide avec la population, le tout permettant une utilisation à grande échelle. tels que: Data Mining, Data Mart, BI (Business Intelligence), Artificial Intelligence, entre autres, capable de générer de l'intelligence et des connaissances pour soutenir la prise de décision stratégique des décideurs et la planification de la santé publique. Cette recherche multidisciplinaire, impliquant les domaines de l'informatique, de l'administration, des statistiques et de la santé publique, vise à développer des modèles prédictifs de la mortalité infantile à l'aide d'algorithmes d'intelligence artificielle et d'exploration de données, capables de générer des connaissances jusque-là non perçues ou non utilisées par la direction de la santé publique pour soutenir leurs actions, leurs politiques et leur planification, visant à réduire la mortalité infantile dans l'état de Ceará. La méthodologie utilisée était la méthode quantitative qui extrait les informations statistiques pertinentes des algorithmes appliqués, mais avec des recherches bibliographiques à l'appui des théories. C'est aussi une recherche appliquée, car ses résultats seront appliqués dans le monde réel au profit de notre communauté. Au Ceará, plus de 130000 enfants naissent chaque année et plus de 1600 enfants décèdent avant l'âge d'un an, soit un coefficient moyen de 12,3. Les principales causes de mortalité infantile sont notamment: faible poids à la naissance, prématurité, consultations prénatales peu nombreuses ou inexistantes, faible âge et faible niveau d'éducation maternelle. Lorsque nous avons appliqué les algorithmes de réseaux neuronaux des milliers de fois dans 20 variables des caractéristiques de 8159 décès infantiles et de 10235 naissances vivantes ayant survécu à la première année de vie, nous avons atteint une précision moyenne de 97,41% (+/- 0,40%) en prédiction de l'avenir. On peut déduire de ces résultats qu'il est possible de prédire la mortalité infantile en utilisant des algorithmes d'intelligence artificielle avec une marge d'erreur satisfaisante, ce qui permet d'utiliser cette technologie peu coûteuse pour appuyer les stratégies de santé publique au Brésil.

Mots Clés: santé publique; Mortalité infantile; Intelligence artificielle; Algorithmes.

LISTA DE SIGLAS E ABREVIATURAS

AI	Artificial Intelligence
AMPS	Advanced Mobile Phone Service
ANS	Agência Nacional de Saúde
BI	Business Intelligence
BPM	Business Process Management
BPMN	Business Process Model and Notation
CEP	Conselho de Ética e Pesquisa
CRISP-DM	Cross Industry Standard Process for Data Mining
IA	Inteligência Artificial
IEA	Associação internacional de Epidemiologia
KDD	Knowledge Discovery in Databases
KPI	Key Performance Indicator
MER	Modelo de Entidades e Relacionamentos
ML	Machine Learning
MLP	Multi-Layer Perceptron
MMDS	Multichannel Multipoint Distribution Service
ONU	Organização das Nações Unidas
OPS	Operadora de Planos de Saúde
OPAS	Organização Pan-Americana de Saúde
PC	Personal Computer
RAS	Rede de Atenção à Saúde
RC	Rede Cegonha
RIPAS	Rede Interagencial de Informações para a Saúde
RMF	Região Metropolitana de Fortaleza
RMM	Razão de Mortalidade Materna
SIM	Sistema de Informações de Mortalidade
SUS	Sistema Único de Saúde
TIC	Tecnologia da Informação e Comunicação
TMI	Taxa de Mortalidade infantil

LISTA DE FIGURAS

Figura 1	- Um modelo de algoritmo	14
Figura 2	- Relação entre as áreas: Profissionais da Saúde e da Computação	15
Figura 3	- Mapa do estado do Ceará por Macrorregiões de Saúde	18
Figura 4	- Modelo de abrangência da filosofia	20
Figura 5	- Valores da filosofia	21
Figura 6	- Tabela Verdade	25
Figura 7	- Modelo de Funcionalidades do SUS	31
Figura 8	- Métodos epidemiológicos	34
Figura 9	- As ciências da saúde e suas relações complementares	34
Figura 10	- Faixas do IDH (Índice do Desenvolvimento Humano)	38
Figura 11	- Cronologia das Taxas de Mortalidade	43
Figura 12	- A evolução tecnológica computacional	50
Figura 13	- Exemplo de árvore de decisão	54
Figura 14	- Representação da Lógica Clássica e Lógica Fuzzy	55
Figura 15	- Exemplo de grafo, que liga vértices por meio de arestas	56
Figura 16	- Modelo de Neurônio Biológico	58
Figura 17	- Modelo de neurônio artificial de McCulloch e Pitts	60
Figura 18	- Função Sigmoidal	60
Figura 19	- Rede Perceptron de uma camada	61
Figura 20	- Rede MLP (Multi-Layer Perceptron) com mais de uma camada	62
Figura 21	- Rede Neural Artificial de duas camadas	62
Figura 22	- Principais técnicas e algoritmos de aprendizado de máquina	63
Figura 23	- Processo de geração do BI	67
Figura 24	- Etapas operacionais do processo KDD	69
Figura 25	- Atividades das etapas do processo KDD	70
Figura 26	- O processo KDD	70
Figura 27	- Processo para obtenção de certificado de um curso	71

Figura 28 – Mapa da Região Metropolitana de Fortaleza por população bruta	76
Figura 29 - MER (Modelo de Entidades e relacionamentos)	82
Figura 30 - Georreferenciamento do estado do Ceará por TMI em 2017	93
Figura 31 - Fases para geração dos modelos preditivos	99
Figura 32 - Técnica de <i>K-folds</i> e <i>Cross Validation</i>	102
Figura 33 - Processo de Aprendizado e geração da Matriz Final	105

LISTA DE GRÁFICOS

Gráfico 1	- Taxa de Mortalidade Infantil, Neonatal e Pós-neonatal, Ceará:2000 a 2017.	13
Gráfico 2	- Coeficiente de Mortalidade Geral no Brasil de 2000 a 2016.	36
Gráfico 3	- Taxa de Mortalidade Infantil (TMI) no Brasil de 2000 a 2017.	37
Gráfico 4	- Mortalidade Geral no Brasil em 2017.	38
Gráfico 5	- Correlação IDHM x TMI dos municípios do Ceará em 2015.	40
Gráfico 6	- Evolução do IDHM do estado do Ceará.	41
Gráfico 7	- Taxa de Mortalidade Infantil e Componentes. RMF, 2000 a 2016.	42
Gráfico 8	- Coeficiente de Mortalidade infantil por causas evitáveis por região.	46
Gráfico 9	- Taxa de Mortalidade Materna no Ceará, de 2000 a 2014.	48
Gráfico 10	- Nascidos vivos no Ceará de 2013 a 2017 por sexo.	83
Gráfico 11	- Óbitos infantis no Ceará de 2013 a 2017 por sexo.	84
Gráfico 12	- Gráfico de Pareto por tempo de sobrevida.	85
Gráfico 13	- Gráfico de Dispersão - Óbitos: Peso ao nascer x Semanas de Gestação.	86
Gráfico 14	- Gráfico de Dispersão - NV: Peso ao nascer x Semanas de Gestação.	87
Gráfico 15	- Histograma: Idade das mães - óbitos infantis no Ceará de 2013 a 2017.	88
Gráfico 16	- Distribuição dos Óbitos Infantis a partir do Apgar1 e Apgar5.	96
Gráfico 17	- Gráfico de Dispersão - Óbitos infantis: Apgar1 x Peso ao nascer.	97
Gráfico 18	- Gráfico de Dispersão - Óbitos infantis: Apgar1 x Semanas de gestação.	98
Gráfico 19	- Gráfico de Acurácia do melhor treinamento (K-fold = 2).	103

LISTA DE TABELAS

Tabela 1	- Comparativo “Idade da Mãe” entre óbitos infantis x nascidos vivos.	88
Tabela 2	- Perfil das gestantes (Óbitos infantis) por Faixa etária x Escolaridade.	90
Tabela 3	- Perfil das gestantes (Nascidos Vivos) por Faixa etária x Escolaridade.	90
Tabela 4	- Comparativo do estado civil das mães: óbitos infantis x nascidos vivos.	91
Tabela 5	- Consultas pré-natal: óbitos infantis x nascidos vivos.	92
Tabela 6	- Comparativo: Faixa etária x Consultas pré-natal.	92
Tabela 7	- Relação de localidades do Ceará com os maiores TMI em 2017.	94
Tabela 8	- Resultados obtidos em 10 treinamentos com 100 épocas cada.	104

LISTA DE QUADROS

Quadro 1	- Lista resumida de mortes evitáveis por intervenções do SUS.	46
Quadro 2	- Relação de Municípios que formam a RMF.	76
Quadro 3	- Variáveis independentes.	80
Quadro 4	- Cálculo da pontuação do Apgar.	95
Quadro 5	- Classificação da pontuação do Apgar no 1º e 5º minuto.	95
Quadro 6	- Variáveis explicativas e variável Alvo.	101

SUMÁRIO

I - INTRODUÇÃO	11
1.1 O PROBLEMA DA PESQUISA	16
1.3 DELIMITAÇÃO DO ESTUDO	18
1.4 RELEVÂNCIA DO ESTUDO.....	19
1.5 BASE EPISTEMOLÓGICA.....	19
1.6 ORGANIZAÇÃO.....	26
II - REFERENCIAL TEÓRICO	28
2.1 SAÚDE PÚBLICA	28
2.1.1. <i>SUS (Sistema Único de Saúde).....</i>	<i>30</i>
2.1.2. <i>Saúde Pública x Saúde Coletiva.....</i>	<i>32</i>
2.2 EPIDEMIOLOGIA	32
2.2.1. <i>Indicadores de Saúde.....</i>	<i>35</i>
2.3 MORTALIDADE INFANTIL	41
2.3.1. <i>Principais causas da mortalidade Infantil.....</i>	<i>44</i>
2.3.2. <i>Óbitos evitáveis.....</i>	<i>45</i>
2.4 MORTALIDADE MATERNA	47
2.4.1. <i>Near Miss.....</i>	<i>49</i>
2.4.2. <i>A importância do Pré-natal.....</i>	<i>50</i>
2.5 A EVOLUÇÃO DA TECNOLOGIA.....	50
2.5.1. <i>Primeira Plataforma.....</i>	<i>50</i>
2.5.2. <i>Segunda Plataforma.....</i>	<i>51</i>
2.5.3. <i>Terceira Plataforma.....</i>	<i>51</i>
2.5.4. <i>Quarta Plataforma.....</i>	<i>52</i>
2.6 INTELIGÊNCIA ARTIFICIAL (IA).....	52
2.6.1. <i>Árvore de decisão.....</i>	<i>54</i>
2.6.2. <i>Lógica Fuzzy.....</i>	<i>55</i>
2.6.3. <i>Grafos.....</i>	<i>56</i>
2.6.4. <i>Teorema de Bayes.....</i>	<i>57</i>
2.6.5. <i>Redes Neurais.....</i>	<i>58</i>
2.6.6. <i>Machine Learning (Aprendizado de máquina).....</i>	<i>62</i>
2.6.7. <i>Deep Learning (Aprendizagem profunda).....</i>	<i>65</i>
2.7 BUSINESS INTELLIGENCE (BI).....	66
2.7.1. <i>Data Warehouse.....</i>	<i>67</i>
2.8 DATA MINING (MINERAÇÃO DE DADOS)	67
2.9 O MÉTODO KDD (KNOWLEDGE-DISCOVERY IN DATABASES).....	69
2.10 MAPEAMENTO DE PROCESSOS.....	71
III - METODOLOGIA.....	73
3.1 TIPO E NATUREZA DO ESTUDO	73
3.2 LOCAL E PERÍODO	75

3.3	POPULAÇÃO E AMOSTRA DO ESTUDO	75
3.3.1	<i>Critérios de elegibilidade: Inclusão e Exclusão</i>	77
3.4	COLETA DE DADOS E EVIDÊNCIAS	77
3.4.1.	<i>Fonte de dados</i>	77
3.4.2.	<i>Procedimentos e instrumentos de coleta de dados</i>	78
3.4.3.	<i>Controle de qualidade de dados</i>	78
3.5	VARIÁVEIS DE ESTUDO	79
3.5.1.	<i>Variáveis dependentes</i>	79
3.5.2.	<i>Variáveis independentes</i>	79
3.6	TRATAMENTO E ANÁLISE DAS EVIDÊNCIAS (ANÁLISE DOS DADOS)	80
3.7	LIMITAÇÃO DO MÉTODO	81
3.8	ASPECTOS ÉTICOS	81
IV	- LEVANTAMENTO DE DADOS	82
4.1	MER – MODELO DE ENTIDADES DE RELACIONAMENTOS	82
4.2	DADOS COLETADOS	83
4.3	MEDIDAS DE APGAR	94
4.4	PROCESSO DE GERAÇÃO DO MODELO DE PREDIÇÃO	98
V	- ANÁLISE DOS DADOS	99
VI	- CONCLUSÕES	106
VII	- PERSPECTIVAS E PROPOSTAS DE NOVOS TRABALHOS	108
	REFERÊNCIAS BIBLIOGRÁFICAS	110
	ANEXO I – DECLARAÇÃO DE NASCIDO VIVO	118
	ANEXO II – DECLARAÇÃO DE ÓBITO	119
	ANEXO III – RELAÇÃO DOS MUNICÍPIOS DO ESTADO DO CEARÁ	120
	APÊNDICE I – PROCESSO PARA GERAÇÃO DE MODELOS PREDITOS	124
	APÊNDICE II – DICIONÁRIO DE DADOS DO ARQUIVO DE ÓBITOS INFANTIS	125
	APÊNDICE III – FERRAMENTA DE PREDIÇÃO	126
	APÊNDICE IV – SIMULADOR PARA A EDUCAÇÃO NA ÁREA DA SAÚDE	127
	APÊNDICE V – SCRIPT PYTHON PARA EXECUÇÃO	128
	APÊNDICE VI – EXECUÇÃO DO SCRIPT PYTHON	136
	APÊNDICE VII – MATRIZ DE CORRELAÇÃO DAS VARIÁVEIS <i>FEATURES</i>	147
	APÊNDICE VIII – DESEMPENHO DOS TREINAMENTOS ($K = 10$)	148
	APÊNDICE IX – EXEMPLO DOS PESOS SINÁPTICOS OBTIDOS	152

I - INTRODUÇÃO

No Brasil, quando falamos de políticas públicas de saúde isso nos remete aos sucessivos fracassos e sofrimentos da nossa população menos favorecida durante décadas, pois quem pode pagar, tem um plano de saúde privado, que surgiu com a lacuna deixada pelo poder público que nos anos 40 não conseguia suprir as demandas da população, abrindo este espaço para a iniciativa privada.

O Sistema Único de Saúde (SUS) tem sua fundamentação legal na Lei Nº 8.080, de 19/09/1990, com o objetivo de resolver os principais problemas da saúde pública brasileira. Entretanto, são evidentes as dificuldades deste sistema, principalmente quanto ao acesso aos serviços de saúde e às precárias condições de atendimento na maioria dos casos, abrindo desta forma espaço para a consolidação dos planos e dos seguros-saúde (FERREIRA, 2011).

A Constituição Federal de 1988, no art. 196, define:

A saúde é direito de todos e dever do estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e outros agravos e ao acesso universal e igualitário às ações e serviços para sua promoção, proteção e recuperação (BRASIL, 1988).

A Razão de Mortalidade Materna (RMM) passou de 143,2 óbitos por 100 mil nascidos, em 1990, para 68,2 óbitos em 2010. Em se tratando da Taxa de Mortalidade Infantil (TMI), a redução foi de 47,1 óbitos por mil nascimentos, em 1990, para 16 em 2010 (BRASIL, 2004).

É possível destacar alguns fatos que contribuíram para estas reduções e mudanças:

- Em 2000, a Organização das Nações unidas (ONU) estabeleceu metas de redução dos índices de mortalidade materno e infantil.

- Em 2004, foi lançado no Brasil o Pacto Nacional pela Redução da Mortalidade Materna e Neonatal, para monitorar a implementação de ações de proteção de saúde da criança e da mulher (BRASIL, 2004).

- Em 2009, foi firmado o Pacto da Redução da Mortalidade Infantil no Nordeste e na Amazônia Legal.

- Em 2010, foi lançada portaria que dava diretrizes para a Rede de Atenção à Saúde (RAS), que instituía a Rede Cegonha (RC) para assegurar à mulher direitos universais para o planejamento reprodutivo e à atenção humanizada à gravidez, ao parto e ao para o puerpério,

bem como à criança o direito ao nascimento seguro, ao crescimento e ao desenvolvimento saudável (BRASIL, 2011).

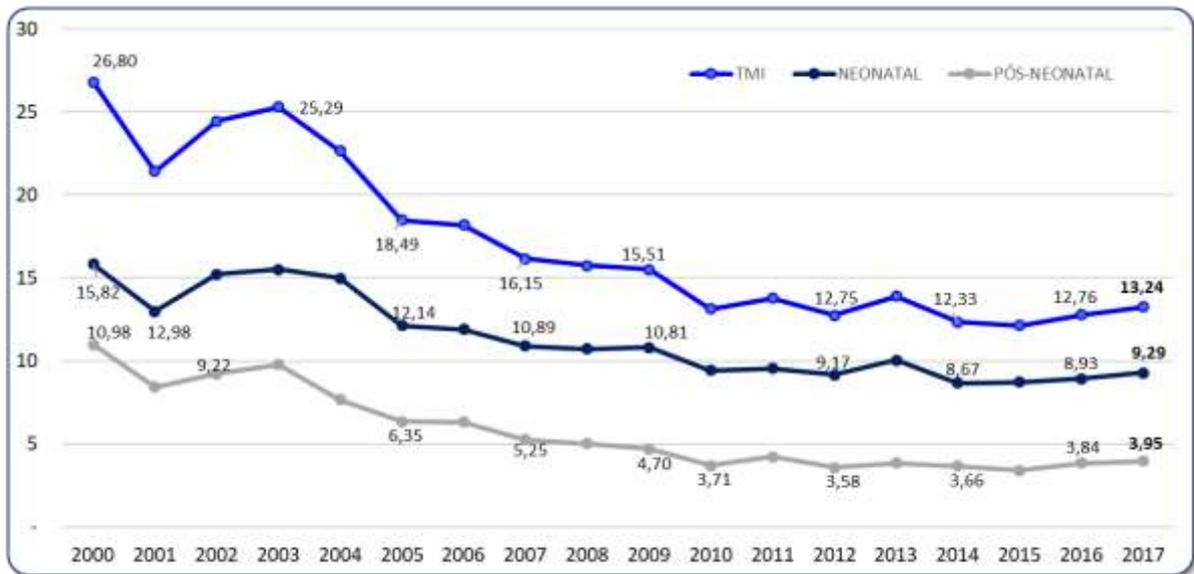
A rede cegonha está presente em mais de 4000 municípios do Brasil e já atendeu mais de dois milhões de gestantes. Em 2012, foram realizadas mais de 18 milhões de consultas pré-natais pelo SUS e um milhão e seiscentas mil mulheres fizeram, no mínimo, sete consultas. (Fonte: Site www.unasus.gov.br).

No Brasil, a saúde materno e infantil ainda é motivo de preocupação. Apesar de grandes avanços nas políticas públicas de saúde no país, os indicadores de mortalidade materna e infantil são elevados, principalmente quando comparados aos países mais desenvolvidos (MORSE, 2011; TEIXEIRA; GOMES & SOUZA, 2012).

Os números de 2017, em preparação, já apresentam uma retomada no crescimento da mortalidade infantil no Brasil que não se via há mais de 25 anos. O Jornal The Lancet relata que em 2017 mais de um milhão de famílias foram excluídas do Programa Bolsa Família, voltado para a erradicação da pobreza e da fome no Brasil, que transferia dinheiro direto para as famílias mais pobres (DONIEC; DALL'ALBA; KING, 2018). As medidas de austeridade que os programas sociais estão sofrendo tendem a piorar a morbidade e mortalidade infantil no Brasil (RASELLA *et. al.*, 2018).

A edição do jornal Folha de São Paulo de 16 de julho de 2018, relata uma alta inédita na taxa de mortalidade infantil de 2016, que não crescia desde 1990, e pior, com tendência de aumento em 2017. A epidemia do vírus da Zika e a crise econômica são as principais causas apontadas pelo Ministério da Saúde. Neste período tivemos queda no número de nascimentos, impactando no cálculo da taxa de mortalidade, e mortes de bebês por malformações graves (FOLHA DE SÃO PAULO, 2018).

Conforme o Gráfico 1, podemos observar que no estado do Ceará a Taxa de mortalidade Infantil (TMI) reduziu drasticamente nos últimos anos, mas não avançou de 2010 a 2017 e pior, dá mostras de retrocesso deste indicador em 2017 e 2018 (dados em processamento). Este é o nosso grande desafio: como quebrar esta barreira e voltar a derrubar esses índices? Podemos usar recursos tecnológicos nesta batalha pela vida de nossas crianças?

Gráfico 1: Taxa de Mortalidade Infantil, Neonatal e Pós-neonatal, Ceará:2000 a 2017.

Fonte: Elaborado pelo autor com dados do Datasus/SIM/SINASC.

Vale lembrar o maior acesso da população aos serviços de saúde graças à criação do SUS (Sistema Único de Saúde), ressaltando a grande contribuição do Programa de Saúde da Família, que desde 1994 vem promovendo e expandindo a atenção territorializada, contribuindo de forma expressiva para a redução dos óbitos infantis em todas as regiões brasileiras (SILVA *et.al.*, 2016).

Nos países desenvolvidos, o acesso à educação de qualidade, boas condições de moradia, saneamento básico e serviços de saúde disponíveis, são fatores que influenciaram a redução da mortalidade infantil nas primeiras décadas do século XX, mesmo antes das grandes descobertas técnico-científicas da medicina moderna (SIMÕES; MONTEIRO, 1995).

A Rede Cegonha (RC) é uma estratégia que visa assegurar à mulher o direito ao planejamento reprodutivo e à atenção humanizada à gravidez, ao parto e ao puerpério, bem como à criança o direito ao nascimento seguro e ao crescimento e desenvolvimento saudável.

Os recursos tecnológicos e de comunicações estão disponíveis na maioria das comunidades brasileiras, principalmente nas regiões metropolitanas, com boa qualidade técnica e altos níveis de disponibilidade. Estes recursos tecnológicos já são utilizados em larga escala pelas operadoras de planos de saúde (OPS) para obter qualidade nos serviços prestados aos seus clientes, baixos custos de operação e extração de conhecimento estratégico para a tomada de decisão. As redes de atenção à saúde também precisam e devem usar estes recursos tecnológicos para melhorar sua qualidade de serviços, reduzir custos e atuar estrategicamente em suas decisões.

Estamos usando todo o conhecimento disponível ? Como poderíamos otimizar os recursos e benefícios da RC com mais tecnologia ? Existe gestão estratégica na saúde coletiva como fazem as organizações privadas ? Essas questões precisam de respostas e a baixa produção de estudos acadêmicos sobre o uso da IA (Inteligência Artificial) junto com a IC (Inteligência Competitiva) aplicada no apoio à tomada de decisão nas políticas públicas, menos ainda na saúde pública, só ressaltam as preocupações, mas também nos mostra um mar azul de oportunidades.

É neste contexto da saúde coletiva no estado do Ceará que este projeto propõe o uso de tecnologias inovadoras e modernas na geração de conhecimento estratégico através da inteligência artificial para apoiar uma gestão baseada no conhecimento e na tecnologia, possibilitando uma elevação no patamar da gestão pública da saúde em nosso estado.

A Informática Biomédica é uma realidade, fruto da colaboração de profissionais de áreas distintas. Não é necessário que um profissional da área da saúde entenda de algoritmos ou programação, mas é importante que esses profissionais de saúde aprendam a trabalhar em conjunto com profissionais da Ciência da Computação e da Informação, auxiliando no levantamento dos requisitos e definição de regras para trazer melhores e maiores benefícios à saúde. Esta tese se baseia nessa união multidisciplinar de profissionais de áreas distintas, mas capazes de gerar ganhos reais para a sociedade.

Segundo Turban e Volonino (2013), um algoritmo é um conjunto de regras ou instruções elaboradas para resolver um problema em um número finito de etapas.

Um algoritmo pode ser associado a uma estratégia para resolver um determinado problema. Os dados do problema constituem a ENTRADA ou os dados do algoritmo. A solução do problema constitui a SAÍDA. Um algoritmo pode ser representado por uma função f , a qual associa uma saída $S = f(E)$, a cada entrada E . Diz-se então que o algoritmo computa a função f , conforme a Figura 1 abaixo (SZWARCFITER, 2018).

Figura 1 – Modelo de um algoritmo.

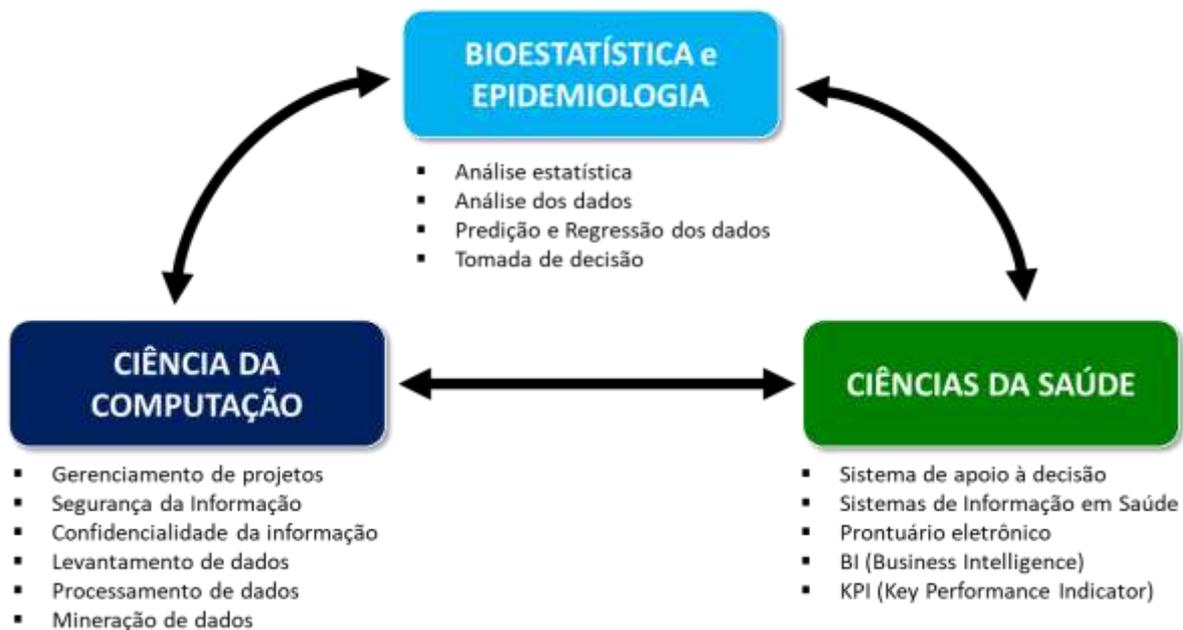


Fonte: Szwarcfiter, 2018.

O estudo de algoritmos e de lógica de programação é essencial no contexto de criação de um software ou programa de computador. Como definição geral, um algoritmo representa um conjunto de regras para a solução de um problema. Dessa forma, uma receita para elaborar um alimento é um exemplo de algoritmo (informal e impreciso), pois descreve as regras necessárias para a conclusão do seu objetivo (FURLAN *et. al.*, 2019).

São várias as oportunidades dessa união de esforços. Na Figura 2 podemos ver como a relação entre profissionais de saúde, estatísticos, epidemiologistas e profissionais da computação pode gerar benefícios reais para saúde pública brasileira.

Figura 2 – Relação entre as áreas: Profissionais da Saúde e da Computação.



Fonte: Elaborado pelo autor.

São várias as atividades dos profissionais de saúde para prestar serviços de qualidade, munir os usuários de informações úteis, usar sistemas de informação e tomar decisões baseados em dados e sistemas confiáveis. Nesse contexto, a ciência da computação, a estatística, a epidemiologia e as ciências da saúde, precisam andar juntas e compartilhar conhecimentos para atingir objetivos complexos, como a saúde pública de qualidade que desejamos.

A evolução tecnológica computacional tem origem nas décadas de 50 e 60 com os *Mainframes*, que eram disponíveis só para grandes organizações públicas e privadas. Somente na década de 80 chega à democratização da tecnologia com os PCs (*Personal Computers*) e a TV por assinatura. Na década de 90 a internet, a telefonia móvel e a privatização das telecomunicações no Brasil. Todo esse caminho foi necessário para que a partir de 2010

podéssemos assistir à popularização das tecnologias com a chegada da computação em nuvem, *smartphones* modernos, mobilidade, e assim criar um ambiente tecnologicamente adequado para processar grandes volumes de dados com capacidade para manipular textos, fotos, vídeos, TVs, links, tudo em alta velocidade e disponibilidade. É neste contexto que a IA (Inteligência Artificial) se desenvolve, possibilitando que qualquer instituição possa se beneficiar dos seus poderosos recursos de geração de conhecimento a partir de grandes volumes de dados.

A inteligência Artificial começou a ser estudada no pós-guerra na década de 50, mas ficou adormecida por várias décadas a espera das condições ideais para o seu desenvolvimento, que encontramos agora, como: computação em nuvem, algoritmos complexos, Big Data e alto poder dos computadores atuais em armazenar e processar grandes volumes de dados em curto espaço de tempo e com alta disponibilidade.

Segundo Luger (2013), a Inteligência Artificial (IA) revela um campo de estudo jovem e promissor, onde o principal interesse é encontrar um modo efetivo de entender e aplicar técnicas inteligentes para a solução de problemas e para o planejamento de uma gama de problemas práticos.

1.1 O PROBLEMA DA PESQUISA

O principal problema objeto desta pesquisa é o desconhecimento do padrão de modelos preditivos da Mortalidade Infantil com base em modelos de Inteligência Artificial no estado do Ceará. Este problema faz emergir a questão central da pesquisa:

“Quais são os padrões de modelos preditivos da Mortalidade Infantil com base em modelos de Inteligência Artificial no estado do Ceará?”.

O desconhecimento do padrão gera a necessidade de desenvolver, testar e aplicar modelos preditivos para auxiliar as políticas públicas. O uso da inteligência Artificial é uma oportunidade tecnológica já utilizada por grandes organizações públicas e privadas pelo mundo, aproveitando a onda de barateamento dos recursos computacionais e tecnológicos.

Segundo Hessen (2012), “Conhecer significa apreender espiritualmente um objeto. Essa apreensão, via de regra, não é um ato simples”. Gerar conhecimento sobre os problemas de saúde pública não é tarefa fácil, mas podemos usar também os recursos tecnológicos que as grandes corporações já utilizam para conhecer seus clientes e o mercado.

1.1.1. Principais Causas detectadas:

- Baixa qualidade das informações no contexto da mortalidade infantil.
- Baixo nível tecnológico das áreas responsáveis pela gestão.
- Baixo nível tecnológico das regiões de maior ocorrência da mortalidade infantil.
- Distanciamento da saúde pública no Brasil do estado da arte em técnicas modernas da computação, como: Mineração de dados dispersos, refinamento de dados e algoritmos de Inteligência Artificial.

1.1.2. Efeitos esperados do projeto:

- Desenvolvimento de modelos preditivos com base na Inteligência Artificial;
- Geração de conhecimentos importantes para a gestão estratégica ao combate da mortalidade infantil;
- Desenvolvimento de técnicas inovadoras de manipulação de dados no apoio à saúde coletiva, em especial na redução da mortalidade infantil

1.2 OBJETIVOS E METAS

1.2.1. Objetivo Geral

O principal objetivo desta pesquisa é caracterizar os padrões de modelos preditivos da Mortalidade Infantil com base em modelos de Inteligência Artificial no estado do Ceará.

1.2.2. Objetivos Intermediários.

O objetivo geral se desdobra nos seguintes objetivos específicos:

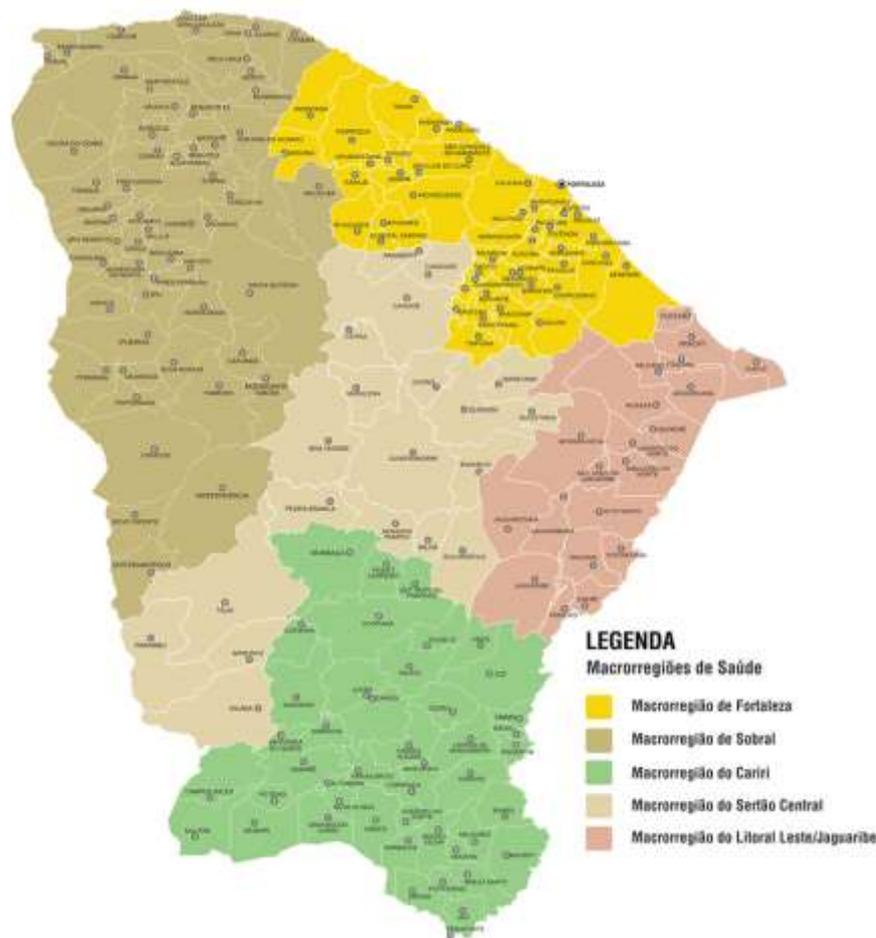
- Verificar a consistência e completude das bases de dados de mortalidade infantil;
- Aplicar algoritmos de predição da mortalidade infantil com base em modelos de Inteligência Artificial;
- Verificar o modelo preditivo da Mortalidade infantil e seus componentes cronologicamente;
- Identificar os padrões das variáveis preditivas e sua aplicação espacial entre municípios do estado do Ceará.

Como meta, esta pesquisa busca a geração de modelos preditivos testados e aplicados à Mortalidade infantil no estado do Ceará.

1.3 DELIMITAÇÃO DO ESTUDO

A pesquisa foi realizada utilizando dados de nascidos vivos e óbitos infantis ocorridos no estado do Ceará no período de 2013 a 2017. O Ceará é um dos estados que formam a República Federativa do Brasil e está situado na Região Nordeste, sendo Fortaleza a sua capital. No último censo de 2010, o Ceará apresentou uma população de 8.452.381 habitantes distribuídos em seus 184 (cento e oitenta e quatro) municípios, conforme Anexo III (Relação dos municípios do estado do Ceará, que ocupam área de 148.886,31 km² e Densidade demográfica de 56,76 habitantes por km². A Figura 3 apresenta o mapa do estado do Ceará com as suas Macrorregiões de Saúde.

Figura 3 – Mapa do estado do Ceará por Macrorregiões de Saúde.



Fonte: SESA – Secretaria de Saúde do estado do Ceará (2014).

Delimitação temporal: serão considerados os eventos de “Nascimentos” e “Mortalidade Infantil” ocorridos no período de 2013 a 2017 registrados na base de dados Datasus/SIM e Datasus/SINASC.

1.4 RELEVÂNCIA DO ESTUDO

A mortalidade infantil é historicamente um indicador diretamente relacionado às condições de vida de um país (KOONTZ; KATHLEEN; RUDERMAN, 2004). O tema em questão está também ligado diretamente nas relações com as condições de vida da saúde da mulher (BRASIL, 2004).

O começo da vida deve ser uma prioridade para a sociedade. Precisamos sensibilizar a todos da importância do começo da vida e de seus impactos na futura população adulta.

As ações para redução da mortalidade infantil são reconhecidas e incorporadas aos diversos compromissos assumidos pelo Brasil internacionalmente, como os Objetivos do Milênio, e nacionalmente como o Pacto pela Redução da Mortalidade Materna e Neonatal, o Pacto pela Vida e, mais recentemente, o Programa Mais Saúde. A redução da mortalidade infantil é prioridade no Brasil (MS, 2009).

Como cidadão e pesquisador, acredito que todo esforço deve ser empreendido para reduzir as mortes de crianças no Brasil, ainda estamos longe dos países do primeiro mundo nesse indicador, mas cuidar do nascimento, acolhimento e crescimento das crianças brasileiras, é missão de toda a população adulta, especialmente dos profissionais de saúde, do poder público e dos que fazem a academia e pesquisa no Brasil, que buscam entender e conhecer os problemas para propor soluções viáveis de implementação. Como cita Cecília Minayo (2000), “Nada pode ser considerado intelectualmente um problema se não tiver sido em última instância, um problema na vida prática”.

A mortalidade infantil é um problema grave da vida prática que convivemos no Brasil com maior prevalência no Norte e Nordeste. Essa tese de caráter multidisciplinar (Computação, Administração, Estatística e Saúde Pública) quer contribuir com a modernização tecnológica da saúde pública, na geração de informações relevantes e possibilitar uma nova forma de olhar o problema da mortalidade no Brasil através da geração de modelos preditivos usando algoritmos de Inteligência Artificial, pois todo esforço deve ser realizado, não para prever mortes, mas para evitá-las.

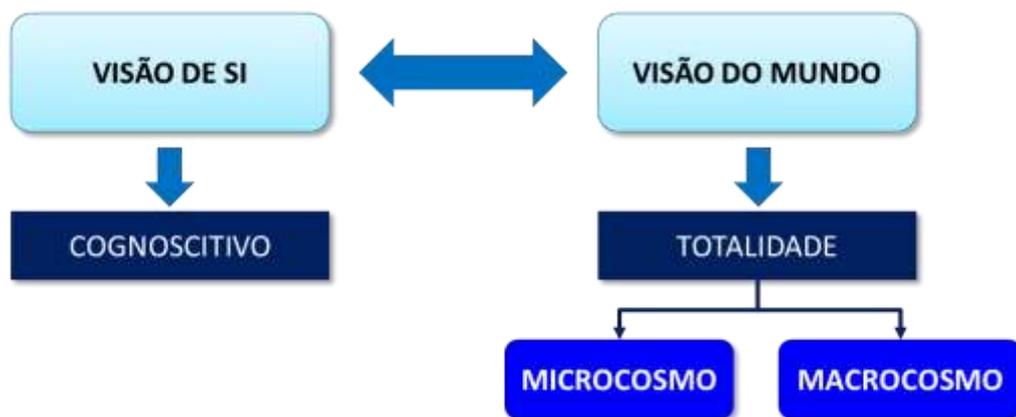
1.5 BASE EPISTEMOLÓGICA

A teoria do conhecimento é uma disciplina filosófica. A fonte e o fundamento do conhecimento humano é a razão ou a experiência ? Essa é a questão sobre a origem do conhecimento (HESSEN, 2012).

Para citar a filosofia nesta pesquisa, devemos inicialmente entender a sua essência, qual a sua aplicação e quais são as suas vertentes. A essência da Filosofia, segundo Hessen (2012): É a autorreflexão do espírito. “É a tentativa do espírito humano de atingir uma visão do mundo, através da *AUTO-REFLEXÃO* sobre suas funções valorativas: *TEÓRICAS e PRÁTICAS*”.

A Filosofia está representada na Figura 4, mostrando um modelo simplificado da sua abrangência quanto à visão de *SI* e do *MUNDO*.

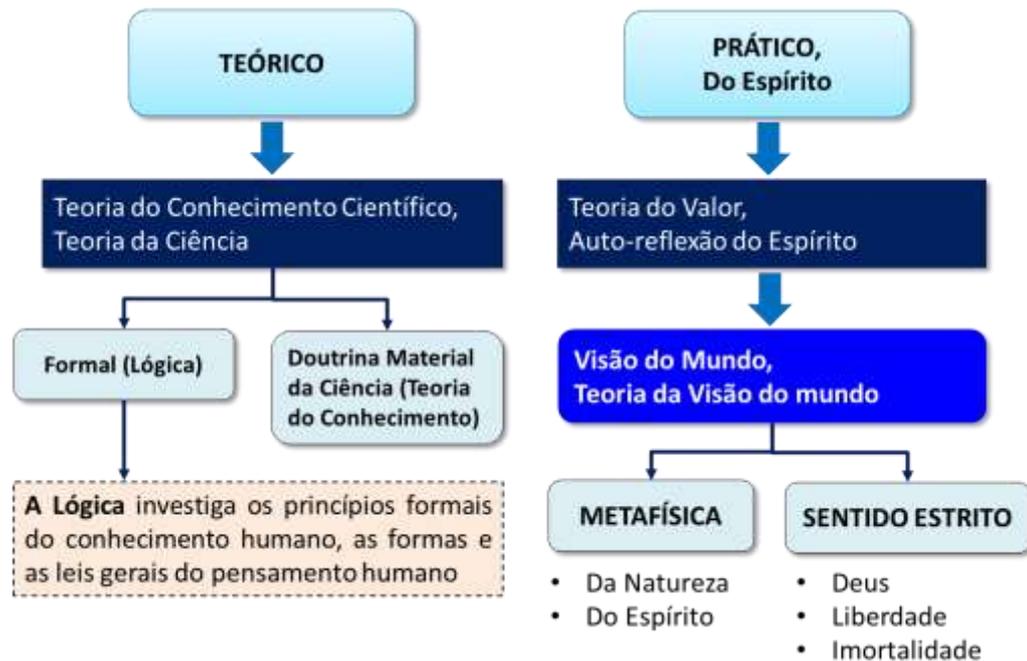
Figura 4 – Modelo de abrangência da filosofia.



Fonte: Elaborado pelo autor.

Ainda segundo Hessen (2012), a filosofia possui valores Teóricos e Práticos, conforme está representado na Figura 5 abaixo.

Figura 5 – Valores da filosofia.



Fonte: Elaborado pelo autor.

Entre os valores teóricos estão a **TEORIA DO CONHECIMENTO CIENTÍFICO** e a **TEORIA DA CIÊNCIA**, é nesta vertente que a lógica se desenvolve para construir algoritmos que vão desde o formalismo simples de Aristóteles até os programas modernos de computadores executando algoritmos de Inteligência Artificial como propõe esta tese.

Platão, Sócrates e Aristóteles criaram as bases do pensamento filosófico ocidental. Aristóteles desenvolveu o sistema de silogismos (conexão de ideias, raciocínio) para criar um raciocínio organizado, permite estabelecer um processo de geração de conclusões a partir de premissas verdadeiras, é uma forma de raciocínio irrefutável e lógico, uma forma de pensar racionalmente. Os silogismos dão uma estrutura de argumentação que sempre fornecem conclusões corretas, dadas premissas corretas.

Geralmente, em um raciocínio perfeito deve haver três proposições, das quais duas funcionam como antecedentes, sendo assim chamadas de premissas, e a terceira é a consequente, isso é, a conclusão que brota das premissas. Um exemplo clássico do silogismo de Aristóteles é: [a] “Sócrates é um homem”; [b] “todo homem é mortal”, logo, [c] “Sócrates é mortal”. Neste exemplo, a sentença [c] é a conclusão obtida a partir das informações coletadas antecipadamente nas sentenças [a] e [b] (REALE; ANTISERI, 2003).

E assim começou o campo da lógica, que posteriormente expandiu com Blaise Pascal (1623-1662, matemático, físico e filósofo francês), Thomas Bayes (1702-1761, matemático

britânico. O teorema de Bayes aplicado na estatística influenciou diretamente o algoritmo Naive Bayes utilizado na IA), George Boole (1815-1864, matemático e filósofo britânico, o pai da Álgebra Booleana) entre outros.

A ideia da lógica formal pode ser traçada até os filósofos da Grécia antiga, mas o desenvolvimento matemático começou mesmo com o trabalho de George Boole, que definiu os detalhes da lógica proposicional ou lógica booleana em 1847. Gottlob Frege (1848-1925) ampliou a lógica de Boole para incluir objetos e relações, criando a lógica de primeira ordem, e Alfred Tarsky (1902-1983) introduziu a teoria de referência que mostra como relacionar os objetos de uma lógica a objetos do mundo real (NORVIG; RUSSELL, 2013).

René Descartes (1596-1650), filósofo, físico e matemático francês, autor da célebre frase: “Penso, logo existo”, também conhecido pelo seu nome latino *Renatus Cartesius*. Descartes é reconhecido como o pai da filosofia moderna, mas também é estudado na física e matemática, onde criou as fundações da geometria analítica. Até hoje ainda são ensinadas nas escolas as “Coordenadas cartesianas”, que recebeu esse nome em homenagem ao seu nome latino.

Descartes (2010), em sua obra “*Discours de la Méthode*”, traduzindo do francês, “Discurso do Método”, introduz através de seis diferentes ensaios que visam a definição da verdade: Os fundamentos da dúvida metódica; a análise; a síntese; a enumeração e a prova; os valores morais e a aplicação do seu método às ciências.

Gottfried Wilhelm Leibniz (1646-1716), filósofo, cientista, matemático, diplomata e bibliotecário alemão, elaborou o Cálculo Diferencial e Cálculo Integral, utilizados em todas as ciências exatas. Leibniz propôs como a fonte última da verdade dos teoremas das ciências demonstrativas o “axioma de identidade”. Leibniz formula esse princípio de várias maneiras, por exemplo, “A é idêntico a A”, “A é o mesmo que A”, “A não é não A”, “cada coisa é o que é” e “cada coisa é igual a si mesma”. Entretanto, a versão predominante do axioma em alguns escritos que registram seus esforços em lógica simbólica é a fórmula “ $A = A$ ” – ou seja, “A é igual (ou idêntico) a A”. De acordo com o axioma de identidade, todas as sentenças desse tipo são verdadeiras (CASS, 2013).

Thomas Bayes (1702-1761), pastor presbiteriano e matemático inglês, conhecido por formular o Teorema de Bayes, “A Inferência Estatística constitui um método estruturado e científico para a tomada de decisões a partir de valores observáveis, tendo uma medida de incerteza (probabilidades, não observáveis) acerca do fenômeno”. A teoria bayesiana atribui à

probabilidade um grau de crença, subjetivo, portanto, à veracidade de alguma proposição, o que se distingue fundamentalmente do que se convencionou chamar de teoria clássica (CARVALHO; CHIANN, 2013).

David Hume (1711– 1776), filósofo britânico, nascido na Escócia, conhecido por seu ceticismo filosófico, criou a teoria da indução: “A indução é a generalização da informação pela persistência dos dados” ou “Nós adquirimos um conjunto de regras através da exposição repetida a associações entre elementos”. O problema da indução está em que, segundo Hume:

“A razão jamais pode nos mostrar a conexão entre dois objetos, mesmo com a ajuda da experiência e da observação de sua conjunção constante em todos os casos passados. Portanto, quando a mente passa da ideia ou impressão de um objeto à ideia de outro objeto, ou seja, à crença neste, ela não está sendo determinada pela razão, mas por certos princípios que associam as ideias desses objetos, produzindo sua união na imaginação” (HUME, 2009, p. 121).

A dúvida de Hume em relação à racionalidade da indução parece repousar sobre a questão de como podemos justificar de maneira racional uma inferência acerca de um objeto observado pela experiência de algo que escapa ao alcance da observação, ou seja, algo que escapa ao testemunho dos sentidos ou da linguagem (PIMENTEL, 2017).

Segundo Russell (2016), o princípio lógico também define princípios óbvios conhecidos tradicionalmente de “Leis do pensamento”, são elas:

- a) A lei da identidade: “O que é, é”;
- b) A lei da contradição: “Nada pode conjuntamente ser e não ser”;
- c) A lei do terceiro excluído: “Tudo tem de ou ser ou não ser”.

A inferência é um sinal de inteligência que marca a superioridade do homem em relação a máquina. Ao mesmo tempo, o uso da inferência e da lógica tradicional é tão insensato que deixa dúvidas com respeito a esta superioridade; a inferência silogística desde Aristóteles até Bacon, é uma coisa que uma simples máquina de calcular pode realizar melhor do que um professor (RUSSELL, 2009a).

A inferência indutiva é um processo de raciocínio para concluir sobre todo o conjunto através do estudo de parte dele, mas da mesma classe e características, isto é, raciocinar do particular para o geral. Por exemplo: [a] pacientes com déficit de atenção atendidos em 1990 sofriam de ansiedade; [b] pacientes com déficit de atenção atendidos em 1991 sofriam de ansiedade, logo posso inferir logicamente que todos os pacientes com déficit de atenção

sofrem de ansiedade. Esta conclusão pode ser ou não verdadeira, mas nos permite uma boa generalização.

Usando formalismo podemos descrever a inferência dedutiva assim, para um conjunto de objetos, $X = \{a, b, c, d, \dots\}$, se a propriedade P é verdadeira para a , e se P é verdadeira para b , e se P é verdadeira para c , então P é verdade para todo X . A inferência indutiva é um dos principais métodos utilizados para gerar conhecimento novo e prever eventos futuros.

Para Russell (2016), o princípio da indução tem duas partes formuladas assim:

- a) Quando se descobriu que uma coisa de um certo gênero A está associada a uma coisa de outro gênero B, e nunca se a viu dissociada de uma coisa do gênero B, quanto maior for o número de casos em que A e B estão associados, maior a probabilidade de estarem associados num caso novo no qual uma delas se saiba estar presente;
- b) Nas mesmas circunstâncias, um número suficiente de casos de associação tornará a probabilidade de uma nova associação quase uma certeza, e fará que se aproxime da certeza sem limite.

Bertrand Russell e seu colega de Cambridge, o filósofo matemático Alfred North Whitehead, tentaram mostrar a filosofia conhecida como “Logicismo” (1910-1913), que as verdades obtidas da matemática são em última instância, consequências dedutivas das leis da lógica. Infelizmente, essa filosofia era quase impossível de ser validada devidos a certos paradoxos. Entretanto, esse é um dos casos em que o fracasso possui mais valor do que o sucesso, pois o trabalho de Russell em torno da lógica influenciou imensamente o pensamento filosófico daquele século. Ele e outros do seu círculo simplesmente mudaram a direção de nosso pensamento (RUSSELL, 2009b).

Continuando esse percurso, encontramos Ludwig Wittgenstein (1889-1951), austríaco, um dos filósofos mais influentes do século XX, com contribuições para a filosofia da linguagem e da mente, bem como para a lógica e a epistemologia. Wittgenstein inventou uma forma elegante de representar funções-de-verdade de proposições elementares, conhecida como Tabela-de-verdade (CHILD, 2013).

Por exemplo, ele representa a proposição complexa “p e q” por meio da seguinte Tabela-verdade, representada abaixo na Figura 6:

Figura 6 - Tabela Verdade.

	p	q	
1	V	V	V
2	V	F	F
3	F	V	F
4	F	F	F

Fonte: Child, 2013.

Na figura acima, cada linha numerada representa uma situação possível, uma combinação possível de verdade e falsidade das proposições elementares de p e q . A Tabela nos diz (V ou F na coluna do lado direito) se, naquela situação, a proposição complexa que ela representa é verdadeira ou falsa. Somente a linha 1 é verdadeira, pois as duas proposições são verdadeiras, nas demais, pelo menos uma das proposições é falsa, tornando a proposição complexa também falsa.

E assim, ao longo dos tempos o homem usou a filosofia, que influenciou a lógica, que desenvolveu a matemática e a estatística para desenvolver soluções práticas usando algoritmos. Os algoritmos por sua vez ganharam força com a computação, que aplica os algoritmos em grandes volumes de dados, processando-os em alta velocidade para gerar conhecimento ao homem de forma rápida. Neste contexto a IA se desenvolveu rapidamente para gerar conhecimentos que a mente humana não consegue capturar a partir de grande volume de dados.

A palavra *algoritmo* vem de Al-Khowarazmi, um matemático persa do século IX, cujos escritos também introduziram os numerais arábicos e a álgebra na Europa. Acredita-se que o primeiro algoritmo não trivial seja o algoritmo de Euclides para calcular o maior divisor comum (NORVIG; RUSSELL, 2013).

Segundo Turban e Volonino (2013), um algoritmo é um conjunto de regras ou instruções elaboradas para resolver um problema em um número finito de etapas.

A partir da segunda grande guerra mundial a IA começa seu desenvolvimento embora as máquinas não estivessem preparadas para o seu potencial. Segue abaixo a cronologia dos principais passos da IA:

- 1943, McCulloch e Pitts desenvolvem as Redes Neurais, um modelo booleano do cérebro humano;

- 1950, Turing publica “Computing Machinery and Intelligence” onde descreve um teste interessante para determinar se um computador exibe um comportamento inteligente. De acordo com o teste de Turing, um computador só poderia ser considerado “inteligente” quando um entrevistador humano, conversando tanto com um humano invisível como com um computador invisível, não consegue determinar quem é quem (TURBAN *et. al.*, 2010).

- 1956, o termo “Inteligência Artificial” foi usado pela primeira vez por John McCarthy em uma conferência no Dartmouth College em Hanover, New Hampshire;

- 1950s, surgem os primeiros programas de IA, o jogador de damas de Samuel, o Logic Theorist de Newell E Simon e o Geometry Theorem Prover de Gelernter;

- 1965, Robinson descobre um método de raciocínio lógico completo;

- 1965, Lotfi Zadeh (1921-2017) propõe a Teoria dos conjuntos Difusos (*Fuzzy Sets*);

- 1969-1979, Desenvolvimento de Sistemas Especialistas;

- 1986, Redes Neurais ganham força com a maior capacidade dos computadores;

- 1987, IA se torna uma ciência;

- 1995, surgem os Agentes Inteligentes.

Agora que a IA já se desenvolveu e pode ser aplicada em vários problemas do nosso cotidiano e em várias áreas do conhecimento, podemos observar que na sua formação e evolução foram necessários conhecimentos multidisciplinares e complementares, como: Filosofia, Matemática, Estatística, Psicologia, Linguística e Engenharia da Computação, sendo neste contexto que esta tese se desenvolve.

1.6 ORGANIZAÇÃO

A fim de estudar em detalhes o tema da mortalidade infantil usando os benefícios da Inteligência Artificial, será realizada uma introdução ao tema através do Capítulo 1, o capítulo 2 apresentará a fundamentação teórica usada para o desenvolvimento dos modelos preditivos de IA sobre a Mortalidade Infantil, baseado em revisão da bibliografia existente em livros, artigos, teses e dissertações, além de relatos obtidos na experiência profissional e acadêmica do autor da tese, sobre os seguintes temas:

- Saúde pública;
- Epidemiologia;
- Mortalidade infantil;

- Mortalidade Materna;
- Evolução da tecnologia;
- Inteligência Artificial (IA);
- Business intelligence (BI);
- Data Mining (Mineração de dados);
- Processo KDD;
- Mapeamento de Processos.

O capítulo 3 aborda a metodologia adotada. Neste capítulo são apresentadas as estratégias de coleta de dados e o processo utilizado para o tratamento, limpeza e processamento dos dados.

O capítulo 4 detalha através da estatística descritiva os dados coletados e tratados, dando uma visão estatística das principais informações. O capítulo 5 detalha os passos do modelo preditivo executado com as principais informações geradas.

Para finalizar o trabalho, o capítulo 6 detalha as conclusões a que chegamos com as propostas de continuação da pesquisa e suas aplicações futuras no capítulo 7.

Na seção Referências bibliográficas são listadas todas as fontes utilizadas na pesquisa, como: teses, dissertações, livros, jornais, sites e artigos.

II - REFERENCIAL TEÓRICO

Para entender o contexto desta pesquisa e o seu caráter multidisciplinar, é necessário a revisão teórica dos conceitos que formam o campo da Saúde Coletiva onde estão inseridos os cuidados básicos com a saúde materna e infantil, e das várias técnicas que sustentam a IA (Inteligência Artificial), a computação e a Epidemiologia baseada na Estatística.

2.1 SAÚDE PÚBLICA

Antes mesmo de Aristóteles conceituar política como uma atividade do ser humano, é possível observar que algumas civilizações já se organizavam para utilizar estratégias coletivas de cuidados e enfrentamentos das ameaças à saúde de sua população. Os egípcios planejavam e construíam cidades com drenagem pública de eflúvios há mais de 3000 anos. Os gregos que muito contribuíram para a definição e uso de políticas públicas de saúde, debatendo questões da saúde e da doença como fenômenos naturais, superando o modelo mágico-religioso, enfrentando os problemas de saúde com soluções concretas (SIEGERIST, 2011).

Somente no século XIV em Portugal e na França e Inglaterra entre os séculos XVI e XVIII, como mediação dos diferentes interesses de uma sociedade mais complexa, industrializada e democrática, é que surgem políticas públicas de saúde como as que conhecemos nos dias de hoje (SAMPAIO *et. al*, 2016).

A Medicina da Idade Média fundamentava-se em práticas eminentemente individuais e provadas. Até o século XVII, o modelo predominante era suscitado pela lepra. O leproso deveria ser descoberto e expulso do espaço público. O modelo era de exclusão e exílio do doente para purificar a cidade. Posteriormente o exílio é substituído pela vigilância, o doente deveria permanecer em local isolado (FOUCAULT, 1993).

Ainda tinha muito a evoluir, mas na Inglaterra do início do século XIX já havia uma preocupação com a saúde da população e sua relação com a economia. Quanto custavam as doenças e as mortes para o estado. Os germânicos usavam a polícia médica para vigiar e controlar as epidemias. Neste período as maiores preocupações eram com as cidades, como limpeza de ruas, drenagem de eflúvios, abastecimento de água, fiscalização de alimentos, quarentena de epidemias e internação de pobres nos recém-criados hospitais (SAMPAIO *et. al*, 2016).

Para entender a saúde pública no Brasil precisamos voltar ao período do Brasil colonial, onde os profissionais que exerciam a Medicina eram físicos e cirurgiões-barbeiros. Em um país tão grande e com poucos profissionais habilitados, outros profissionais exerciam a Medicina, como: boticários, aprendizes de cirurgião, rezadores, curandeiros, parteiras, curiosos e entendidos, entre outros (GURGEL, 2010).

Em 1768 é fundado no Rio de Janeiro o Hospital Real Militar, mostrando um interesse do Estado: a doença não é uma questão médica, mas uma questão militar. Em 1808, com a chegada da Corte portuguesa ao Brasil são criadas duas escolas médicas na Bahia e no Rio de Janeiro para formar profissionais de saúde. Durante o século XIX, o objeto da Medicina passa da cura para a prevenção, a ação médica para se preocupar com a fonte da doença, com o meio físico, a conduta moral, os hábitos de higiene. Em 1841 foi criado o Hospício Dom Pedro II como espaço de atenção à saúde dos alienados e isolamento total (SAMPAIO *et. al*, 2016).

No governo Rodrigues Alves (1902-1906), Oswaldo Cruz assume a Diretoria Geral de Saúde Pública. Em 1904, ele propôs uma reforma na organização sanitária, ampliando as competências. Instituiu a vacinação contra a varíola como obrigatória, que apesar das arbitrariedades e abusos, conseguiu erradicar a varíola da cidade do Rio de Janeiro, que fortaleceu o modelo e o tornou hegemônico, usada na saúde pública durante décadas (PAIM, 2003).

Após a Segunda Grande Guerra (1939-1945), os países do chamado terceiro mundo iniciam uma recuperação financeira e social. No Brasil é uma época de grandes realizações, como a inauguração de Brasília em 1961, mas em 1º de abril de 1964, o presidente João Goulart é deposto por um golpe militar e começa uma ditadura que vai durar mais de duas décadas. O início dos anos 1970 é de crescimento, com o chamado “Milagre brasileiro”, mas que dura pouco. Em 1974, a inflação volta, a miséria aumenta, e a saúde é muito afetada, por exemplo, a mortalidade infantil em São Paulo, estado mais rico da nação, atinge 10% (NUNES, 2012).

Com a reabertura política iniciada em 1974, novos atores sociais começam a questionar o modelo de saúde vigente no país. Ao mesmo tempo, cresce a centralização hospitalar e a dicotomia assistência x prevenção. O processo de redemocratização e as novas propostas políticas trazem a ideia de descentralização e desconcentração de ações que irá compor o novo modelo de atenção à saúde. A esse movimento social que participavam profissionais de saúde e sociedade civil, deu-se o nome de Reforma Sanitária, que marchava

para a criação do SUS (Sistema Único de Saúde) dentro da Constituinte de 1988 (GONÇALVES; MORAES, 2007).

2.1.1. SUS (Sistema Único de Saúde).

Foi a partir da VIII Conferência Nacional de Saúde, palco de debates sobre: Saúde como direito de todos; Reformulação do sistema de saúde e seu financiamento; Garantia de participação dos setores organizados da sociedade, que se criou as bases político-institucionais do SUS (Sistema Único de Saúde), baseado nos seguintes princípios e diretrizes (BRASIL, 1988) e (GONÇALVES; MORAES, 2007):

- Descentralização na gestão dos serviços – Criar vários centros de poder, principalmente pela municipalização, dando poder da gestão aos estados e municípios decidirem as suas prioridades em saúde;
- Integralização das ações – serviços integrais preventivos e curativos, dando força a prevenção de doenças e promoção da saúde;
- Regionalização e hierarquização das unidades prestadoras de serviços – o usuário só deverá ser encaminhado para unidades mais complexas do sistema quando necessário, mas que possa voltar a sua unidade de origem para o acompanhamento;
- Participação da população – Inclusão representativa da população e dos trabalhadores de saúde no processo decisório e no controle dos serviços;
- Fortalecimento do papel do município;
- Universalização em relação à cobertura populacional – igualdade de acesso aos serviços, o Estado como provedor da atenção em saúde;
- Equidade quanto ao acesso – além de ter acesso, todos devem acessar a mesma qualidade de atenção. Todos acessando o mesmo conjunto de ações, à mesma rede de serviços, sem diferenças de classe social, credo, cor, lugar onde vive etc.

Enfim, passamos de um modelo de atenção excludente, em que poucas pessoas tinham acesso à atenção em saúde, para um modelo incluyente, em que o acesso, a partir de 1988, foi universalizado (GONÇALVES; MORAES, 2007).

Foi a Constituição de 1988 que incorporou mudanças no Sistema Público de Saúde, originando assim o SUS (Sistema Único de Saúde). O SUS encontra-se no Título VIII, da Ordem Social, capítulo II, da seguridade social.

A Constituição Federal de 1988, define:

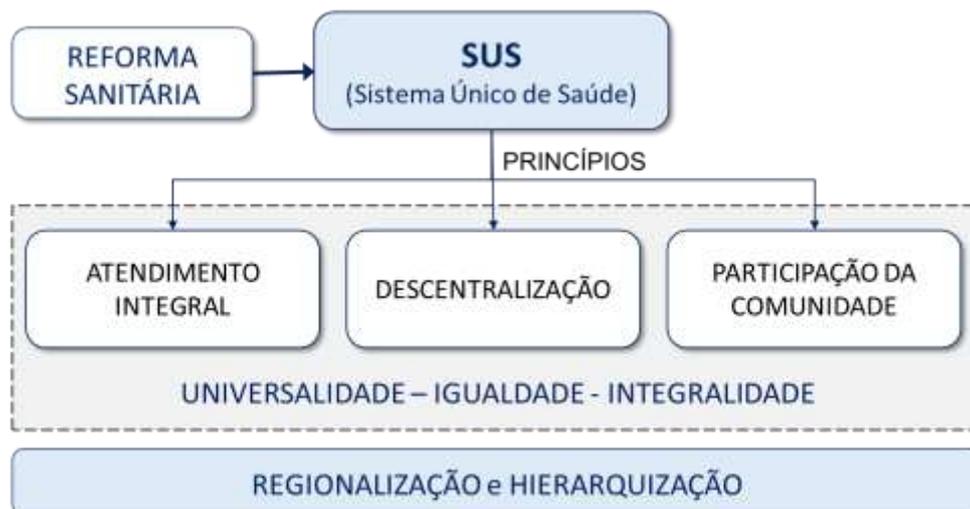
Artigo 196 - A saúde é direito de todos e dever do estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e outros agravos e ao acesso universal e igualitário às ações e serviços para sua promoção, proteção e recuperação (BRASIL, 1988).

Artigo 197 – São de relevância pública as ações e serviços de saúde, cabendo ao poder público dispor, nos termos da lei, sobre sua regulamentação, fiscalização e controle, devendo sua execução ser feita diretamente ou através de terceiros e, também, por pessoa física ou jurídica de direito privado (BRASIL, 1988).

Artigo 198 – As ações e os serviços públicos de saúde integram uma rede regionalizada e hierarquizada e constituem um sistema único, organizado de acordo com as seguintes diretrizes: descentralização, atendimento integral e participação da comunidade.

A Figura 7 apresenta o modelo de funcionalidades do SUS com as suas principais características e diretrizes.

Figura 7 – Modelo de Funcionalidades do SUS.



Fonte: Elaborado pelo autor.

Em 2018 o SUS completou 30 (trinta) anos, um marco na saúde pública brasileira, muito comemorado no 12º ABRASCÃO – Congresso Brasileiro de Saúde Coletiva, realizado de 26 a 29 de julho no Campus da Fiocruz com a participação de quase 10.000 pessoas, isto é, uma mobilização significativa da sociedade brasileira pela luta de melhores dias para a saúde, com debates, pesquisas científicas e decisões importantes.

A Revista POLI – Saúde, Educação e trabalho, em sua edição especial de julho de 2018, divulgou números expressivos fornecidos pelo Ministério da Saúde do Brasil da atuação do SUS em 2017, como:

- Dos 209,2 milhões de habitantes no Brasil, 162 milhões dependem exclusivamente do SUS, isto é, mais de 77% da população brasileira;
- Mais de 120 milhões de vacinas aplicadas de 25 tipos diferentes;
- Quase 300 milhões de atendimentos de urgência e emergência;
- Mais de 42 mil equipes de estratégia de saúde da família, cobertura de 67,7% do país.
- Mais de 322 mil hospitais ativos;
- Mais de 11 milhões de internações, entre outros números expressivos.

2.1.2. Saúde Pública x Saúde Coletiva.

A Saúde Pública tem um caráter sanitário, com ações e serviços para prevenir doenças ou situações que coloquem em risco a saúde da população. O termo saúde pública é mais conhecido e mais usado do que saúde coletiva, pois é dever do estado garantir esses serviços e políticas.

A Saúde Coletiva é um movimento sanitário de caráter social que surgiu no SUS. Esse movimento é formado pela integração das ciências sociais com as políticas de saúde pública. Neste campo são identificadas variáveis sociais, econômicas e ambientais que possam determinar cenários de epidemias em uma região, auxiliando com dados epidemiológicos elaborar políticas de prevenção.

Em termos comparativos podemos dizer que enquanto a saúde pública se preocupa com a saúde individualizada, preocupada com a presença da doença, a saúde coletiva está preocupada com o processo coletivo e não só individual. Enquanto o enfoque da saúde pública está no estado, programas e indivíduo, a saúde coletiva tem enfoque na participação popular e nas demandas da população.

2.2 EPIDEMIOLOGIA

O termo epidemiologia não é fácil de definir, a temática é dinâmica e o seu objeto é complexo. Simplificando, pode-se conceituá-la como: Ciência que estuda o processo saúde-doença em coletividades humanas, analisando a distribuição e os fatores determinantes das enfermidades, danos à saúde e eventos associados à saúde coletiva, propondo medidas específicas de prevenção, controle, ou erradicação de doenças, e construindo indicadores que

servam de suporte ao planejamento, administração e avaliação das ações de rotina, em consonância com as políticas de promoção da saúde (ROUQUAYROL, 1989).

A Epidemiologia está preocupado com eventos em grande escala que envolvam doença e não doença de comunidades, grupos demográficos ou classes sociais. A ausência de saúde é estudada sob a forma de doenças infecciosas como: malária, doença de chagas ou verminoses, doenças não infecciosas como diabetes e outras, e agravos à integridade física como acidentes de trânsito, homicídios e suicídios. Estuda ainda o processo saúde-doença ou saúde-adoecimento (ROUQUAYROL, 2012).

A Associação Internacional de Epidemiologia (IEA) define Epidemiologia como: “O estudo dos fatores que determinam a frequência e a distribuição das doenças nas coletividades humanas. Enquanto a clínica estuda a doença no indivíduo, analisando caso a caso, a Epidemiologia estuda os problemas de saúde em grupos de pessoas, às vezes grupos pequenos, mas na maioria das vezes, populações numerosas. Os principais objetivos são: [a] Descrever a magnitude dos problemas de saúde nas populações humanas; [b] Gerar dados para o planejamento, controle e tratamento das doenças; [c] Identificar fatores etiológicos na gênese das enfermidades (ROUQUAYROL, 2012).

A Epidemiologia é um método científico com o objetivo de descrever como as doenças se distribuem nas populações humanas. Os estudos epidemiológicos são classificados como (LIMA; SOARES; SANTANA, 2016):

Descritivos – estudos que se propõem a descrever como as doenças ocorrem no tempo, lugar e pessoas, sem querer entender os fatores que influenciam na ocorrência das doenças.

Analíticos – estudos que pretendem entender quais são os fatores (determinantes, de risco ou variáveis independentes) que podem influenciar na ocorrência da doença.

Os métodos epidemiológicos podem ser classificados quanto à forma de seleção da amostra ou de acordo com a relação temporal entre a medição da exposição e a verificação do desfecho (LIMA; SOARES; SANTANA, 2016).

A Figura 8 descreve os três tipos de métodos epidemiológicos.

Figura 8 – Métodos epidemiológicos.



Fonte: Elaborado pelo autor a partir de (LIMA; SOARES; SANTANA, 2016).

Existe ainda a epidemiologia clínica, mas está mais preocupada com o indivíduo. A Epidemiologia Clínica é a ciência que faz predições sobre pacientes individuais, utilizando a contagem dos eventos clínicos em grupos de pacientes semelhantes e valendo-se de métodos científicos sólidos para garantir que as predições sejam corretas. O objetivo da epidemiologia clínica é desenvolver e aplicar métodos de observação clínica que conduzam a conclusões válidas, evitando o engano por erros sistemáticos e aleatórios (FLETCHER; FLETCHER; FLETCHER, 2014).

A Figura 9 apresenta as ciências da saúde e suas relações complementares, através dos campos de pesquisa e o foco de cada campo, destacando a posição da Epidemiologia e da Epidemiologia Clínica.

Figura 9 – As ciências da saúde e suas relações complementares.



Fonte: FLETCHER; FLETCHER; FLETCHER, 2014.

Epidemias e Endemias estão nesse campo de estudo, assim conceituadas:

- A Epidemia é a ocorrência da doença em grande número de pessoas ao mesmo tempo. O termo vem de “*epidemeion*”, no sentido de visitar, temporário ou provisório.

- A Endemia é a ocorrência coletiva de uma doença em um longo período e que mantém a incidência constante. O termo “*endemeion*”, no sentido de habitar, residir por longas datas e se instalar.

2.2.1. Indicadores de Saúde.

Para medir e avaliar a saúde das populações ou comunidades de uma região ou país, foi necessária a criação de vários indicadores internacionais, a Organização mundial de Saúde (OMS) recomendou em 1957 o uso de dados de óbitos, para assim medir de forma indireta os indicadores de saúde, então surgem os “Coeficientes de mortalidade”.

Segundo Rouquayrol (2012), em 2002 a Organização Pan-Americana de Saúde (OPAS) junto com a Rede Interagencial de Informações para a Saúde (Ripas/MS), lançam uma das publicações mais completas sobre indicadores de saúde, descrevendo o significado de uma centena de indicadores categorizados como demográficos, socioeconômicos, de mortalidade, de morbidade, de fatores de risco, de recursos e de cobertura. A publicação conceitua: “*Em termos gerais, os indicadores são medidas-sínteses que contém informações relevantes sobre alguns atributos e dimensões do estado de saúde, bem como do desempenho do sistema de saúde. Vistos em conjunto, devem refletir a situação sanitária de uma população e servir para a vigilância das condições de saúde*”.

Os indicadores de mortalidade são muito usados no Brasil e no mundo. Estes indicadores representam estatisticamente os óbitos ocorridos sob vários ângulos: pelo local do óbito (na residência, nos hospitais, na via pública, etc.), pelas características do sujeito (idade, sexo, cor, profissão, grau de instrução, etc.), pela geografia do óbito (bairro, distrito, município, estado, país, etc.) e pelos motivos (violência, acidentes de trânsito, doenças, causas naturais, etc.). Através da combinação destes vários tipos de indicadores a gestão pública é capaz de conhecer, acompanhar e planejar ações de combate à mortalidade.

O termo coeficiente representa a relação entre o número de casos reais e o número de possíveis casos. Por exemplo: o Coeficiente de Mortalidade representa o quociente entre os casos reais de óbitos e o número de expostos ao risco de morrer. No caso da Mortalidade

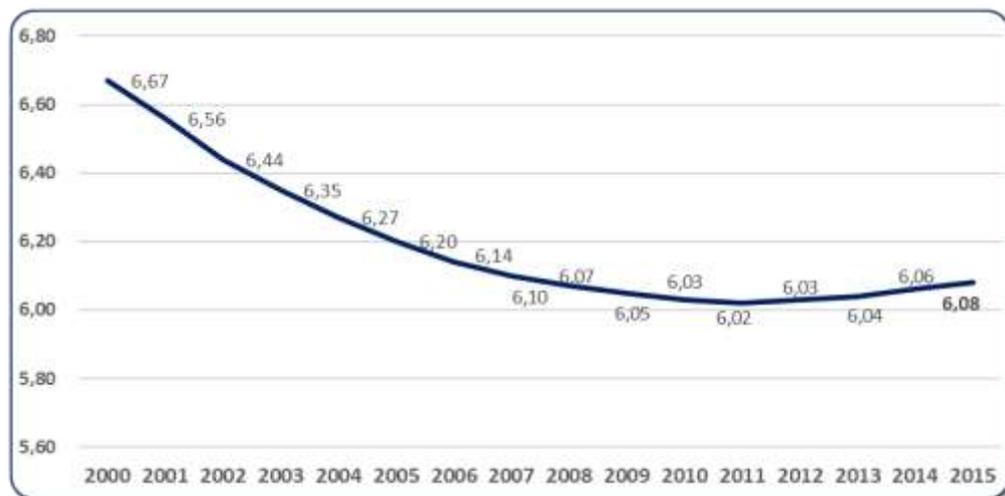
Infantil, os expostos são as crianças de até um ano, mas na Mortalidade Geral, os expostos aos riscos são todos os indivíduos da população em um determinado ano (ROUQUAYROL, 2012).

O Coeficiente de Mortalidade Geral (CMG) é calculado dividindo-se todos os óbitos de todas as causas, em um determinado ano, pela população naquele ano, ao final multiplicando-se por 1000. É um indicador muito utilizado na saúde pública para estudos comparativos.

$$CMG = \frac{\text{Número Total de Óbitos, no período}}{\text{População Total, no período}} \times 1000$$

O Gráfico 2, representa o Coeficiente de Mortalidade Geral no Brasil entre os anos de 2000 e 2015 segundo dados do IBGE. Neste gráfico, observa-se uma redução constante até 2015.

Gráfico 2 – Coeficiente de Mortalidade Geral no Brasil de 2000 a 2015.



Fonte: Elaborado pelo autor com dados do IBGE.

A Taxa de Mortalidade Infantil (TMI) é calculada dividindo-se o número de óbitos de crianças com menos de um ano pelos nascidos vivos naquele ano em uma determinada região, ao final multiplica-se por 1000.

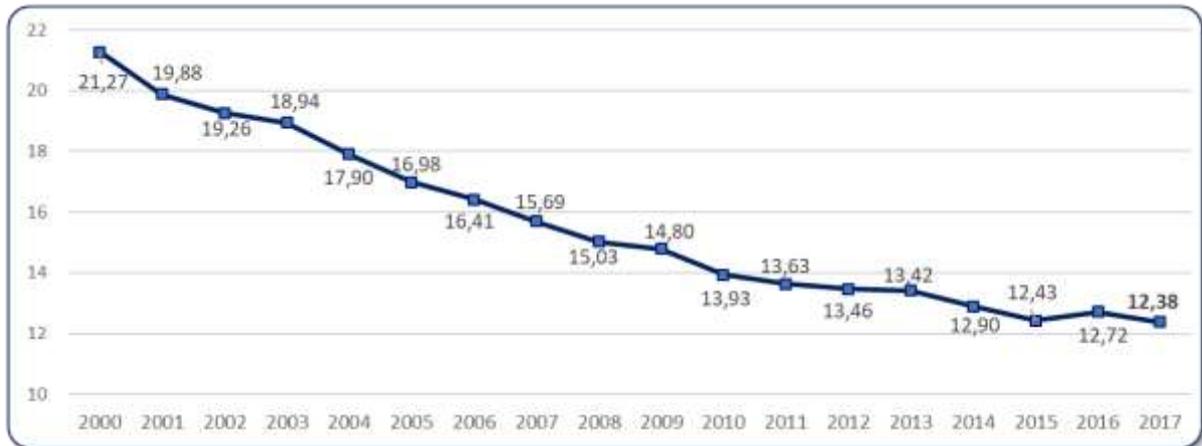
$$TMI = \frac{\text{Número de óbitos de crianças menores de 1 ano de idade, no período}}{\text{Número de nascidos vivos, no período}} \times 1000$$

O TMI é classificado como do grupo coeficientes gerais, pois seu objetivo principal é avaliar o estado sanitário geral de uma população, em associação com outros

indicadores de promoção da saúde, como: escolaridade, renda, disponibilidade da água, saneamento básico, moradia digna, entre outros (ROUQUAYROL, 2012).

O Gráfico 3, representa a Taxa de Mortalidade Infantil (TMI) no Brasil entre os anos de 2000 e 2017 segundo dados do Datasus, sistema SIM, apresentando redução significativa de 41,8% da TMI durante esse período.

Gráfico 3 – Taxa de Mortalidade Infantil (TMI) no Brasil de 2000 a 2017.

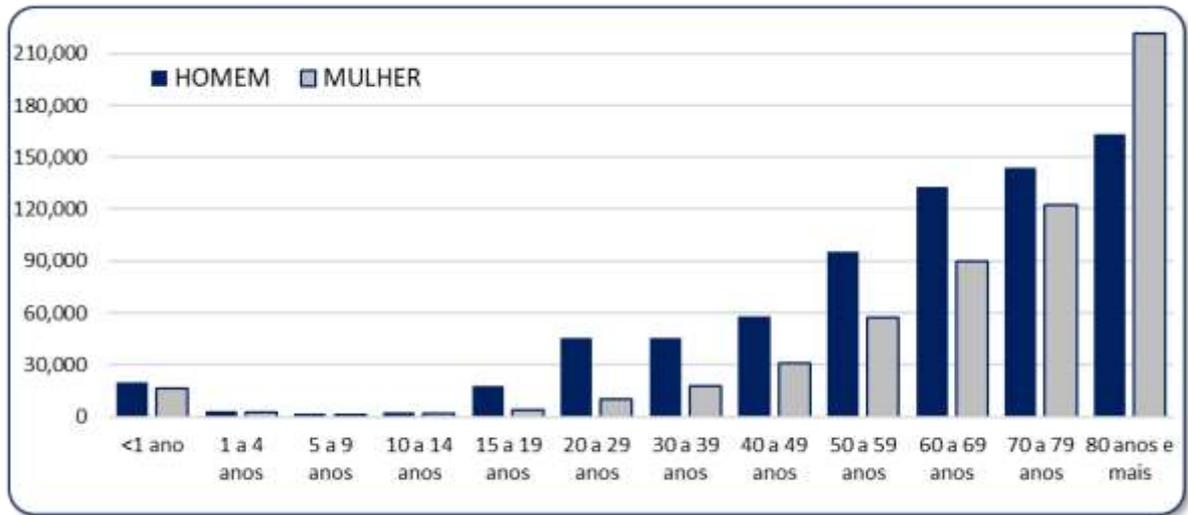


Fonte: Elaborado e calculado pelo autor a partir de dados do Datasus/SIM/SINASC.

A epidemiologia busca entender por que os índices caem ou sobem, quais são os fatores de maior impacto, quando são fatores positivos, serão mantidos e incentivados, já os fatores negativos, serão atacados para conseguir sua redução ou eliminação. No caso específico da TMI apresentada no Gráfico 3, observa-se uma queda considerável na mortalidade infantil no Brasil entre 2000 e 2017, provavelmente pela melhoria nas condições de vida da população no período, com mais saneamento básico, água potável nas residências, reduzindo assim os óbitos infantis causados por diarreias.

Em estudos epidemiológicos podemos extrair conhecimentos variados a partir da análise dos resultados gráficos estatísticos. Por exemplo quando combinamos algumas variáveis para inferir sobre os resultados. O gráfico 4, apresenta o Mortalidade Geral no Brasil no ano de 2017 categorizado por sexo e faixa etária.

Gráfico 4 –Mortalidade Geral no Brasil em 2017.

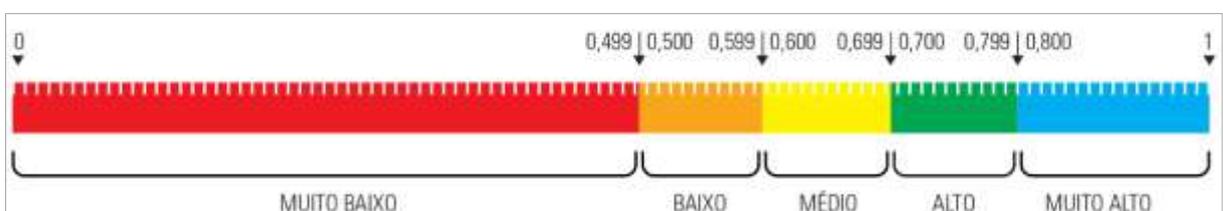


Fonte: Elaborado e calculado pelo autor a partir de dados do Datasus/SIM.

Observe nos dados acima, a mortalidade Infantil (até 1 ano de idade) se destaca no início da vida dos brasileiros com volumes de óbitos equivalente a todas as mortes ocorridas de 1 a 19 anos. Outra informação se destaca, os homens morrem muito mais no Brasil do que as mulheres até os 79 anos, com destaque negativo para a faixa entre 15 e 19, onde os homens morrem 4,7 vezes mais do que as mulheres, e na faixa entre 20 e 29 anos, onde os homens morrem 4,6 vezes mais do que as mulheres, sugerindo aqui a vulnerabilidade dos homens jovens quanto à segurança pública que apresenta altos índices de homicídios de jovens em todo território nacional.

O IDH (Índice de Desenvolvimento Humano) varia de 0 (zero) a 1 (um) e mede o grau de desenvolvimento de uma sociedade ou comunidade a partir de três indicadores básicos: saúde (expectativa de vida), educação (anos de escolaridade) e renda (renda per capita). Em 2017 os países mais desenvolvidos atingiram IDHs próximos de 1,0: Noruega (0,949), Austrália (0,939), Suíça (0,939) e Alemanha (0,926), já os menores índices foram dos países com índices abaixo de 0,5: Burundi (0,404), Burkina Faso (0,402), Chade (0,306), Níger (0,353) e República Centro-Africana (0,352). O Brasil ficou na posição 79ª, com um IDH de 0,754 considerado “Alto”, conforme a classificação apresentada na Figura 10 (MOTA, 2018).

Figura 10 – Faixas do IDH (Índice de Desenvolvimento Humano).



Fonte: Atlas do Desenvolvimento Humano no Brasil, 2018.

No Brasil é utilizado o IDHM (Índice de Desenvolvimento Humano Municipal) para calcular o IDH dos municípios brasileiros usando a mesma metodologia do IDH. Dessa forma foi possível correlacionar o IDHM dos municípios do Ceará com o TMI destes municípios, permitindo inferir que quanto maior o IDHM, menor será o TMI.

Na Estatística a análise bivariada tem como objetivo o estudo das relações entre duas variáveis (associações para variáveis qualitativas e correlações para variáveis quantitativas). Para variáveis quantitativas são usadas representações gráficas como o diagrama de dispersão, além de medidas de correlação como a covariância e o coeficiente de Pearson (BELFIORE, 2015).

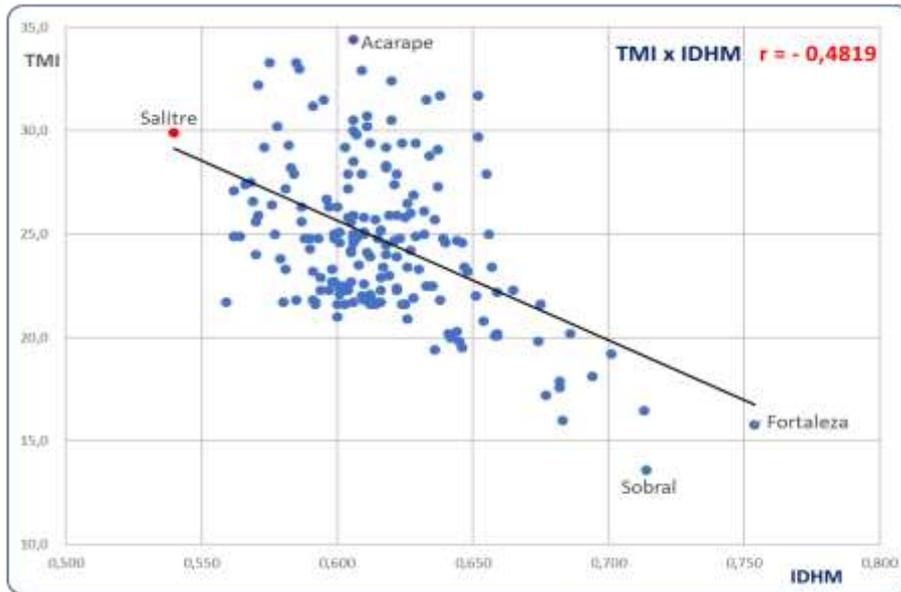
Segundo Larson e Farber (2010), interpretar a correlação usando um diagrama de dispersão pode ser subjetivo. Uma maneira mais precisa de medir o tipo e a força de uma correlação linear entre duas variáveis é calcular o coeficiente de correlação. O coeficiente de correlação (r) é uma medida de força e direção entre duas variáveis. O nome formal de “ r ” é “coeficiente de correlação produto-momento de Pearson”. Recebeu esse nome em homenagem ao estatístico inglês Karl Pearson (1857-1936), que propôs esta fórmula em 1896.

O coeficiente de correlação tem duas características especiais. A primeira, por razões matemáticas, trata-se de um número único que varia de -1 a 1. Uma correlação de 1 é chamada de correlação perfeita, isto é, qualquer alteração em uma das variáveis está associada a uma alteração equivalente na outra variável e no mesmo sentido. Uma correlação de -1, ou correlação negativa perfeita, significa que uma alteração em uma variável está associada a uma alteração equivalente na outra variável no sentido oposto. A segunda característica, é que não está ligado a nenhuma unidade, podemos calcular correlação entre altura e peso, embora a altura seja medida em centímetros e o peso em quilogramas. O coeficiente de correlação faz uma coisa aparentemente milagrosa: reduz uma complexa bagunça de dados medidos em unidades diferentes numa única e elegante estatística descritiva (WHEELAN, 2016).

A regressão linear é uma das técnicas estatísticas mais utilizadas. É uma ferramenta que sempre será utilizada em pesquisa quantitativas ou no seu cotidiano. Por exemplo: quando você lê em um jornal que pessoas com mais primos ou irmãos são mais felizes, ou que cortar pela metade o consumo de vitamina B3 duplica seu risco de ter pé de atleta, nestes casos, você está usando os resultados de uma regressão linear (ELLENBERG, 2015).

O Gráfico 5 apresenta o diagrama de dispersão entre as variáveis quantitativas TMI e IDHM de todos os municípios do estado do Ceará em 2015. Este gráfico nos faz supor que, quanto maior o IDHM, menor será a TMI, caracterizando uma regressão linear negativa, confirmada pelo índice negativo de correlação de Pearson ($r = -0,4819$), embora esse valor próximo de 0,5 (meio) indique uma correlação fraca.

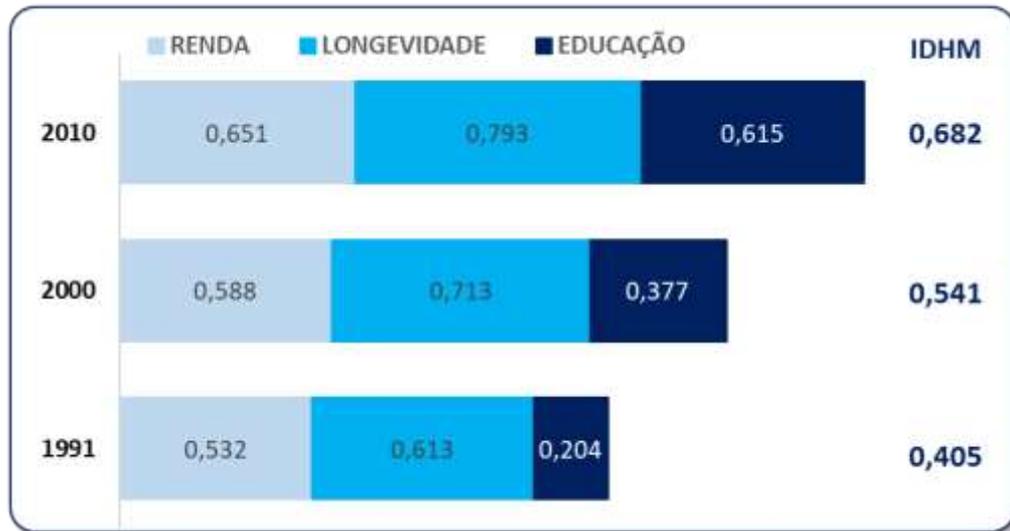
Gráfico 5 – Correlação IDHM x TMI dos municípios do Ceará em 2015.



Fonte: Atlas do Desenvolvimento Humano no Brasil, 2018.

O gráfico acima nos mostra que existe correlação negativa e fraca entre as variáveis, próximo de -0,5 (meio), mas quando calculamos o índice de correlação de Pearson usando somente as localidades da Região Metropolitana de Fortaleza esse índice sobe para -0,8680 configurando uma correlação forte, isto é, na RMF, podemos afirmar com maior segurança que, quanto maior a qualidade da saúde, educação e renda, menor será a taxa de mortalidade infantil. Fortaleza e Sobral possuem as menores TMI e os melhores IDHM, enquanto a localidade de Salitre tem baixo IDHM e alta TMI.

No estado do Ceará o IDHM vem crescendo nas últimas avaliações censitárias, de 0.405 em 1991 para 0.541 em 2000 e no último censo de 2010 atingiu 0.682, classificação considerada “Média” (entre 0.600 e 0.699) e se aproximando da classificação considerada “Alta” (entre 0.700 e 0.799). A dimensão que mais contribui para o IDHM do Ceará é a Longevidade, com índice de 0.793, seguida da Renda, com índice de 0.651 e de Educação, com índice de 0.615, conforme mostra o Gráfico 6, abaixo.

Gráfico 6 – Evolução do IDHM do estado do Ceará.

Fonte: Atlas do Desenvolvimento Humano no Brasil, 2018.

2.3 MORTALIDADE INFANTIL

No início do século 20, o entendimento da complexidade das questões relacionadas com a mortalidade infantil passou a preocupar. Observou-se que ela estava vinculada não só às condições de vida da população, mas também na existência de uma relação mãe-filho do qual depende a saúde e o bem-estar da criança. O primeiro ano de vida passou então a ser dividido em faixas etárias, com o objetivo de refinar e analisar os óbitos infantis. O termo mortalidade neonatal foi introduzido em 1938 e na década de 50 foram ratificadas as classificações que ainda vigoram até hoje: natimortalidade ou mortalidade fetal, que considera os óbitos de fetos com mais de 22 semanas, neonatal precoce, perinatal, neonatal e pós-neonatal (ARMSTRONG, 1986).

No contexto da mortalidade infantil é importante conceituar: [a] O nascimento é a expulsão ou extração do organismo materno de um feto independente do cordão ter sido cortado ou de a placenta estar inserida. Fetos pesando menos de 500 gramas não são viáveis, por isso não são considerados para estatísticas perinatais. [b] A vida é considerada presente ao nascimento quando o recém-nascido respira ou apresenta qualquer outra evidência vital. [c] O Nascido vivo é o produto de um nascimento no qual existe evidência de vida ao nascer. [d] A morte infantil é o óbito ocorrido no primeiro ano de vida e divide-se em neonatal (primeiros 28 dias incompletos) e pós-neonatal (nos meses seguintes até um ano).

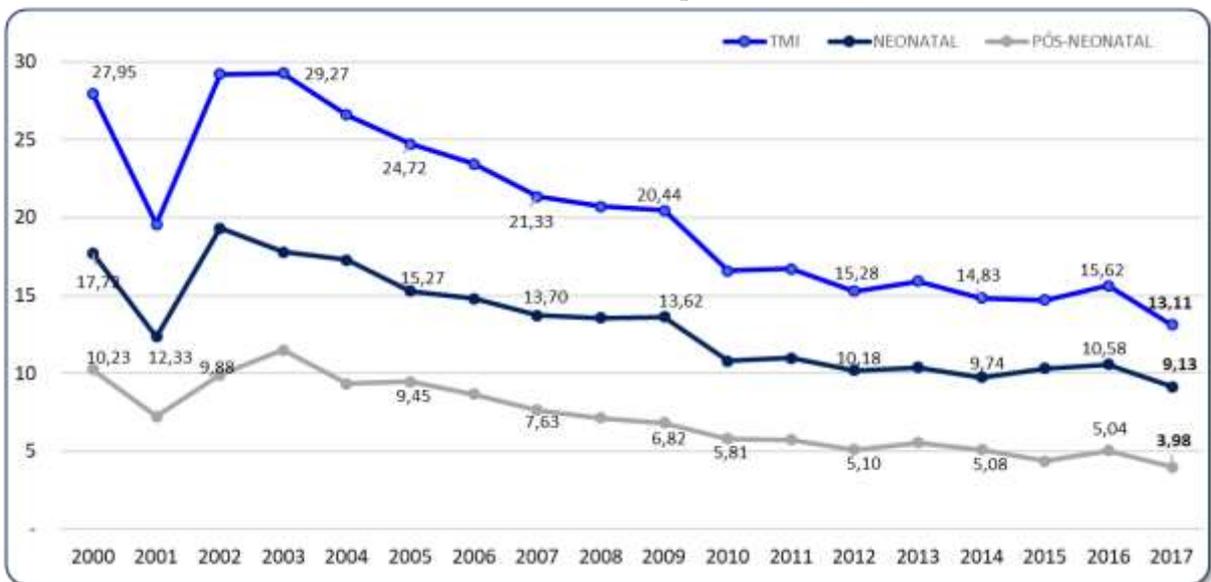
O Brasil, acompanhando a tendência mundial, experimentou uma redução marcante na mortalidade infantil, mas esse declínio não foi percebido na taxa de mortalidade neonatal, o que sugere falhas na qualidade de assistência obstétrica e neonatal. O componente neonatal da

mortalidade infantil reflete muito mais a qualidade dos serviços ofertados às mães e recém-nascidos durante o período ante natal, parto e neonatal (OPAS, 2006).

Infelizmente, notícias recentes dão conta do retrocesso nas conquistas de redução da taxa de mortalidade infantil nos últimos 25 anos, retrocessos causados principalmente pelas medidas de austeridade que os programas sociais sofreram desde 2016 juntamente com crises econômicas, políticas e sociais que nosso país sofreu nos últimos anos.

Na Região Metropolitana de Fortaleza (RMF), abrangendo 11 (onze) municípios, podemos observar no Gráfico 7 que a TMI teve uma queda considerável de 2002 a 2017, mas entre 2010 e 2016 não houve redução, o que é pior, em 2016 observa-se um crescimento da TMI e da TMI Neonatal, causando um alerta para os gestores e profissionais da saúde pública no estado do Ceará, embora os dados de 2017 anuncie uma nova redução.

Gráfico 7 - Taxa de Mortalidade Infantil e Componentes. RMF, 2000 a 2017.



Fonte: Elaborado e calculado pelo autor a partir de dados do Datasus/SIM/SINASC.

As adequadas assistências durante a pré-gravidez, durante a gestação, ao parto e ao recém-nascido são consideradas intervenções imprescindíveis à promoção da saúde neonatal. No entanto, profundas desigualdades nacionais e regionais, o acesso limitado aos cuidados necessários, as falhas na distribuição dos leitos intensivos neonatais e a fragilidade dos serviços prestados, contribuem como fatores associados ao risco de morte infantil no período neonatal (OPAS, 2006; BARBOSA, 2004; RIBEIRO; SILVA, 2000; SCHRAMM; SZWARCOWALD, 2000).

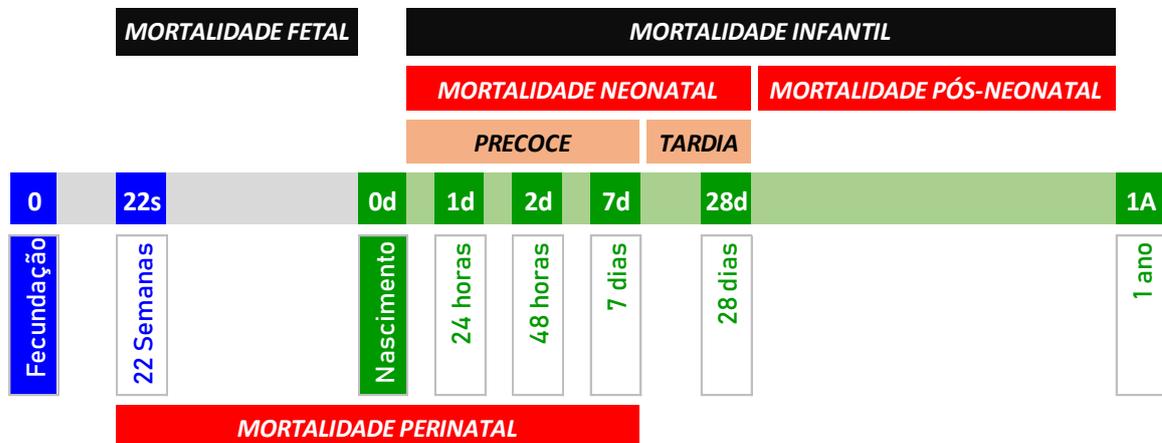
O termo Mortalidade Infantil está associado aos óbitos ocorridos em crianças nascidas vivas que tiveram óbito entre um minuto de vida a um ano de vida. A epidemiologia

acompanha esses eventos através dos indicadores Taxa da Mortalidade Infantil, Taxa da Mortalidade Neonatal e Taxa de Mortalidade Pós-Neonatal. Esses indicadores medem os riscos de morte em crianças menores de um ano. Outro indicador importante de avaliação é a Taxa de Mortalidade Fetal, isto é, antes do nascimento.

Os óbitos fetais são de grande importância por serem em maior número do que os óbitos neonatais precoces (menos de 7 dias de vida). Estas mortes traduzem falhas relacionadas ao pré-natal, ao manejo obstétrico e as horas que antecedem o parto. Grande parte dos óbitos fetais como dos óbitos neonatais precoces poderiam ser evitadas em mais da metade dos casos (SILVA *et.al.*, 2016).

A Mortalidade infantil pode ser medida e avaliada através de vários índices, o mais importante e mais conhecido é o TMI (Taxa de Mortalidade Infantil), mas existem outros indicadores importantes, como: TMF (Taxa de Mortalidade Fetal), TMN (Taxa de Mortalidade Neonatal) e a TMP (Taxa de Mortalidade Pós-neonatal), onde cada uma tem suas implicações, características, causas e efeitos. A Figura 11 apresenta a distribuição cronológica que avalia cada uma dessas taxas.

Figura 11 – Cronologia das Taxas de Mortalidade.



Fonte: Elaborada pelo autor.

Na figura acima podemos observar que as classificações são de ordem cronológicas, com taxas de acompanhamento para cada fase, desde a fecundação até o primeiro ano de vida da criança. Elas se classificam assim:

- A Mortalidade Fetal é calculada pelo número de óbitos fetais (ocorridos a partir da 22ª semana completa de gestação, ou 154 dias ou fetos com peso igual ou superior a 500g ou estatura a partir de 25cm) por mil nascimentos totais, em uma região e período de estudo.

- A Mortalidade Perinatal envolve a Mortalidade Fetal e a Mortalidade Neonatal Precoce, quando a morte ocorre a partir da 22ª semana de gestação até o 7º dia de nascido.

- A Mortalidade Infantil é classificada como: Mortalidade Neonatal (do nascimento até o 28º dia de vida) ou a Mortalidade Pós-neonatal (do 29º dia de vida até completar o primeiro ano de vida).

- Por sua vez, a Mortalidade Neonatal também pode ser classificada como Neonatal Precoce (do nascimento até o 7º dia de vida) ou Neonatal Tardia (do 8º dia até o 28º dia de vida).

Em cada uma das classificações acima é possível identificar características predominantes quanto à causalidade. Por exemplo: os fatores idade materna, escolaridade materna, baixa renda familiar, história reprodutiva desfavorável e qualidade da atenção pré-natal. Nos países onde estas condições são mais precárias e desfavoráveis, há maior ocorrência de óbitos perinatais e mais da metade é representada pelos óbitos fetais (SILVA *et.al.*, 2016).

2.3.1. Principais causas da mortalidade Infantil.

Como foi visto na classificação acima, a epidemiologia estuda todas as fases do ciclo de vida da criança, desde a fecundação até o primeiro ano de vida. A partir dos vários estudos que já foram realizados neste campo da Saúde Pública alguns fatores causais já são conhecidos a décadas, mas a sociedade em vários países continua a ignorá-los ou tratá-los paliativamente.

Há várias décadas foi estudada a influência da educação materna no declínio da mortalidade infantil. O estudo ressalta que o melhor nível de instrução fortalece a mulher, tornando-a mais apta na prestação de cuidados aos filhos, além de permitir o desenvolvimento de novas maneiras de cuidar e fortalecer a ruptura com costumes e hábitos tradicionais (SILVA *et.al.*, 2016).

As principais causas da mortalidade neonatal precoce estão ligadas as condições de saúde da mulher durante a gravidez, acompanhamento pré-natal e o parto. Importante também é a assistência dada ao recém-nascido, já na mortalidade neonatal tardia, que vai até o 28º dia de vida, as principais causas são as doenças infecciosas.

Nos meses seguintes até o primeiro ano, as principais causas são relativas as doenças infecciosas, como diarreia, pneumonia e desnutrição. Nesta fase o combate é principalmente

através de saneamento básico, educação materna e aumento do aleitamento materno, imunização, terapia de reidratação, entre outros (LEITE *et. al.*, 2013).

2.3.2. Óbitos evitáveis.

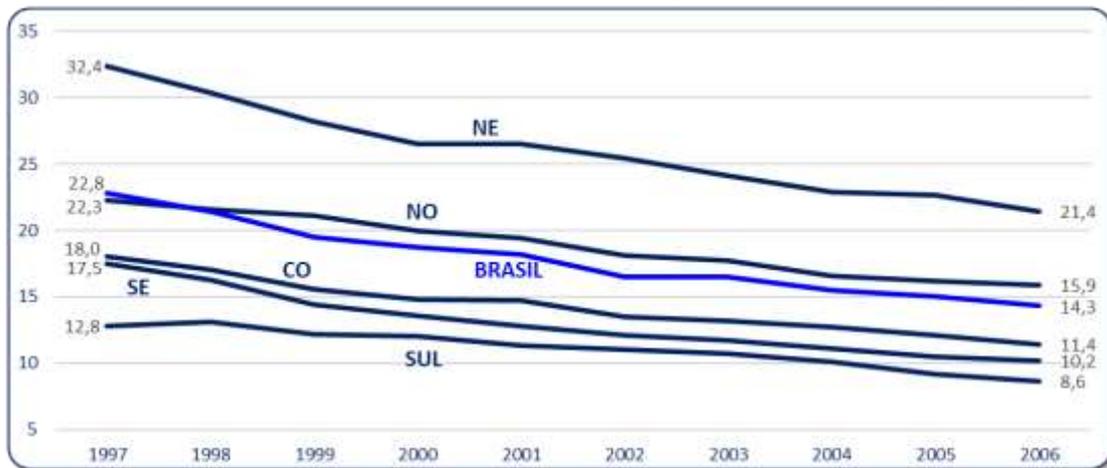
Óbitos evitáveis são aqueles que poderiam ter sido evitados com a oferta de serviços de saúde efetivos. Estes óbitos foram classificados em três grupos, abaixo descritos, que merecem análises distintas para avaliar sua evitabilidade (PORTO, 2012).

- Óbitos claramente evitáveis: são os óbitos chamados sentinelas, sua ocorrência mesmo isolada demonstra uma falha na assistência em saúde oferecida ao indivíduo, merecendo investigação.

- Uso limitado: um número de mortes, ao invés de um único evento, justificaria uma investigação mais profunda.

- Problemas médico-sociais: exige um esforço maior para esclarecer, dado que o conhecimento na época não permitiria determinar com certeza as contribuições que a saúde poderia dar para evitar essas mortes.

Sobre a frequência da ocorrência de óbitos evitáveis no país, o coeficiente de mortalidade infantil por óbitos evitáveis vem reduzindo no Brasil como um todo, embora a participação dos óbitos evitáveis tenha se mantido praticamente constante ao longo dos últimos anos do período estudado. Entretanto, essa informação não nos esclarece o suficiente a respeito do real cenário vivenciado nas diversas regiões do país. Na mesma pesquisa são apresentados resultados da medição da ocorrência dos óbitos evitáveis também por região geográfica. Podemos observar que o coeficiente vem caindo em todas as regiões, mas em taxas desiguais. A queda foi mais acentuada na região Sudeste, com uma variação no período de 42,0 % e menos marcante na região norte, com uma queda de 28,7 %, conforme apresenta o Gráfico 8 (MALTA *et. al.*, 2010).

Gráfico 8 – Coeficiente de Mortalidade infantil por causas evitáveis por região.

Fonte: Elaborado pelo autor com dados do Datasus/SIM/SINASC.

Após debates entre pesquisadores de diversas áreas do conhecimento sob a coordenação da Secretaria de Vigilância em Saúde do Ministério da Saúde do Brasil, o tema foi sistematizado quanto aos conceitos e metodologias com o propósito de construir uma lista brasileira de mortes evitáveis, segundo grupos etários (MALTA *et. al.*, 2007). As doenças, de acordo com a CID-10, foram enquadradas em três categorias, sendo a primeira dividida em quatro subcategorias de evitabilidade conforme apresenta o Quadro 1.

Quadro 1 – Lista resumida de mortes evitáveis por intervenções do SUS.

1.	Causas evitáveis
1.1.	Reduzíveis por ações de imunoprevenção
1.2.	Reduzíveis por adequada atenção à mulher na gestação e parto e ao recém-nascido
1.3.	Reduzíveis por ações adequadas de diagnóstico e tratamento
1.4.	Reduzíveis por ações adequadas de promoção à saúde
2.	Causas mal definidas
3.	Demais causas (não claramente evitáveis)

Fonte: Elaborado pelo autor a partir de Malta *et.al.*, 2007.

Entre as causas evitáveis por ações de imunoprevenção podemos citar: tuberculose, tétano neonatal, difteria, sarampo, rubéola, caxumba, meningite, hepatite B, entre outras. Quando se trata das causas oriundas da adequada atenção à mulher na gestação e parto, podemos citar: sífilis congênita, problemas na placenta e nas membranas, afecções maternas e etc. E quando as causas são decorrentes da falta de ações adequadas de atenção à saúde, podemos citar: febres por arbovírus, febres hemorrágicas, raiva, doenças bacterianas e zoonóticas, e outras doenças infecciosas.

Na pesquisa de Dias, Neto e Andrade (2017), realizada no estado do Espírito Santo, Brasil, a maioria dos óbitos ocorridos em menores de um ano entre 2006 e 2013, é evitável, independentemente do método de classificação utilizado. Destacando-se, principalmente a qualidade da assistência à mulher e à criança no período do pré-natal, do parto e do puerpério como causas evitáveis de óbitos.

2.4 MORTALIDADE MATERNA

A saúde da mulher é motivo de preocupação em várias partes do mundo, em especial nos países subdesenvolvidos onde por muito tempo foi negligenciada. O cuidado, a orientação e a educação podem evitar vários problemas de saúde, como: gravidez indesejada na adolescência pela falta de orientação sexual, HIV/AIDS, câncer de mama e do colo uterino, entre outros.

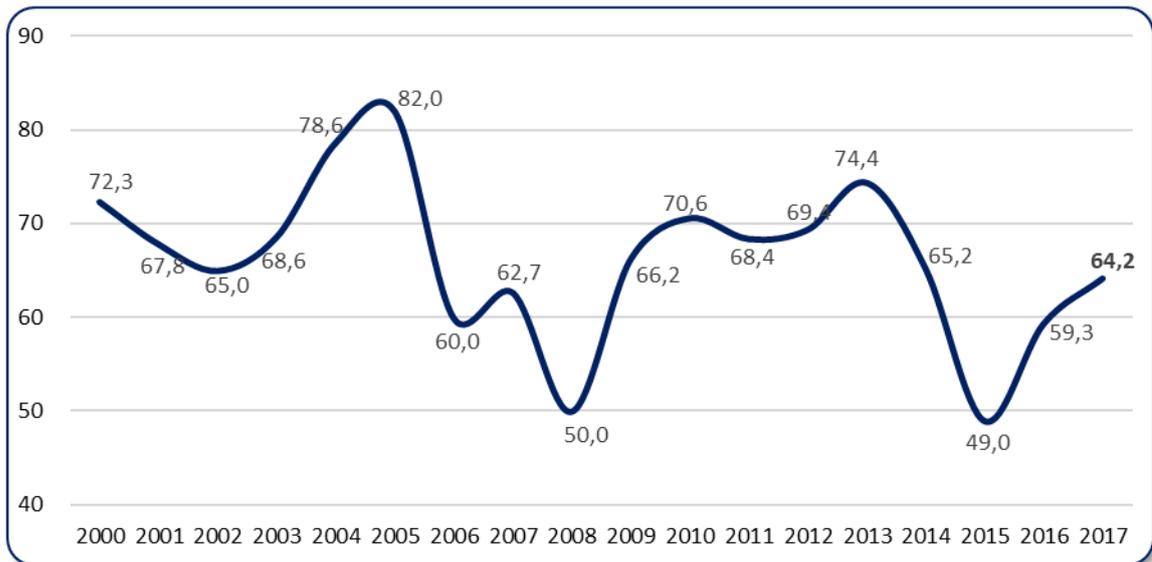
A mortalidade materna é um problema de saúde pública e se constitui um reflexo da situação da saúde da população em geral e revela-se como um indicador da realidade socioeconômica e da qualidade da assistência à saúde da mulher no seu ciclo Gravídico Puerperal (CARVALHO *et. al.*, 2016).

Segundo a OMS, a morte materna é definida como: de uma mulher durante a gestação ou dentro de um período de 42 dias após o término da gestação, independentemente da duração ou da localização da gravidez, devido a qualquer causa relacionada com ou agravada pela gestação ou por medidas tomadas em relação a ela, porém não devidos a causas acidentais ou incidentais (OMS, 1993).

A Taxa de Mortalidade Materna (TMM) é calculada dividindo-se o número de óbitos maternos de um ano pelos nascidos vivos naquele ano em uma determinada região, ao final multiplica-se por 100.000.

$$TMM = \frac{\text{Número de óbitos maternos, no período}}{\text{Número de nascidos vivos, no período}} \times 100.000$$

No Ceará, de 2000 a 2017, foram notificados 44.658 óbitos de MIF (Mulheres em Idade Fértil) e destes foram confirmadas 1.601 mortes maternas (causas obstétricas diretas, indiretas, não obstétricas, não especificadas e tardias), com uma média da Razão da Mortalidade Materna (RMM), no período de 66,3 mortes maternas por 100.000 nascidos vivos, índice considerado alto segundo parâmetros da OMS. O Gráfico 9 apresenta o desempenho da TMM no Ceará de 2000 a 2017.

Gráfico 9 – Taxa de Mortalidade Materna no Ceara, de 2000 a 2017.

Fonte: Elaborado pelo autor com dados do Datasus/SIM/SINASC.

No período analisado, observa-se uma média anual de 88,9 óbitos no ciclo gravídico-puerperal por causas obstétricas em nosso estado, com uma média de 7,4 mortes por mês. A maior TMM foi registrada em 2005 (82,0), e a menor taxa em 2015 (49,0), dados compilados do sistema Datasus/SIM/SINASC.

Os principais problemas relacionados com a alta taxa de mortalidade materna, infantil e fetal em alguns países podem estar relacionados aos seguintes fatores (CEARÁ, 2014):

- Baixo acesso ao planejamento reprodutivo;
- Baixa qualidade do pré-natal;
- Demora na identificação da necessidade de encaminhamento para o pré-natal de alto-risco;
- Falta de busca ativa de gestantes faltosas às consultas;
- Complicações relacionadas à gravidez, parto e puerpério;
- Baixa valorização das queixas e do quadro clínico;
- Demora na tomada de decisões no atendimento das intercorrências e urgências;
- Não realização das consultas de puerpério.

A cultura da Cesariana é outro fator importante de estudo, enquanto a OMS recomenda 15% de partos sejam cesarianas, o Brasil o índice é de 52%, chegando a 88% na rede privada. O Excesso de cesariana está entre os motivos que impede o Brasil de atingir as

metas de redução da mortalidade materna (CARVALHO *et. al.*, 2016). A prevalência de cesariana cresceu de 2000 a 2011 em todas as regiões do Brasil.

Uma das estratégias utilizadas pelo Ministério da Saúde/Secretaria de Vigilância em Saúde (MS/SVS) para redução da mortalidade materna foi o lançamento em 2011 da Rede Cegonha (RC), que tem como objetivo garantir às mulheres e as crianças o direito ao parto e ao nascimento seguro.

Algumas ações são imprescindíveis para a saúde da mulher, começando com a assistência pré-natal adequada, acolhimento da gestante e classificação de riscos, atenção humanizada ao parto e nascimento, puerpério e atenção integral à saúde da criança, transporte sanitário e regulação de qualidade, entre outras ações.

Quanto à mortalidade materna podemos refletir que melhorou no Brasil nos últimos anos (1990-2015), pois houve grande queda de fecundidade, universalização da atenção pré-natal a hospitalar ao parto, aumento de acesso à contracepção e aleitamento materno, e redução de hospitalizações por aborto e da subnutrição. Mantém-se em excesso a sífilis congênita, taxa de cesarianas e nascimentos prematuros (LEAL *et. al.*, 2018).

2.4.1. Near Miss.

A morte materna é um dos eventos mais raros e dramáticos de uma cadeia de acontecimentos que pode acometer qualquer mulher durante o ciclo gravídico-puerperal. Estes eventos começam quando uma mulher apresenta algumas das diversas condições potencialmente ameaçadoras da vida (CPAV). Quando não adequadamente tratada, ela pode evoluir para condições ameaçadoras da vida, em que há disfunção ou falência orgânica e que se definem a partir de critérios clínicos, laboratoriais ou de manejo. Mulheres que sobrevivem a tais eventos graves serão classificadas como “*near miss*” materno (SANTANA *et. al.*, 2018).

Quando complicações com maior potencial de evoluir mais gravemente para a morte for detectada a tempo e medidas adequadas forem tomadas pelos serviços e profissionais de saúde, há a possibilidade de se alterar a evolução do caso e a mulher pode sobreviver. Ao mesmo tempo, a chegada de uma mulher aos serviços de referência em condições de gravidade que podem levá-la à morte pode representar a desorganização dos serviços de saúde (SOUSA *et. al.*, 2008).

Pesquisa realizada em Fortaleza, concluiu que as principais ações para reduzir os eventos de *near miss* são: fortalecer a rede de referência e contra referência, realizar

capacitação profissional, melhorar a cobertura do pré-natal e investir na infraestrutura, na gestão de processos e em auditorias clínicas. E os fatores que influenciam os casos de *near miss* materno englobam desde o atraso nos cuidados até a não realização do pré-natal, cuja melhoria na gestão constitui o principal direcionamento (MONTE *et. al.*, 2018).

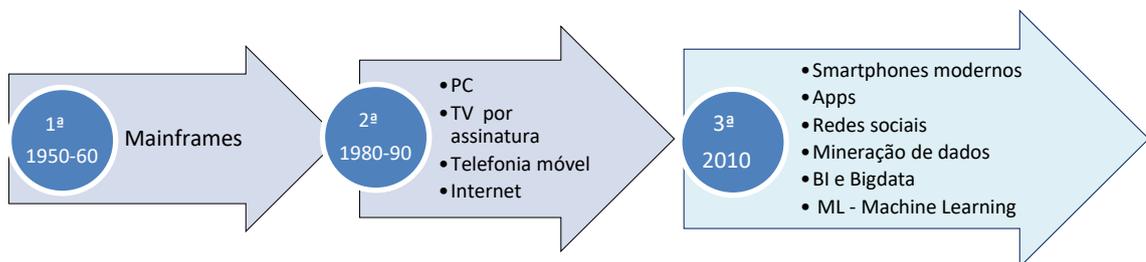
2.4.2. A importância do Pré-natal.

A assistência pré-natal é um importante componente da atenção à saúde das mulheres no período gravídico-puerperal. Práticas realizadas rotineiramente durante essa assistência estão associadas a melhores desfechos perinatais. O Ministério da Saúde recomenda que a assistência pré-natal deve ser por meio de condutas acolhedoras; através do desenvolvimento de ações educativas e preventivas e sem intervenções desnecessárias; da detecção precoce de patologias e de situações de risco gestacional; de estabelecimento de vínculo entre o pré-natal e o local do parto; e do fácil acesso a serviços de saúde de qualidade, desde o atendimento ambulatorial básico ao atendimento hospitalar de alto risco. (VIELLAS *et.al.*, 2014).

2.5 A EVOLUÇÃO DA TECNOLOGIA

A evolução tecnológica computacional normalmente é apresentada em três ou quatro plataformas distribuídas cronologicamente, conforme a Figura 12 apresenta. Não são fases rígidas, mas que de alguma forma um acontecimento tecnológico ou uma nova técnica influenciou esse período mudando o comportamento humano em suas atividades no trabalho ou na vida pessoal, transformando suas vidas e das organizações públicas e privadas para sempre.

Figura 12 – A evolução tecnológica computacional.



Fonte: Elaborado pelo autor.

2.5.1. Primeira Plataforma.

Identificada no final dos anos 50 e na década de 60 foi marcada pelo uso dos Mainframes (grandes computadores que ocupavam grandes salas), mas de uso restrito das grandes corporações e de empresas públicas devido seus altos custos de operação. Estes

equipamentos se destacavam pela possibilidade de processamento de grandes volumes de dados e foi nessa época que foi desenvolvido o “Processamento em paralelo”, quando um computador executa várias tarefas computacionais ao mesmo tempo, uma inovação para a época e tão comum hoje em dia nos *Gadgets* (relógios, leitores de músicas MP3, smartphones e etc.).

2.5.2. Segunda Plataforma.

Esta fase é marcada pela democratização da tecnologia com a chegada dos PCs (*Personal Computers*) no início dos anos 80, possibilitando o uso pessoal dos computadores nas escolas, universidades, residências e pequenas empresas. Esses computadores (PCs) ainda tinham pouca capacidade de armazenamento e processamento, bem inferior aos *Smartphones* de hoje.

Ainda no final dos anos 80 chega ao Brasil a primeira TV por assinatura, transmitida pela empresa TV SHOW BRASIL em Fortaleza-CE via sinal MMDS (*Multichannel Multipoint Distribution Service*), a segunda empresa a transmitir TV por assinatura foi a TV FILME em Brasília-DF.

A década de 90 foi marcada também por grandes revoluções tecnológicas como a chegada da internet, ainda com baixas velocidades, até cem vezes menores que as utilizadas hoje pelos *smartphones* 4G. A telefonia móvel chega aos estados brasileiros: Rio de Janeiro (fev/1990), Paraíba, São Paulo (1992), Ceará (nov/1993), Alagoas e Pernambuco (mar/1994), Maranhão (jun/1994), Pará e Amazonas (jul/1994), e até o final de 1994 chegava em outros estados do Brasil, como: Mato Grosso, Mato Grosso do Sul, Roraima, Amapá e Rondônia, todos usando a 1ª geração da tecnologia: AMPS (*Advanced Mobile Phone Service*).

Um grande marco desta fase foi a privatização das telecomunicações no Brasil nos anos de 1997 e 1998, com a entrada do capital estrangeiro financiando uma revolução na telefonia móvel, telefonia fixa, comunicação de dados e internet, acabando com monopólios, criando ofertas e disputas pelo mercado, antes monopólio estatal através de empresas de economia mista (público-privado).

2.5.3. Terceira Plataforma.

Todo esse caminho relatado foi necessário para que a partir de 2010 pudessemos assistir à popularização das tecnologias com a chegada da computação em nuvem, revolução

dos smartphones e Apps (Aplicativos móveis), conectividade e mobilidade, e assim criar um ambiente tecnologicamente adequado para difundir as redes sociais, com capacidade para manipular textos, fotos, vídeos, TVs, links, tudo em alta velocidade e disponibilidade.

Novas tecnologias começam a ganhar força com a disponibilidade e baixo custo dos equipamentos (servidores, computadores, notebooks, componentes de redes, *smartphones* e etc.) e da computação em nuvem, surge assim o ambiente propício para a Mineração de dados, Bigdata (busca, agrupamento e análise de dados), Algoritmos de IA (Inteligência Artificial), Aprendizado de máquina (*Machine Learning*) e Aprendizagem profunda (*Deep Learning*), tecnologias que já estão sendo usadas ou sendo implantadas em várias organizações no Brasil e no mundo, mas que ainda é incipiente nas organizações públicas, especialmente na Saúde Pública do Brasil.

Outras tecnologias são importantes citar, como: Iot (Internet das Coisas), Realidade Virtual (3D), Realidade aumentada, Ensino a Distância (EAD), os livros digitais (E-books), impressão 3D, BI (*Business Intelligence*), E-commerce e Redes sociais (Facebook, WhatsApp, Instagram, Twitter) e etc.

Temos ainda a revolução nos softwares e linguagens de programação que agora se aproveitam de todo o potencial dos novos hardwares disponíveis, podemos citar: Java, Python, C#, PHP, Javascript, Ruby, R e etc.

2.5.4. Quarta Plataforma.

Provavelmente estamos vivendo essa transição da terceira para a quarta plataforma, com algumas tecnologias novas surgindo e iniciando o processo de popularização e redução dos custos, como: a nanotecnologia, a próxima geração de baterias, Blockchain (repositório público de transações), Computação Quântica, carros autônomos, revolução da robótica, moedas digitais, entre outras tecnologias que a cada dia estarão mais presentes no nosso cotidiano.

2.6 INTELIGÊNCIA ARTIFICIAL (IA)

A Inteligência Artificial tem como objetivo principal entender e aplicar técnicas inteligentes na solução de problemas complexos para as mais variadas aplicações que usam computadores ou máquinas inteligentes. Seguem algumas definições de inteligência artificial usadas na literatura.

“IA é a parte da Ciência da Computação que se preocupa em desenvolver sistemas computacionais inteligentes, isto é, sistemas que exibem características, as quais nós associamos com a inteligência no comportamento humano - por exemplo: Compreensão da linguagem, aprendizado, raciocínio, resolução de problemas etc.” (BARR; FEIGENBAUM, 1981).

“IA é o estudo de faculdades mentais através do uso de modelos computacionais” (CHARNIAK; McDERMOTT, 1987).

“IA é a arte de criar máquinas que executam funções que requerem inteligência quando executadas por pessoas” (KURZWEIL, 1990).

“IA é o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor” (RICH; KNIGHT, 1990).

Combinando todos os conceitos acima, chegamos ao conceito mais usado na academia: “IA é o campo da Ciência da Computação que está preocupado com a automação do comportamento inteligente”.

Embora um computador não possa ter experiências, estudar e aprender como um ser humano consegue, ele pode utilizar os conhecimentos dos especialistas humanos. Esses conhecimentos podem ser fatos, conceitos, teorias, métodos heurísticos, procedimentos e relacionamentos. O conhecimento também é informação organizada e analisada para torná-la compreensível e aplicável à solução de problemas ou à tomada de decisão (TURBAN *et. al.*, 2010).

Várias aplicações de IA suscitam também, questões filosóficas profundas. Em que medida podemos dizer que um computador pode entender expressões em linguagem natural ? Para produzir ou compreender uma linguagem, é necessária a interpretação de símbolos. Não é suficiente ser capaz de dizer que uma cadeia de símbolos está bem formada. Um mecanismo para a compreensão deve ser capaz de atribuir um significado ou interpretar símbolos dentro de um contexto. O que é significado ? O que é interpretação ? Em que medida a interpretação requer responsabilidade ? (LUGER, 2013).

Segundo Luger (2013), são várias as áreas em que a IA pode auxiliar, entre elas podemos citar:

- O uso do computador para executar raciocínio, reconhecimento de padrões, aprendizado ou outras formas de inferência;
- Raciocínio que utiliza as características qualitativas significativas de uma situação;
- Respostas que não são nem exatas nem ótimas, mas que são “suficientes”. São casos que resultados exatos ou ótimos são caros demais ou mesmo impossíveis, entre outras áreas.

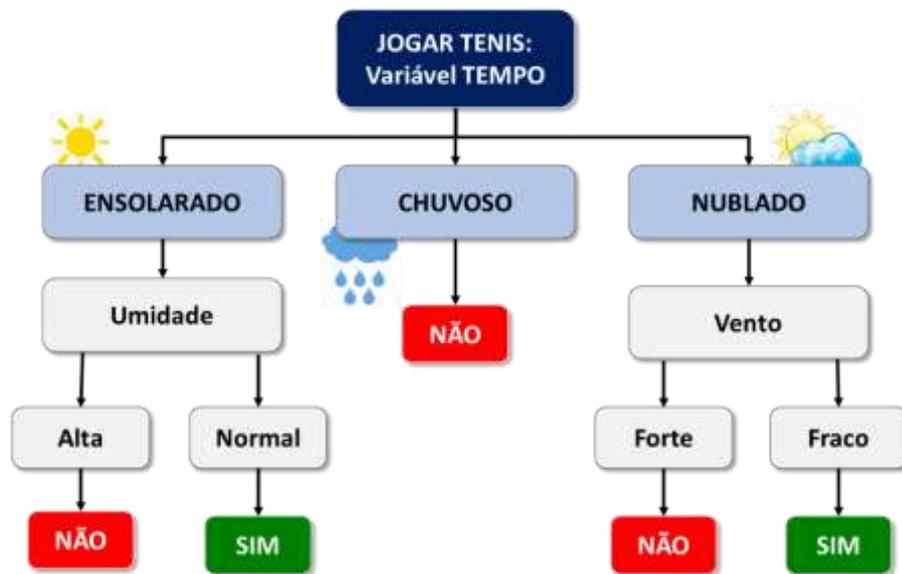
2.6.1. Árvore de decisão

Essa é uma estrutura baseada em uma situação descritiva por um conjunto de atributos e retorna uma decisão, que é o valor resultado para o valor da entrada. A decisão é uma das estruturas básicas de programação, representada pelos comandos: *if...then...else* (SE... ENTÃO..... SENÃO.....).

Segundo Russel e Norvig (2013), a indução de árvores de decisão é uma das formas mais simples, e ainda assim mais bem-sucedidas, de aprendizagem de máquina. Primeiro, descrevemos a representação – o espaço da hipótese – e, em seguida, como aprender uma boa hipótese. Uma árvore de decisão representa uma função que toma como entrada um vetor de valores de atributos e retorna uma “decisão” – um valor de saída único.

A Figura 13 abaixo, representa um exemplo de uso de uma árvore de decisão que responde à pergunta: Está propício para a prática do Tênis ? A partir da análise das variáveis: Condições do tempo, a umidade e a velocidade do vento.

Figura 13 – Exemplo de árvore de decisão.



Fonte – Elaborado pelo autor.

Para construir uma árvore de decisão é necessário desenhar um gráfico de modo que as informações adicionais obtidas pelo tomador de decisão, no transcorrer do processo, possam ser incluídas como novas variáveis (BOUZADA *et. al.*, 2013).

A árvore de decisão oferece uma visão gráfica da tomada de decisão necessária, conforme a Figura 13, acima. Algumas especificam que variáveis são testadas, que ações

devem ser executadas e a ordem que a tomada de decisão deverá ser executada (MARTIN; MCLURE, 1991).

2.6.2. Lógica Fuzzy.

A Teoria dos Conjuntos Difusos (*Fuzzy Sets*) foi proposta em 1965 pelo matemático Lotfi Zadeh (1921-2017), como uma alternativa na representação de informações imprecisas ou vagas. Esta área da IA contabiliza milhares de publicações entre artigos, livros e teses relacionadas a desenvolvimentos teóricos realizados neste campo, e às diversas aplicações desenvolvidas em áreas como: tomada de decisões, agrupamento de dados, classificação de padrões, sistemas de diagnósticos, processamento de sinais (sons, imagens, etc.), sistemas de controle, e muitas outras (LIMA; PINHEIRO; SANTOS, 2014).

Na teoria dos conjuntos formais há duas suposições essenciais: [1] A pertinência a conjuntos: para qualquer elemento e um conjunto pertencente a um universo, o elemento é o membro do conjunto ou é um membro do complemento do conjunto; [2] A Lei do Meio excluído: um elemento não pode pertencer a um conjunto e, também, ao seu complemento. Essas suposições são violadas na teoria dos conjuntos nebulosos (LUGER, 2013).

A Lógica Clássica determina um valor de corte, enquanto que a Lógica Fuzzy permite uma área nebulosa de transição entre uma classificação e a seguinte, por exemplo: a legislação brasileira usando a lógica clássica determina a maioridade penal em 18 (dezoito) anos, isto é, um crime realizado um dia antes do autor completar dezoito anos, não lhe será imputada culpa penal, mas pela lógica Fuzzy, um autor de crime com 17 (dezessete) anos poderá ser considerado 80% maior de idade de acordo com suas características comportamentais e 20% menor de idade, permitindo assim a aplicação de uma penal proporcional a sua maioridade. A Figura 14 abaixo, representa esse exemplo do modelo clássico e o modelo Fuzzy.

Figura 14 – Representação da Lógica Clássica e Lógica Fuzzy.



Fonte: Elaborada pelo autor.

Aplicando-se a lógica Fuzzy no exemplo acima, poderia também considerar que um

jovem de 19 (dezenove) anos que praticou um crime, ainda não possuía todas as características da maioridade penal, neste caso, a pena seria reduzida para 85% por exemplo, pois ele não tinha atingido a plenitude da vida adulta.

Para Luger (2013), a teoria de Zadeh expressa a falta de precisão de uma maneira quantitativa, introduzindo uma função de pertinência a conjuntos, que pode assumir valores entre 0 e 1. Essa noção de um conjunto nebuloso pode ser descrita como: seja “S” um conjunto e “s” um membro desse conjunto. Um subconjunto nebuloso “F” de “S” é definido por uma função de pertinência $mF(s)$ que mede o “grau” com que “s” pertence a “F”.

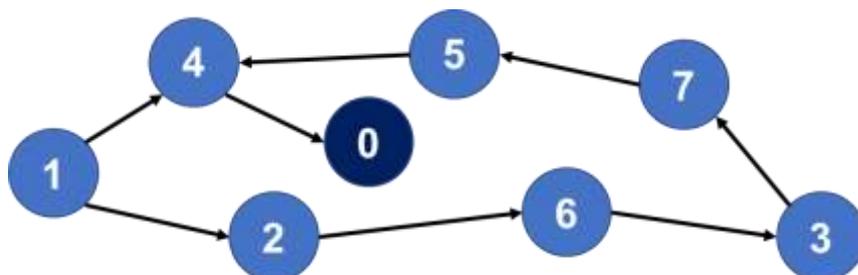
2.6.3. Grafos.

A teoria dos grafos é a melhor ferramenta para representar problemas sobre objetos e suas relações. E foi assim mesmo que surgiu essa técnica, quando no início do século XVIII, em 1736, o matemático suíço Leonhard Euler inventou a teoria dos grafos para resolver o “problema das pontes de Königsberg”. A cidade de Königsberg na Prússia, hoje Kaliningrado, ocupava ambas as margens e duas ilhas do rio Pregel. As ilhas e as margens eram conectadas por sete pontes. O problema questiona se existe um roteiro para que percorra todas as pontes, passando uma única vez em cada ponte (LUGER, 2013).

Entre os vários tipos de problemas que a computação tenta propor soluções, existem os problemas de conexões entre elementos. A teoria dos grafos tenta representá-los e auxiliar em suas soluções, entre eles podemos citar: as estradas que conectam cidades, links de um website, a distribuição entre disciplinas e professores de uma escola, entre outros problemas complexos (ASCENCIO; ARAÚJO, 2010).

Um grafo G é formado pelo par de conjuntos V e E , sendo V o conjunto de vértices de G , e E o conjunto de arestas de G . A representação geométrica dos grafos é feita marcando pontos distintos no plano para representar os vértices, e uma linha ligando dois pontos para representar a aresta, como mostra a Figura 15, abaixo.

Figura 15 – Exemplo de grafo, que liga vértices por meio de arestas.



Fonte – Adaptação do autor

Um grafo pode ser visualizado através de uma representação geométrica, na qual seus vértices correspondem a pontos distintos do plano em posições arbitrária, enquanto que a cada aresta (v, w) , é associada uma linha arbitrária unindo os pontos correspondentes a (v, w) (SZWARCFITER, 2018).

2.6.4. Teorema de Bayes.

O famoso Teorema de Thomas Bayes foi publicado em 1763, poucos anos após a sua morte. Thomas Bayes (1702-1761), matemático britânico e pastor presbiteriano, escreveu o artigo intitulado “Ensaio para a solução de um problema na doutrina das chances”, publicado na *Philosophical Transactions of the Royal Society of London*. Seu teorema relaciona causa e efeito de modo que, compreendendo o efeito, podemos descobrir a probabilidade de suas causas. Como resultado, esse teorema é importante tanto para determinar as causas de doenças, como o câncer, quanto é útil para determinar os efeitos de uma medição em particular sobre essa doença (LUGER, 2013).

A equação descrita abaixo é conhecida como “Regra de Bayes” ou “Lei de Bayes” ou “Teorema de Bayes”, é uma equação simples usada como base de vários sistemas modernos de IA para inferência probabilística.

$$P(b|a) = \frac{P(a|b) P(b)}{P(a)} \quad \text{ou} \quad P(\text{causa}|\text{efeito}) = \frac{P(\text{efeito}|\text{Causa}) P(\text{causa})}{P(\text{efeito})}$$

A probabilidade condicional $P(\text{efeito}|\text{causa})$ quantifica a direção na relação “causal”, enquanto $P(\text{causa}|\text{efeito})$ descreve a direção do “diagnóstico”. Em uma tarefa como o diagnóstico médico com frequência temos probabilidades condicionais sobre relacionamentos causais, isto é, o médico conhece a $P(\text{sintomas}/\text{doenças})$ e quer derivar um diagnóstico $P(\text{doenças}/\text{sintomas})$ (NORVIG; RUSSELL, 2013).

Norvig e Russell (2013) apresentam o seguinte exemplo, um médico sabe que a meningite faz o paciente ter uma rigidez no pescoço, digamos, durante 70% do tempo. O médico também conhece fatos incondicionais: a probabilidade *a priori* de um paciente ter meningite é 1/50.000, e a probabilidade *a priori* de qualquer paciente ter rigidez no pescoço é 1%. Sendo “s” a proposição de que o paciente tem rigidez no pescoço e “m” a proposição de que o paciente tem meningite, temos então:

$$P(s|m) = 0,7$$

$$P(m) = 1/50.000$$

$$P(s) = 0,01$$

$$P(m|s) = \frac{P(s|m) P(m)}{P(s)} = \frac{0,7 \times 1/50.000}{0,01} = 0,0014$$

Ou seja, esperamos que apenas um em 5.000 pacientes com rigidez de pescoço tenha meningite, pois embora a rigidez de pescoço seja uma indicação forte de meningite, a probabilidade de o paciente estar com meningite permanece pequena.

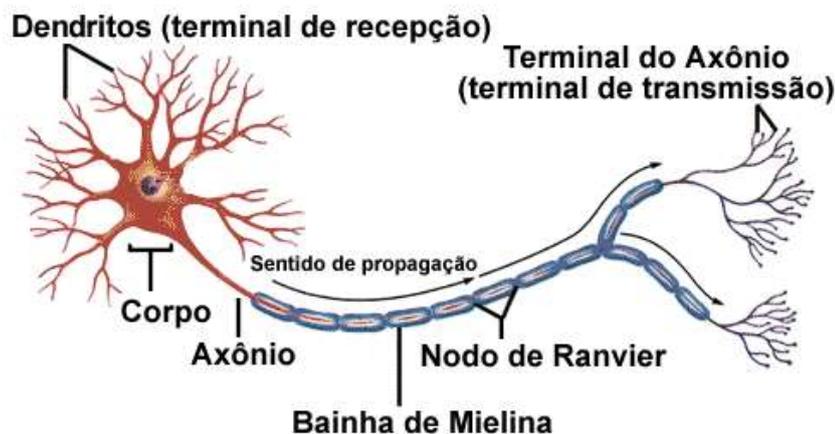
2.6.5. Redes Neurais.

O trabalho sobre redes neurais foi motivado pelo reconhecimento de que o cérebro humano processa as informações de uma maneira diferente do que ocorre com o computador clássico (MITCHEL, 1997).

O neurônio biológico possui terminais de recepção chamados “Dendritos” que recebem estímulos de outros neurônios, esse fenômeno é chamado de “Sinapses”, que pode ser elétrica ou química. Uma sinapse padrão conecta o axônio do neurônio que envia o estímulo, chamado neurônio pré-sináptico ao dendrito do neurônio que recebe o estímulo, chamado de neurônio pós-sináptico, formando assim uma rede de neurônios. Os potenciais pós-sinápticos podem ser: excitatórios ou inibitórios (HAYKIN, 2001).

A Figura 16 representa graficamente um neurônio biológico.

Figura 16 – Modelo de Neurônio Biológico.



Fonte – HAYKIN, 2001.

O primeiro modelo artificial de um neurônio biológico foi resultado de um trabalho pioneiro de um psicólogo e neurofisiologista, Warren McCulloch e do matemático Walter Pitts, em 1943. O trabalho de McCulloch e Pitts não teve, a princípio, o propósito de

apresentar técnicas computacionais de aprendizado e sim de mostrar um modelo artificial de um neurônio e suas capacidades computacionais (BRAGA, 2007).

Donald Hebb, em 1949 desenvolveu o primeiro trabalho que efetivamente demonstra o aprendizado de máquina. No seu trabalho ele demonstra como é possível através da variação de pesos de entrada dos neurônios obter-se a aprendizagem da rede neural. Ele era um biólogo que estudava o comportamento dos animais, propôs um princípio que mostrava que o aprendizado em sistemas nervosos complexos poderia ser reduzido a um processo local, onde as intensidades das conexões sinápticas seriam alteradas apenas considerando os erros detectados localmente (KOVÁCS, 2006).

A regra de Hebb, um formalismo matemático ainda é usado hoje em diversos algoritmos de aprendizado. Trata-se de uma teoria baseada no reforço das ligações sinápticas entre neurônios excitados, que explica o aprendizado (BRAGA, 2007).

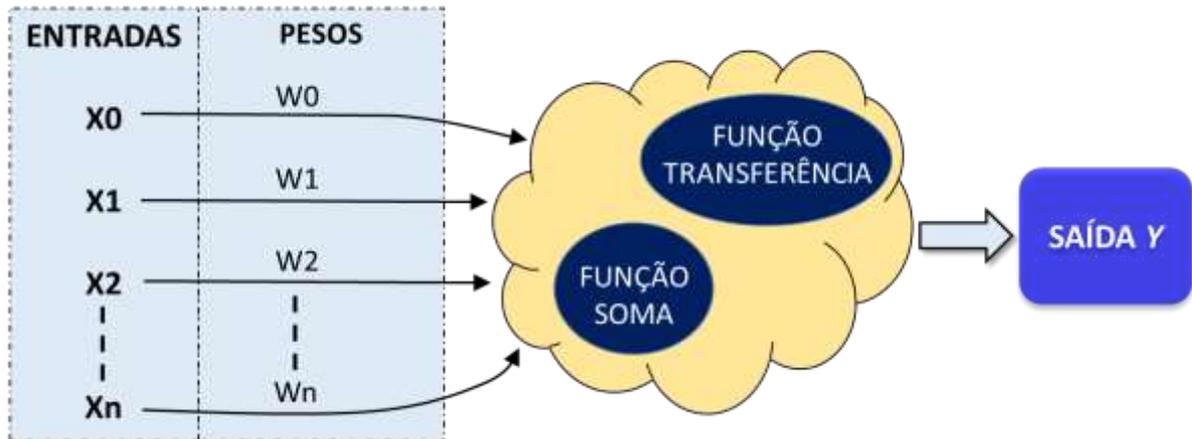
Nosso cérebro funciona como um computador altamente complexo, não-linear e paralelo, tem a capacidade de organizar sua estrutura em torno de 1011 neurônios, de forma a processar por exemplo, o reconhecimento de um rosto familiar inserido em uma multidão em aproximadamente 100-200 ms, ao passo que tarefas muito mais simples levam dias para serem processadas por um computador convencional (HAYKIN, 2001).

De acordo com Haykin (2001), uma rede neural é um processador maciçamente e paralelamente distribuído, constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. A rede neural se assemelha ao cérebro em dois aspectos:

- 1) O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
- 2) Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar conhecimento adquirido.

Um neurônio é uma unidade de processamento de informação fundamental para a operação de uma rede neural (HAYKIN, 2001). O processo utilizado para se obter o conhecimento é chamado de algoritmo de aprendizagem, conforme mostra a Figura 17, abaixo.

Figura 17 – Modelo de neurônio artificial de McCulloch e Pitts.

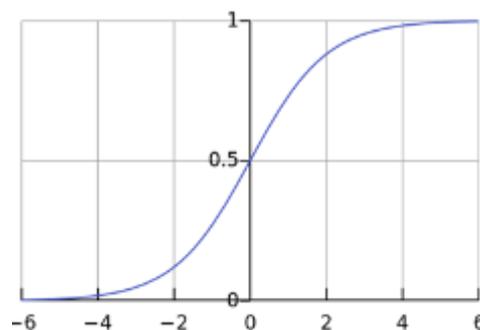


Fonte – Adaptação do autor

De acordo com Braga (2007), o neurônio artificial criado por McCulloch e Pitts é uma simplificação do neurônio biológico como este era conhecido à época. Sua descrição matemática propiciou um modelo com “ n ” nodos de entrada (correspondente aos dendritos) que recebem os valores X_1 , X_2 até X_n (que correspondem as ativações dos neurônios anteriores), e um neurônio de saída “ y ” (que corresponde ao axônio). O comportamento das sinapses dos neurônios biológicos, são representados no modelo artificial pelos pesos W_1 , W_2 até W_n , que se referem aos valores de entrada através do produto destes pelos valores de entrada. Esses pesos podem assumir valores positivos ou negativos correspondendo às sinapses excitatórias ou inibitórias, respectivamente.

Um neurônio dispara quando a soma ponderada dos valores $X_i W_i$ é confrontado com um limiar (*threshold*) que determinará sua excitação ou não, correspondente à soma dos impulsos que dispara a sinapse no modelo biológico. A ativação do neurônio de saída é obtida através de uma função de ativação, normalmente uma função logística, sendo as mais comuns a sigmoide e degrau, conforme a Figura 18 (BRAGA, 2007).

Figura 18 – Função Sigmoide



Fonte – Adaptação do autor

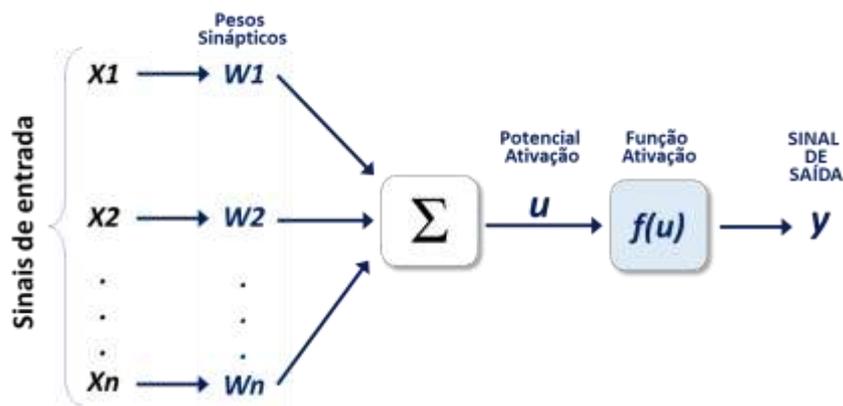
$$f(x) = \frac{L}{1 + e^{-k(x - x_0)}}$$

Onde x_0 é o valor de x no ponto médio da curva sigmoide.

L = o valor máximo da curva sigmoide.

Segundo Castro e Ferrari (2016), a Rede Perceptron de uma camada é um tipo de rede neural artificial que foi inventada em 1957 pelo psicólogo americano Frank Rosenblatt (1928-1971) no Cornell Aeronautical Laboratory, ela é o tipo mais simples de uma rede neural artificial capaz de classificar padrões linearmente separáveis. O Perceptron é uma rede neural com uma única camada de pesos, ou seja, um conjunto de neurônios de entrada e um conjunto de neurônios de saída, com pesos sinápticos e bias ajustáveis, conforme apresenta a Figura 19.

Figura 19 – Rede Perceptron de uma camada.

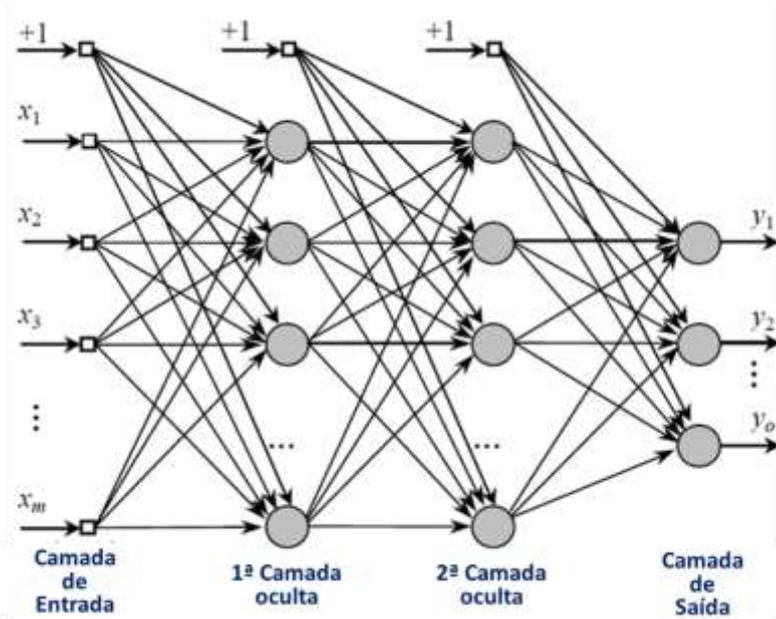


Fonte – Elaborado pelo autor.

A rede MLP (*Multi-Layer Perceptron*) ou *Perceptron* de Múltiplas Camadas é uma rede do tipo *Perceptron* com pelo menos uma camada intermediária. Trata-se de uma generalização do *Perceptron* simples. A proposta da Rede *Perceptron* com múltiplas camadas foi uma tentativa de superar as limitações encontradas no *perceptron* simples. O algoritmo de treinamento dessa rede é uma generalização da regra delta proposta por Widrow & Hoff para o treinamento da Rede Adaline (CASTRO; FERRARI, 2016).

A Figura 20 nos apresenta uma Rede MLP com uma camada de entrada, duas camadas ocultas e uma camada de saída.

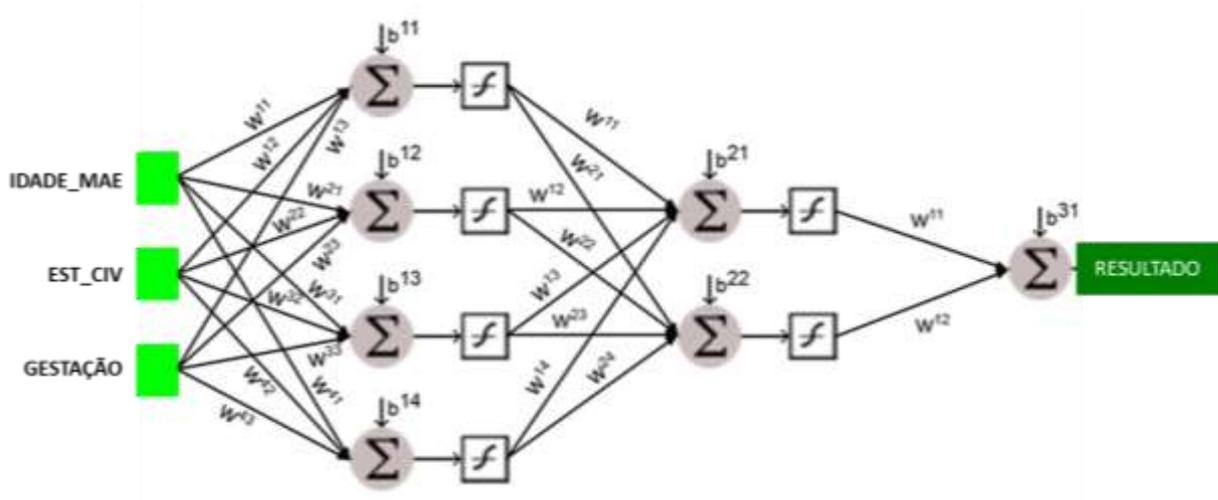
Figura 20 – Rede MLP (Multi-Layer Perceptron) com mais de uma camada.



Fonte – Castro e Ferrari (2016).

A Figura 21 mostra uma Rede Neural Artificial (RNA) de duas camadas para classificar características de uma gestação baseada nas variáveis: idade da mãe, estado civil e semanas de gestação até o parto.

Figura 21 – Rede Neural Artificial de duas camadas.



Fonte – Elaborado pelo autor.

2.6.6. Machine Learning (Aprendizado de máquina)

Aprendizado de máquina é quando um programa aprende a partir da experiência E , em relação a uma classe de tarefas T , com medida de desempenho P , se seu desempenho em T ,

medido por P, melhora com E, isto é, melhorar a realização de uma tarefa a partir da experiência (MITCHELL, 1997).

O aprendizado de máquina pode ser supervisionado ou não supervisionado. Existem algoritmos específicos para cada tipo de aprendizado, como mostra a Figura 22 abaixo, os principais algoritmos de aprendizado supervisionado são: K-NN, Árvore de decisão, Naive Bayes, Perceptron/Adaline e MLP (Multi-Layer Perceptron), já os principais algoritmos de aprendizado não supervisionado são: K-means, Métodos hierárquicos e SOM.

Diz-se que um algoritmo usa aprendizado supervisionado quando ele utiliza características prévias para prever o futuro, o modelo mais conhecido é a regressão linear que minimiza os erros quadráticos entre a variável predita e o valor real da variável. Já o aprendizado não supervisionado busca agrupar objetos ou fatos com características semelhantes ou próximas, a técnica mais conhecida é a clusterização ou agrupamento.

Na nossa pesquisa utilizamos aprendizado supervisionado para gerar modelos preditivos com MLP (Multi-Layer Perceptron).

Figura 22 – Principais técnicas e algoritmos de aprendizado de máquina.



Fonte: Elaborado pelo autor.

Segue abaixo a descrição e finalidade das principais técnicas e algoritmos:

- a) Descoberta de associações – é a busca por eventos que ocorrem simultaneamente fazendo emergir conhecimento dos eventos combinados. Os algoritmos mais conhecidos para detectar associações são: Apriori, GSP e DHP (ZAKI, 2000).

Exemplo de uso: Para identificar grupos de produtos adquiridos juntos por um cliente para usar esse conhecimento para vender mais.

b) Classificação – Essa tarefa consiste em descobrir uma função que mapeie um conjunto de registros em um conjunto de classes. Os algoritmos mais conhecidos para essa tarefa são: Redes Neurais, Algoritmos Genéticos, Rede Bayesiana e Lógica Nebulosa.

Exemplos de uso: quando bancos e financeiras avaliam um pedido de empréstimo; um e-mail que foi direto para a caixa de spam; no reconhecimento de voz e face ou no Classificador de notícias;

c) Regressão – É semelhante a classificação, só que o atributo-alvo assume valores numéricos. As técnicas mais usadas são Redes Neurais e estatística. Os algoritmos mais conhecidos são: K-NN, Adaline e MLP (Multi-Layer Perceptron).

Exemplo de uso: quando um banco aumenta o seu limite de crédito.

d) Agrupamento (Clusterização) – Técnica que segmenta uma base de dados em subconjuntos ou clusters, de forma que os elementos de um cluster compartilhem propriedades comuns que os diferenciem dos elementos dos outros clusters. O objetivo é maximizar a similaridade “Intracluster” e minimizar a similaridade “Intercluster” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Exemplos de uso: A área de Marketing das empresas usa para segmentar os clientes, agrupando clientes com as mesmas características. A Netflix já usou para prever preferências dos usuários com base nos usuários similares;

e) Sumarização – Identifica características comuns entre registros do conjunto de dados. Essa tarefa é aplicada a cada agrupamento obtido na tarefa de clusterização. A Lógica Indutiva e os Algoritmos Genéticos são as tecnologias mais utilizadas para a sumarização (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Exemplo de uso: identificar as características dos clientes que gastam muito e procurar semelhantes no mercado.

f) Detecção de Desvios – Consiste em identificar registros cujas características NÃO atendam aos padrões considerados normais no contexto, são os chamados valores atípicos (*outliers*), aqui são usadas técnicas estatísticas (WEISS; INDURKHYA, 1998).

Exemplos de uso: Bancos usam para detectar fraudes em cartão de crédito; na seleção de atletas para exames antidoping e para detectar invasão em redes de computadores.

g) Descoberta de sequência – É uma extensão da Descoberta de Associações, neste caso pode ser ampliada para considerar a ordem (ou sequência) que os fatos ocorreram (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Exemplos de uso: A Netflix utiliza para recomendar seu próximo filme ou como a Amazon utiliza, informando, quem comprou o livro A também comprou o livro B e o livro C.

O problema fundamental da análise é entender o que realmente deve ser resolvido. É preciso entender a situação, os processos, os dados e as circunstâncias. É preciso categorizar tudo em volta do problema da melhor forma que puder a fim de compreender exatamente qual é a solução ideal. Em Ciência de dados, você encontrará com frequência o “problema visto da forma errada” (FOREMAN, 2016).

2.6.7. *Deep Learning* (Aprendizagem profunda)

A inteligência artificial vem se desenvolvendo há anos, mas está entrando em um estágio crucial em seu desenvolvimento e adoção. A última década viu uma confluência de fatores - em particular, técnicas aprimoradas como o *Deep Learning* (aprendizagem profunda), e o crescimento dos dados disponíveis e do poder de processamento dos computadores, habilita essa tecnologia para ser implantado de forma muito mais abrangente, isso traz uma série de oportunidades, mas também riscos e desafios (UK, 2018).

Deep Learning ou aprendizagem profunda é uma subárea de *Machine Learning* que aprimora e otimiza o uso das redes neurais. Nesta técnica são usadas muitas camadas de neurônios artificiais, o Deep Learning trabalha com dados brutos e pode resolver problemas mais complexos do que as demais técnicas de Inteligência Artificial.

Esta técnica tem sido empregada em visão computacional, reconhecimento de voz e processamento de linguagem. Ao usar várias camadas de processamento de dados não lineares, a saída de uma camada é entrada de uma camada mais profunda e assim por diante. Empresas como a Google e a Microsoft tem usado esta técnica em duas ferramentas muito populares atualmente: Google Translate, tradutor entre várias línguas e o Cortana da Microsoft que funciona como um assistente pessoal.

Conforme Deng e Yu (2014), o conceito de Deep Learning é abrangente, podendo ser definido como uma subárea de Aprendizado de Máquina caracterizado por:

[a] Utilização de várias camadas de informações não-lineares para extração de características (de forma supervisionada ou não-supervisionada), transformação e análise de padrões;

[b] Uso de algoritmos de aprendizado de múltiplos níveis de representação para identificar relações entre os dados através de modelos estatísticos, normalmente fazendo uso de redes neurais.

Deep Learning refere-se a presença de muitas camadas na RNA (Rede Neural Artificial), mas esse significado mudou ao longo do tempo. Em 2012, 10 (dez) camadas já eram suficientes para considerar uma rede neural como profunda. Contudo, em 2017 é mais comum considerar uma rede neural profunda quando esta possui mais de 100 (cem) camadas (DRAELOS *et. al.*, 2017).

2.7 BUSINESS INTELLIGENCE (BI)

A inteligência de negócios (IN) ou *Business Intelligence* (BI), é uma estratégia que combina arquiteturas, ferramentas, banco de dados, aplicações e metodologias (RAISINGHANI, 2004).

A arquitetura do Business Intelligence (BI), traduzido por Inteligência de Negócios ou Inteligência Empresarial, é um conjunto de técnicas e ferramentas de manipulação dos bancos de dados para transformar dados primários em informações e conhecimentos úteis para apoiar a gestão pública ou privada nas tomadas de decisões.

O principal benefício do BI para uma organização é a capacidade de fornecer informações precisas quando necessário, incluindo uma visão em tempo real do desempenho corporativo. Estas informações são necessárias para todos os tipos de decisões, principalmente para o planejamento estratégico (TURBAN *et. al.*, 2010).

O BI é composto por uma série de ferramentas que podem auxiliar os usuários finais, como: relatórios e consultas, mineração de dados ou textos, gestão do desempenho do negócio, painéis de controle (*dashboards*) e etc.

Existem no mercado várias ferramentas de BI que auxiliam na implantação de uma inteligência de negócios e competitiva para as organizações públicas ou privadas.

O BI moderno deixou de ser apenas um sistema de relatórios quando foram adicionadas atividades como: [1] Análise preditiva sofisticada; [2] Alertas por eventos em tempo real e [3] Suporte à tomada de decisão. Usar um BI só para relatar é como dirigir um carro olhando para o retrovisor, a visão é sempre do passado. A análise preditiva permite que a gestão reaja conforme as coisas acontecem e seja proativa com relação ao próprio futuro (TURBAN; VOLONINO, 2013).

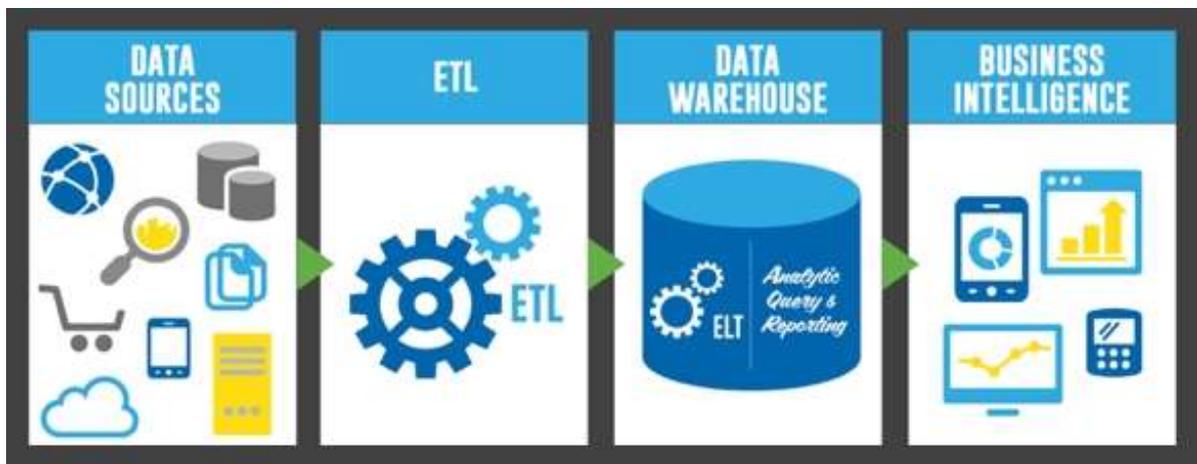
2.7.1. Data Warehouse.

O *Data Warehouse* é uma fonte de dados consultáveis da organização, formado pela união de todos os *data marts* correspondentes (KIMBALL *et. al.*, 2008).

O processo de *Extract, Transform and Load* (ETL) é o processo mais crítico e demorado na construção de um *Data Warehouse* (DW), pois consiste na extração dos dados de bases heterogêneas, na transformação e limpeza desses dados, e na carga dos dados na base do DW. As decisões gerenciais são tomadas com base nas informações geradas pelas ferramentas do tipo *front-end*. Essas informações são geradas através dos dados armazenados no DW. Se esses dados não forem corretamente trabalhados no processo de extração, as informações geradas através deles farão com que decisões sejam tomadas erroneamente, podendo afetar diretamente os negócios da organização (SILVA; ZAIDAN, 2015).

A Figura 23 abaixo, mostra o processo para geração do BI, desde a seleção dos dados primários, extração, tratamento e carga, geração do *Data Warehouse* e conseqüentemente, a geração de dados do BI.

Figura 23 – Processo de geração do BI.



Fonte – site <http://www.keylink.net.au/> acessado em 20/08/2017

2.8 DATA MINING (Mineração de dados)

Data mining (DM), também popularmente referido como descoberta de conhecimento em banco de dados (*KDD – Knowledge Discovery in Databases*), é o processo de descoberta automática de informações úteis em grandes depósitos de dados (PLLANA ET al., 2011).

O *Data Mining* (DM) é uma das fases do KDD (*Knowledge-Discovery in Databases*). É uma técnica que faz a extração de padrões utilizando os dados disponíveis (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

“[...] Mineração de Dados é a expressão geralmente utilizada quando os dados se apresentam de forma estruturada. *Text Mining* (Mineração de Textos), por sua vez, costuma ser usada para designar a análise de dados em formato parcialmente estruturado ou não estruturado (textual). A busca por padrões em mídias, tais como imagem, vídeo, som, gráficos, combinadas ou isoladas, tem sido referenciada como *Multimedia Data Mining* (Mineração de Dados Multimídia)”, (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Dado, Informação e Conhecimento, embora pareçam sinônimos, tem definições bem diferentes quando se trata do contexto da Mineração de Dados. O “dado” é um fato, um valor documentado ou um valor de medição. Quando um sentido semântico ou um significado é atribuído aos “dados”, gera-se “informação”. Quando estes significados se tornam familiares, ou seja, quando um agente os aprende, este se torna consciente e capaz de tomar decisões a partir deles, e surge o “conhecimento” (SILVA; PERES; BOSCARIOLI, 2016).

Segundo Amaral (2016), O processo de mineração de dados é feito a partir de várias etapas, que vão desde entender o negócio ou contexto até a sua implementação. Os dois padrões mais conhecidos são: CRISP-DM (*Cross Industry Standard Process for Data Mining*) e o KDD (*Knowledge-Discovery in Databases*).

Nesse contexto, a mineração de dados é feita em todos os dados existentes e que trafegam na rede, sejam eles educacionais, sociais ou opiniões, entre outros, que sejam estruturados ou não. Coletando diversos formatos de dados modelados como grafos onde no Big Data é efetuada a coleta, análise, e integração dos dados que trafegam na rede para sua posterior apresentação.

A Mineração de Dados provê um método automático para descobrir padrões em dados, sem a tendenciosidade e a limitação de uma análise baseada meramente na intuição humana. Presidentes de grandes corporações como IBM, Microsoft e Harley-Davidson não foram capazes de prever que o mercado ia preferir PC's, internet e motos populares. Além disso, massas de dados são demasiadamente grandes e intrincadas para tratamento manual (BRAGA, 2005).

No contexto da análise dos dados, o Aprendizado de Máquinas (ML – Machine Learning), vem de forma crescente sendo largamente utilizado. O Big Data permite que os algoritmos de Machine Learning revelem padrões mais refinados e façam previsões mais atualizadas e precisas (ZHOU *et. al.*, 2017).

O termo *Data Mining* é frequentemente utilizado por estatísticos, analistas de dados e

técnicos ligados à Administração de Sistemas de Informação. As técnicas de mineração são estruturadas para extrair de grandes massas de dados os padrões estruturais, úteis e recentes implicitamente armazenados nos bancos de dados e que poderiam, de outra forma, permanecer ignoradas (TAN; STEINBACH; KUMAR, 2009).

2.9 O MÉTODO KDD (*Knowledge-Discovery in Databases*).

O termo KDD foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento em bases de dados. Em 1996 foi proposta a definição abaixo por um grupo de pesquisadores:

“KDD é um processo não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Para executar o processo KDD é necessário a existência de um conjunto de dados sobre o assunto pesquisado com vários atributos (campos) em vários registros (linhas). Quanto mais registros, melhor para identificar padrões. Um padrão é um conhecimento representado segundo as normas sintáticas de alguma linguagem formal. Um exemplo de uma linguagem ou formalismo utilizado para representação de padrões é uma linguagem de equações. Em uma linguagem de equações, operadores de adição e multiplicação podem ser utilizados para relacionar variáveis, por exemplo: $(Y = aX + b)$ (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Para usar o processo KDD são necessários estes três componentes: (a) O problema em que será aplicado o KDD; (b) Os recursos disponíveis para a solução do problema e (c) Os resultados obtidos. As etapas operacionais executadas no Processo KDD são: [1] Pré-processamento; [2] Mineração de dados e [3] Pós-processamento, conforme a Figura 24.

Figura 24 – Etapas operacionais do processo KDD.



Fonte: GOLDSCHMIDT; PASSOS; BEZERRA, 2015.

Cada etapa do Processo KDD tem atividades específicas de suas funções. A Figura 25 apresenta as principais atividades de cada etapa.

Figura 25 - Atividades das etapas do processo KDD.

ETAPAS do KDD	ATIVIDADES
[1] PRÉ-PROCESSAMENTO Funções relacionadas com a captação, a organização e o tratamento dos dados.	Seleção de dados
	Limpeza de dados
	Codificação dos dados
	Enriquecimento dos dados
[2] MINERAÇÃO DE DADOS Compreende a busca efetiva por conhecimentos úteis no contexto da aplicação, é a principal etapa do processo. Aqui são definidas as técnicas e algoritmos a serem utilizados.	Descoberta de associações
	Classificação
	Regressão
	Agrupamento (Clusterização)
	Sumarização
	Detecção de desvios
[3] PÓS-PROCESSAMENTO Compreende o tratamento do conhecimento obtido na Mineração de dados para facilitar a interpretação e avaliação dos dados	Elaboração e organização do conhecimento obtido
	Simplificação de gráficos, diagramas ou relatórios

Fonte: GOLDSCHMIDT; PASSOS; BEZERRA, 2015.

O processo de KDD também pode ser representado pela Figura 26, que consiste em uma sequência iterativa, segundo Han e Kamber (2000).

Figura 26 – O Processo KDD.



Fonte – Adaptação do autor

- 1º. Limpeza dos dados: para remover ruídos e dados irrelevantes;
- 2º. Integração dos dados: onde fontes de dados múltiplos podem ser combinadas;
- 3º. Seleção dos dados: onde dados relevantes para a análise são recuperados do banco de dados;

4°. Transformação dos dados: onde os dados são transformados ou consolidados no formato apropriado para mineração. Este é o momento das categorizações;

5°. Mineração de dados: é um processo onde métodos inteligentes são utilizados a fim de extrair padrões de dados;

6°. Avaliação e representação do conhecimento: onde técnicas de visualização e representação de conhecimento são utilizadas para apresentar o conhecimento extraído para o usuário.

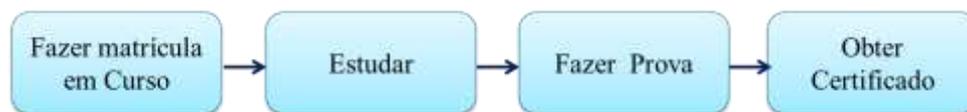
2.10 MAPEAMENTO DE PROCESSOS

Processo é qualquer atividade ou conjunto de atividades que recebe um *input* (entrada), adiciona valor a ele e fornece um *output* (saída) a um cliente específico.

Um processo é uma sequência de atividades com um objetivo específico, isto é, após realizar todos os passos do processo, chega-se a um resultado (CAMPOS, 2013).

Por exemplo, a Figura 27 representa o processo para realização de um determinado curso com o propósito de obter o certificado desse curso.

Figura 27 – Processo para obtenção de certificado de um curso.



Fonte: CAMPOS (2013).

A gestão de processos de negócios, conhecida como BPM – *Business Process Management*, é uma técnica fundamental de gestão que inclui métodos e ferramentas para suportar as fases do ciclo de vida de processos de negócio. Em curto espaço de tempo utilizando a BPM, as empresas públicas ou privadas melhoram a rentabilidade, reduzem desperdícios e custos, no longo prazo, a BPM ajuda a manter as companhias sensíveis às mudanças de negócios (TURBAN *et. al.*, 2010).

Nos últimos anos muito se ouviu falar sobre o mapeamento de processos, com o objetivo de desenhar, executar, documentar, monitorar e controlar a melhoria dos processos esperando alcançar resultados positivos dentro das organizações. A atividade de mapear processos faz com que se esteja a todo momento estudando a melhoria e o funcionamento das atividades, rotinas e processos a fim de atingir os objetivos esperados na organização.

A notação BPMN (*Business Process Model and Notation*) é uma técnica rica em

elementos que possibilita o mapeamento de processos dos mais diversos tipos e de todas as áreas de negócio. Esta notação é hoje a mais utilizada nos ambientes corporativos no mundo todo e incorporada aos principais softwares de gestão de processos (CAMPOS, 2013).

As técnicas de mapeamento e modelagem de processos evoluíram muito nos últimos anos com o apoio de ferramentas tecnológicas, *softwares* de mapeamento, que facilitam e agilizam a tarefa de mapear, modelar e acompanhar a execução dos processos gerenciais e operacionais. No mercado existem vários *softwares* disponíveis pagos ou gratuitos, por exemplo: Bizagi Process Modeler, Bonita Open Solution, Sydle Seed, Dia, Process Maker e Microsoft Visio.

O Apêndice I apresenta o processo proposto para o levantamento de dados, tratamento dos dados, aprendizado de máquina, testes de acurácia e geração do modelo final deste projeto de tese. Esse processo foi mapeado usando o *software* Microsoft Visio Professional 2013.

III - METODOLOGIA

A evolução da Ciência é motivada pela vontade do homem de obter respostas, porém a História nos mostra que respostas erradas já foram aceitas como verdadeiras durante muito tempo. Basta lembrar que por muitos séculos acreditou-se que a Terra era o centro do Universo. Como podemos ter certeza que chegamos à resposta verdadeira? É preciso fazer a investigação científica baseada em três conceitos intimamente relacionados: evidência, verdade e certeza (MASCARENHAS, 2012).

Uma pesquisa é um conjunto de processos sistemáticos, críticos e empíricos aplicados no estudo de um fenômeno. Ao longo da História da Ciência surgiram várias correntes de pensamento, como: empirismo, materialismo dialético, positivismo, fenomenologia, estruturalismo e diversos marcos interpretativos como o construtivismo e a etnografia, dando origem a vários caminhos na busca do conhecimento. No entanto, a partir do século passado essas correntes se “polarizaram” em duas abordagens principais: quantitativo e qualitativo (SAMPLIERI; CALLADO; LUCIO, 2013),

3.1 TIPO E NATUREZA DO ESTUDO

A pesquisa acadêmica visa à busca sistemática de fatos, dados, evidências, argumentos ou indícios que possam contribuir para a solução ou a melhor compreensão de um problema sobre um tema. Dessa forma, gera-se conhecimento a partir da confirmação, da complementação, da refutação ou da soma de condições de contorno à teoria sobre o tema estudado (CAUCHICK et. al., 2017).

A pesquisa quantitativa é um meio para testar teorias objetivas, examinando a relação entre as variáveis. Tais variáveis, por sua vez, podem ser medidas tipicamente por instrumentos, para que os dados numéricos possam ser analisados por procedimentos estatísticos, permitindo aos pesquisadores desta forma de investigação deduzir teorias, explicar alternativas, generalizar e replicar os achados (CRESWELL, 2010).

Esta pesquisa foi desenvolvida utilizando o método quantitativo na geração de modelos preditivos da Mortalidade Infantil no estado do Ceará, restringindo-se ao período de 2013 a 2017. Os modelos preditivos foram gerados a partir da mineração de dados nas bases públicas acessadas para em seguida aplicar algoritmos de Inteligência Artificial. As ferramentas tecnológicas utilizadas foram preferencialmente de uso público e gratuito, como: Python, R, MySql, Java, entre outras, proporcionando baixo custo ao projeto.

Um modelo preditivo pode ser descrito como uma função matemática que aplicada a um grande volume de dados é capaz de identificar padrões e oferecer uma previsão do que pode ocorrer. O modelo preditivo utiliza métodos quantitativos e qualitativos para estabelecer previsões, recomendar otimizações e simular modelos que sejam úteis a uma linha de pesquisa (WALLER; FAWCETT, 2013).

Combinar fatores de risco por meios estatísticos produz um modelo de predição de riscos ou uma ferramenta de predição de riscos, essas ferramentas são cada vez mais comuns nas práticas clínicas (FLETCHER; FLETCHER; FLETCHER, 2014).

Quanto aos fins é uma pesquisa exploratória, pois se deseja desenvolver uma metodologia inovadora com algoritmos específicos para tratamento de dados padronizados e não padronizados, gerando valor estratégico e competitivo para os gestores da saúde pública.

A pesquisa exploratória tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-la mais explícita. Essas pesquisas envolvem análise de exemplos que estimulem a compreensão (GIL, 2007).

Quanto aos objetivos, podemos definir como uma pesquisa aplicada, pois gerou novos conhecimentos estratégicos para auxiliar na tomada de decisões dos gestores públicos. Na pesquisa aplicada, o investigador é movido pela necessidade de contribuir para fins práticos mais ou menos imediatos, buscando soluções para problemas concretos (CERVO; SILVA; BERVIAN, 2007).

Quanto aos meios, foi usada pesquisa bibliográfica para direcionar o trabalho e dar sustentação teórica. A revisão bibliográfica procurou discorrer sobre os temas que cercam a saúde pública, gestão estratégica e inteligência artificial. Organizar uma bibliografia significa buscar aquilo cuja existência ainda se ignora. O bom pesquisador é aquele que é capaz de entrar em uma biblioteca sem ter a mínima ideia sobre um tema e sair dali sabendo um pouco mais sobre ele (ECO, 2005).

A pesquisa bibliográfica procura explicar um problema a partir de referências teóricas publicadas em artigos, livros, dissertações e teses. Pode ser realizada independentemente ou como parte da pesquisa descritiva ou experimental. Em ambos os casos, busca-se conhecer e analisar as contribuições culturais ou científicas do passado sobre determinado assunto, tema ou problema (CERVO; SILVA; BERVIAN, 2007).

Este estudo pesquisou bases de dados do Ministério da Saúde para a geração de modelos preditivos que auxiliem as estratégias públicas de saúde aplicadas à redução da mortalidade infantil, especificamente no estado do Ceará.

3.2 LOCAL E PERÍODO

A pesquisa foi restrita ao estado do Ceará, situado na Região Nordeste da República Federativa do Brasil. Os dados secundários coletados são de 2013 a 2017.

3.3 POPULAÇÃO E AMOSTRA DO ESTUDO

Damos o nome de população ou universo a um conjunto que pode ser formado por pessoas, famílias, empresas, animais ou qualquer outro tipo de elemento. Já a amostra é um pedaço da população que vamos estudar (MASCARENHAS, 2012).

Na estatística, bons resultados são provenientes do uso de bons dados. Não há volume de análise sofisticada que possa compensar dados fundamentalmente falhos. E na estatística, inferências feitas a partir de amostras razoavelmente grandes, adequadamente escolhidas, podem ser tão acuradas quanto tentar extrair a mesma informação da população inteira (WHEELAN, 2016).

O Universo dessa pesquisa são os óbitos infantis ocorridos no Brasil e registrados no Datasus/SIM. São considerados óbitos infantis as mortes de crianças com até um ano de idade.

Como amostra foram utilizados os dados dos óbitos infantis ocorridos no estado do Ceará, no período de 2013 a 2017. A amostra foi extraída da base de dados pública do Ministério da Saúde do Brasil no site do Datasus: Sistema de Informações de Mortalidade (SIM) e contém informações de 8.159 óbitos infantis ocorridos neste período com média de 1.631 óbitos por ano.

Este estudo pesquisou bases de dados do Ministério da Saúde para a geração de modelos preditivos para auxiliar estratégias públicas de saúde aplicadas na redução da mortalidade infantil, especificamente no estado do Ceará.

Para este estudo foi capturada uma base de dados do Datasus com informações secundárias de todos os óbitos infantis ocorridos no Brasil no período de 2013 a 2017, desta base foram selecionados para o estudo os 8.159 óbitos infantis ocorridos nos 184 municípios do estado do Ceará neste período. A RMF (Região Metropolitana de Fortaleza) representa

mais de 45% dos óbitos infantis do estado do Ceará com 3.700 óbitos e somente Fortaleza, a capital, representa mais de 85% da RMF com 3.158 óbitos no período.

A Região Metropolitana de Fortaleza (RMF) é formada por 11 (onze) municípios com a população total estimada pelo IBGE em 2017 de 3,7 milhões de habitantes. O Quadro 2 abaixo apresenta a relação de localidades que formam a RMF e suas respectivas populações em ordem decrescente.

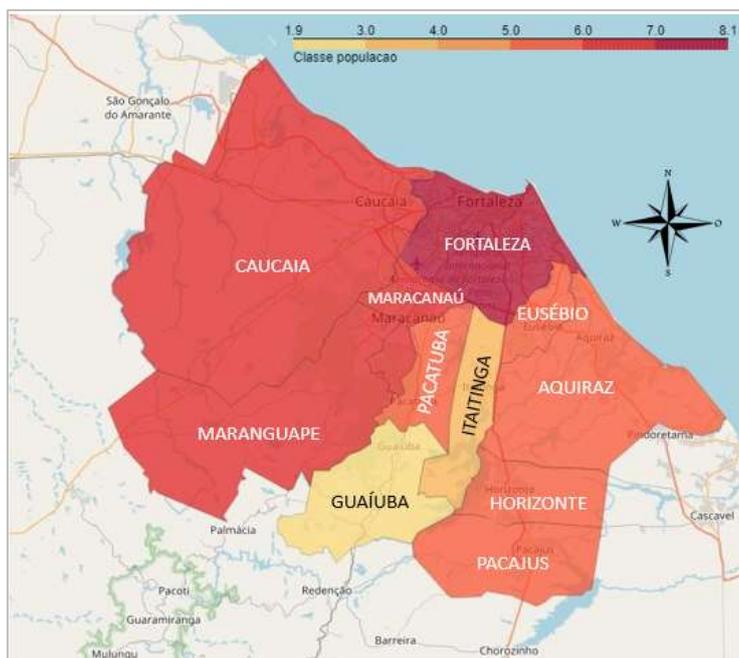
Quadro 2 – Relação de Municípios que formam a RMF.

UF	CD_MUN	NM_MUN	POPULAÇÃO
CE	230.440	Fortaleza	2.627.482
CE	230.370	Caucaia	362.223
CE	230.765	Maracanaú	224.804
CE	230.770	Maranguape	126.486
CE	230.970	Pacatuba	82.824
CE	230.100	Aquiraz	79.128
CE	230.960	Pacajus	70.911
CE	230.523	Horizonte	65.928
CE	230.428	Eusébio	52.667
CE	230.625	Itaitinga	39.310
CE	230.495	Guaiúba	26.331

Fonte: Elaborada pelo autor. (*) População estimada em 2017 (IBGE).

A Figura 28 abaixo, nos mostra o mapa da Região Metropolitana de Fortaleza (RMF) com a classificação por faixa populacional identificada por cores.

Figura 28 – Mapa da Região Metropolitana de Fortaleza por população bruta.



Fonte: Elaborada pelo autor com dados do IBGE (2017).

3.3.1 Critérios de elegibilidade: Inclusão e Exclusão

O estado do Ceará foi escolhido como base do estudo pela facilidade da localização geográfica, posto que o pesquisador nasceu, reside e conhece as características sócio demográficas da região, entretanto esse critério não invalida que estudos futuros sejam desenvolvidos em outras regiões do Brasil.

Os principais critérios de Inclusão são: [a] Óbitos infantis registrados no banco de dados do Datasus/SIM; [b] Óbitos infantis no período de 2013 a 2017; [c] Óbitos infantis com as informações das variáveis categóricas devidamente preenchidas; [d] Óbitos infantis ocorridos no estado do Ceará; [e] Óbitos de crianças com até 01 (um) ano de vida.

Os principais critérios de exclusão são: [a] Óbitos infantis anterior a 2013, pois as características sócio demográficas já não são as mesmas de hoje; [b] Óbitos infantis após 2017, pois as informações ainda não foram homologadas pelo Ministério da Saúde; [c] Registro de óbitos com dados faltantes ou inválidos; [d] Óbitos infantis em localidades fora do estado do Ceará.

3.4 COLETA DE DADOS E EVIDÊNCIAS

A coleta de dados é uma forma fácil e barata de extrair um conjunto de dados de uma fonte *on-line*. Esse método depende da provisão de dados por outras fontes institucionais grandes e pequenas. Muitas organizações governamentais e internacionais oferecem dados para livre acesso e uso pelo público (OLSEN, 2015).

As tabelas que compõem o banco de dados foram coletadas no site de acesso público do Datasus nos sistemas SIM e SINASC do Ministério da Saúde do Brasil. As informações contidas nestas tabelas são oriundas da digitação dos documentos Declaração de nascido vivo e Declaração de óbito, conforme o Anexo I e Anexo II, respectivamente.

3.4.1. Fonte de dados

Os dados foram obtidos nas bases de dados do Ministério da Saúde a partir do banco de dados SIM (Sistema de Informação de Mortalidade) e SINASC (Sistema de Informações Sobre Nascidos Vivos). A partir destas bases foram realizadas consultas usando SQL (*Structure Query Language*) para *linkar* as duas bases e extrair informações estratégicas para o objetivo do estudo.

Para a utilização das bases a pesquisa foi submetida ao Comitê de Ética e Pesquisa (CEP) da Universidade Federal do Ceará (UFC) que liberou o acesso e utilização das informações sem restrições, por tratar-se de dados públicos disponibilizados pelo Ministério da Saúde através do Sistema Datasus/SIM/SINASC.

A partir das bases de dados foi desenvolvido o MER (Modelo de Entidades e Relacionamentos) entre as tabelas das bases de dados capturadas para executar os trabalhos de preparação, limpeza, categorização dos dados e processamento destes dados.

3.4.2. Procedimentos e instrumentos de coleta de dados

Os dados foram coletados no site do Ministério da Saúde: www2.datasus.gov.br/DATASUS que permite a seleção da opção “Arquivo de dados” ou “Download do Tabwin/Tabnet”. Foi utilizada a primeira opção. Em seguida foi usada a caixa de escolha de opções como: Tipo de consulta, Tipos de arquivos, Ano de referência, seleção de uma UF, várias ou todas. Para finalizar a consulta foi executada e ao final disponibilizado um arquivo no formato “dbc” que foi baixado no computador.

O arquivo “dbc” baixado foi convertido para um formato de banco de dados conhecido. Nesse trabalho o arquivo “dbc” foi convertido para o formato “dbf” e posteriormente foi carregado no banco de dados de uso público Mysql Server 5.7, ficando disponível para os processamentos necessários.

3.4.3. Controle de qualidade de dados

Após a carga dos dados brutos no banco de dados Mysql 5.7. A tabela principal com as informações das variáveis (campos) foi validada quanto ao preenchimento dos campos, sintaxe e valoração. Por exemplo: Uma data tem que ter o dia (entre 01 e 31), o mês (entre 01 e 12) e o ano (entre 2013 e 2017) ou a variável RAÇACOR (entre 1 e 5) que representa: 1 - BRANCA; 2 - PRETA; 3 - AMARELA; 4 - PARDA; 5 – INDIGENA. O Apêndice II detalha o Dicionário de dados da Tabela de Óbitos Infantis, explicando cada variável da tabela.

Banco de Dados é um conjunto de arquivos integrados que atendem a um conjunto de sistemas, formado por tabelas e seus atributos, também chamados de campos (HEUSER, 2009).

O dicionário de dados é a ferramenta utilizada para detalhar e explicar cada variável (campo ou atributo) de uma tabela (objeto) de banco de dados, definindo as características e valores possíveis.

Em seguida foram identificados e criados os campos chaves que funcionaram como índices de busca e criação do MER (Modelo de Entidades e Relacionamentos) com tabelas auxiliares para a formação de um banco de dados relacional. A Tabela de Municípios foi relacionada com alguns campos (atributos) da tabela de Mortalidade Infantil.

O uso de um banco de dados relacional evita os erros de restrições de integridades dos dados, por exemplo: Um óbito em uma localidade que não está cadastrada na tabela de localidades ou um código de ocupação da mãe que não conste na tabela de ocupações profissionais do IBGE.

Foram descartados os registros de óbitos com informações incompletas ou preenchidas de forma errada. Essa avaliação foi realizada no momento da carga e durante o processo de preparação e limpeza dos dados.

3.5 VARIÁVEIS DE ESTUDO

3.5.1. Variáveis dependentes

A variável dependente do estudo é a variável óbito confirmado durante o período do nascimento até 12 meses completos de vida da criança, isto é, o desfecho da Mortalidade Infantil, objeto deste estudo.

3.5.2. Variáveis independentes

As variáveis independentes, referentes ao desfecho (Mortalidade infantil) foram hierarquizadas em três níveis de determinação: distal, intermediário e proximal. O posicionamento das variáveis seguiu uma ordem previamente estabelecida baseada em um modelo que descreve as relações lógicas ou teóricas entre as variáveis na determinação do desfecho (RIBEIRO *et. al.*, 2009).

O Quadro 3 relaciona as variáveis com os níveis de determinação.

Quadro 3 – Variáveis independentes.

VARIÁVEIS INDEPENDENTES		
NÍVEL DISTAL (Fatores socioeconômicos)	NÍVEL INTERMEDIÁRIO (Fatores de Atenção à saúde)	NÍVEL PROXIMAL (Fatores Biológicos)
Raça / Cor	Idade Materna	Peso ao nascer
Anos de estudo da Mãe	Número de gestações	Escore Apgar 1º minuto
Ocupação Materna	Número de abortos	Escore Apgar 5º minuto
Nível econômico	Número de filhos natimortos	Sexo
	Número de consultas pré-natal	Duração da gestação
	Tipo de parto	

Fonte: Elaborada pelo autor.

3.6 TRATAMENTO E ANÁLISE DAS EVIDÊNCIAS (ANÁLISE DOS DADOS)

Os dados brutos baixados do site do Ministério da Saúde receberam tratamento de normalização, isto é, deixar os dados padronizados e coerentes com o que eles representam. Os atributos selecionados visam a melhor identificação de informações com peso substancialmente relevante para o evento “óbito infantil”.

Os dados originais baixados já possuem um nível de categorização bem estruturada que não necessitam tratamentos, isto é, os dados já estão preparados para o uso. Não serão utilizadas informações que identifiquem a criança, a mãe, a rua da residência ou a rua do óbito, sempre preservando o anonimato dos atores e seus familiares.

Os dados foram normalizados utilizando-se as técnicas dos bancos de dados relacionais através da técnica conhecida como MER (Modelo de Entidade e Relacionamentos) que evita os erros de integridade referencial dos dados.

O Apêndice I apresenta o processo mapeado de tratamento dos dados desde a extração dos dados até a aplicação dos modelos preditivos e avaliação da acurácia. Os principais passos são: [a] extração dos dados; [b] normalização dos dados; [c] dividir os dados em treinamento, validação e testes; [d] aplicar os algoritmos de aprendizado; [e] avaliar acurácia; [f] testar modelo gerado; [g] salvar modelo de aprendizado e [h] aplicar o modelo em dados reais.

Neste enfoque, se o processo for rigorosamente seguido e algumas regras lógicas forem seguidas, os dados coletados e tratados terão os padrões de validade e confiabilidade e suas conclusões irão contribuir para a geração de conhecimentos (SAMPIERI; CALLADO; LUCIO, 2013).

3.7 LIMITAÇÃO DO MÉTODO

Nesta pesquisa foram tratadas informações relativas ao período de 2013 a 2017, considerando que o contexto atual, onde se pretende inferir, é semelhante ao período estudado.

Os dados coletados são supostamente de todos óbitos ocorridos no período 2013 a 2017, mas sabe-se que alguns óbitos não são registrados pelas entidades de saúde ou cartórios quando crianças são sepultadas em terrenos não autorizados. Outro problema a ser relatado é o preenchimento errado ou incompleto (*missing values*) de boletins de óbitos que já podemos perceber em registros faltando informações.

Na pesquisa de Almeida e Szarcwald (2017), verificou-se grande precariedade das informações em vários municípios brasileiros quanto ao número de óbitos infantis informados de 2011 a 2013. O estudo inferiu que no estado do Ceará apenas 80,1% dos óbitos infantis foram registrados no SIM, isto é, não conhecemos 1/5 (um quinto) dos óbitos infantis daquele período.

Outro fato importante de registrar são os óbitos registrados como “causas mal definidas”, óbitos em que os sintomas e sinais não foram esclarecidos. Em 2001 o Ministério da Saúde identificou mais de 14% dos óbitos como causas mal definidas, variando de 6,2% na região Sul para a 27% na região Nordeste (MS, 2004).

Em uma pesquisa quantitativa o que se pretende é generalizar os resultados encontrados em um grupo ou segmento (amostra) para uma coletividade maior (universo ou população), e se espera que os estudos realizados possam ser replicados (SAMPIERI; CALLADO; LUCIO, 2013).

3.8 ASPECTOS ÉTICOS

Os referenciais da Bioética, preconizados na resolução nº 466/12 do Conselho Nacional de Saúde (CNS) foram consideradas nesse estudo. O projeto foi submetido e ao Comitê de Ética em Pesquisa da Universidade Federal do Ceará – UFC, obtendo a autorização para utilização dos dados visto que os dados são públicos, disponibilizados pelo Sistema Datasus/SIM/SINASC. Serão mantidas a confidencialidade dos dados e o anonimato das crianças e mães envolvidas no estudo.

IV - LEVANTAMENTO DE DADOS

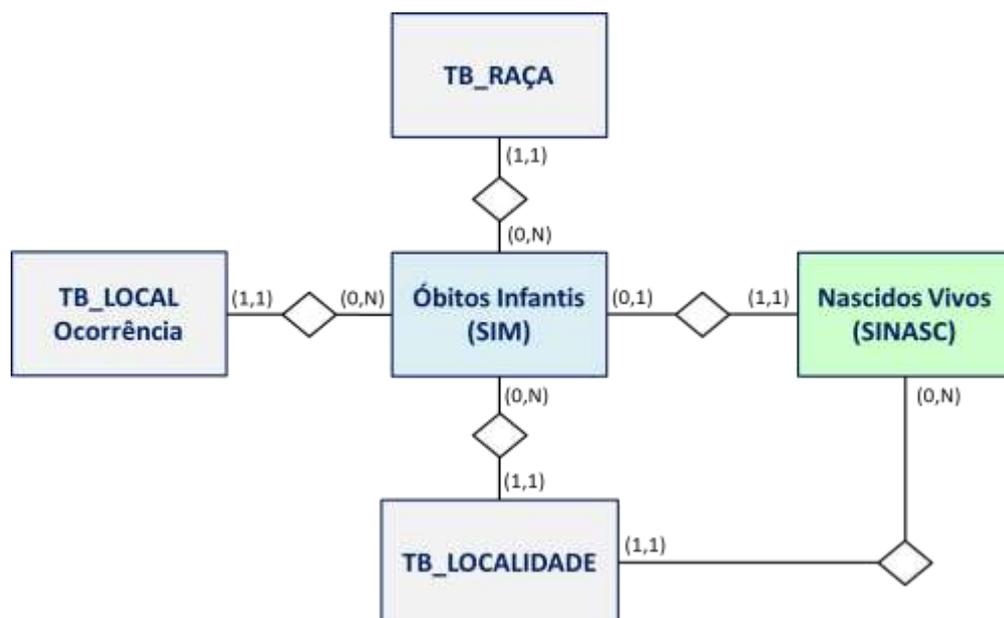
Os dados coletados estão dispostos em 02 (duas) principais tabelas: TB_NASCIDOS_VIVOS e TB_ÓBITOS_INFANTIS, que se relacionam através do atributo (campo) chave NU_DN (Número no cadastro de Nascidos Vivos). Para garantir a qualidade dos dados foi desenvolvido um MER (Modelo de Entidades e Relacionamentos), técnica utilizada em projeto de Banco de Dados Relacional que garante a integridade dos dados.

4.1 MER – MODELO DE ENTIDADES DE RELACIONAMENTOS

Um banco de dados relacional é projetado a partir dos modelos: Conceitual, Lógico e Físico. O modelo Conceitual descreve a estrutura do banco de dados independente do software de gerenciamento de Banco de dados, já o modelo Lógico deve definir quais as tabelas o banco contém, e para cada tabela, quais os atributos (campos). O modelo Físico é a implementação real do banco de dados em SGDB (Sistema de Gerenciamento de Banco de Dados), que no nosso caso foi utilizado o Software MySQL 5.7 (HEUSER, 2009).

A Figura 29, apresenta os relacionamentos entre as tabelas utilizadas no modelo e suas cardinalidades para garantir a integridade referencial do Banco de dados. Por exemplo: um tipo de raça cadastrada na tabela de Raças poderá ter nenhum óbito infantil ou N (vários), já um óbito infantil cadastrado obrigatoriamente terá um tipo de raça cadastrado na Tabela de raças.

Figura 29 – MER (Modelo de Entidades e relacionamentos).



Fonte: Elaborada pelo autor.

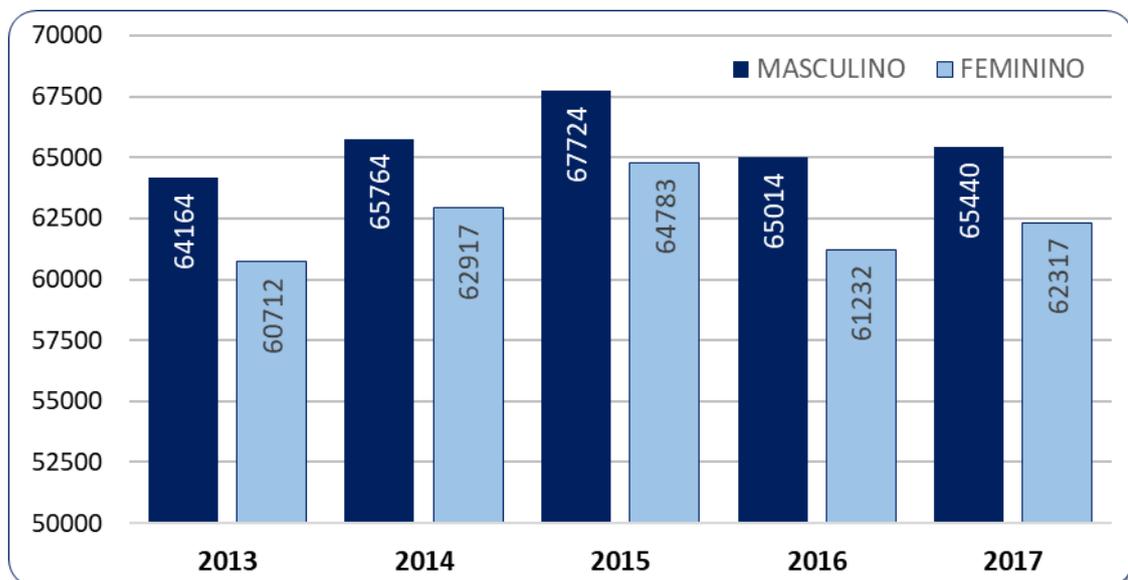
O MER apresentado na figura acima garante que um registro na tabela de nascidos vivos poderá ter ou não um registro correspondente na tabela de óbitos infantis, mas na leitura inversa, todo óbito infantil obrigatoriamente terá um registro na tabela de nascidos vivos. A partir desse relacionamento entre estas tabelas foi possível coletar as informações APGAR1 e APGAR5 que só estavam presentes na tabela de nascidos vivos.

4.2 DADOS COLETADOS

A partir dos dados coletados nas bases do SIM e SINASC foi possível fazer várias análises sobre a situação da Mortalidade Infantil e dos Nascidos Vivos no estado do Ceará entre 2013 e 2017, usando estatística descritiva, ramo da estatística que visa sumarizar e descrever qualquer conjunto de dados.

O Gráfico 10 apresenta a distribuição dos Nascidos Vivos por sexo no estado do Ceará no período estudado. Esse gráfico nos mostra que em todos os anos nasceram mais meninos do que meninas, em média nascem 3.200 meninos a mais do que meninas por ano no Ceará.

Gráfico 10 – Nascidos vivos no Ceará de 2013 a 2017 por sexo.

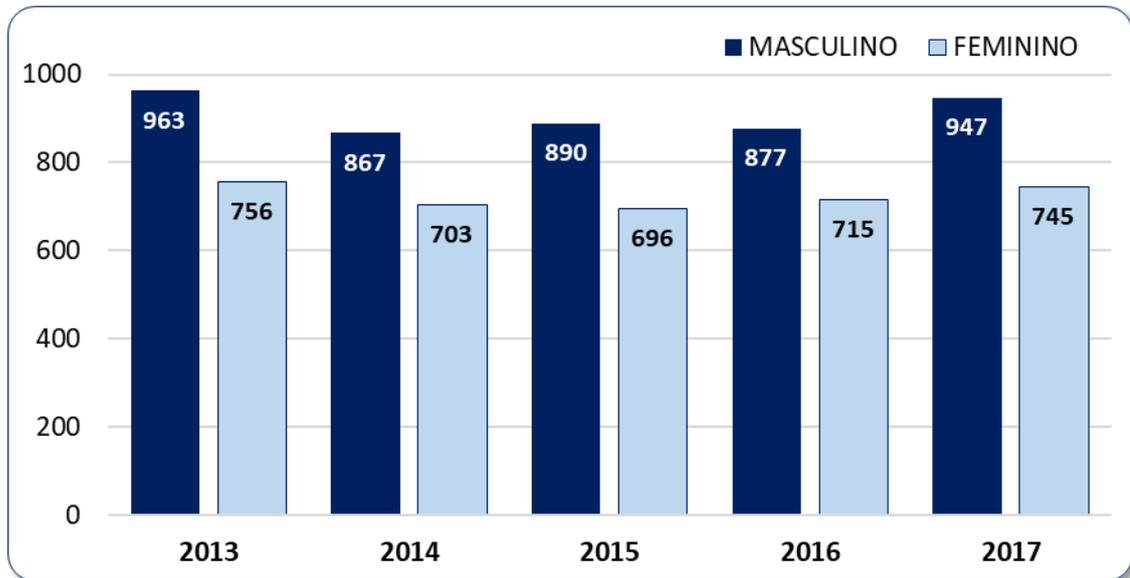


Fonte: Elaborada pelo autor a partir dos dados do Datasus/SINASC.

Importante salientar que 2013 foi o ano de menor natalidade no Ceará dentro do período estudado: em 2013 foram 124.876 crianças contra 132.507 crianças em 2015, ano com maior natalidade no período estudado, 7.631 crianças a mais, representando aumento de 6,1% de 2013 para 2015.

O Gráfico 11 apresenta a distribuição dos Óbitos Infantis por sexo no estado do Ceará no período estudado. Esse gráfico nos mostra que em todos os anos morreram mais meninos do que meninas, em média morreram 186 meninos a mais do que meninas por ano no Ceará.

Gráfico 11 – Óbitos infantis no Ceará de 2013 a 2017 por sexo.



Fonte: Elaborada pelo autor a partir dos dados do Datasus/SIM.

Em 2013 foi o ano com maior quantidade de óbitos infantis com 1.719 óbitos infantis, contra 2016 com 1.592 óbitos infantis, representando uma redução de 7,4% comparando 2013 com 2016, com uma redução absoluta de 127 óbitos. Infelizmente a mortalidade infantil voltou a crescer em 2017, com aumento de 6,3%, representado por 100 óbitos a mais de 2016 para 2017, provavelmente devido a epidemia dos vírus Zika e Chikungunya ocorridos nesse período.

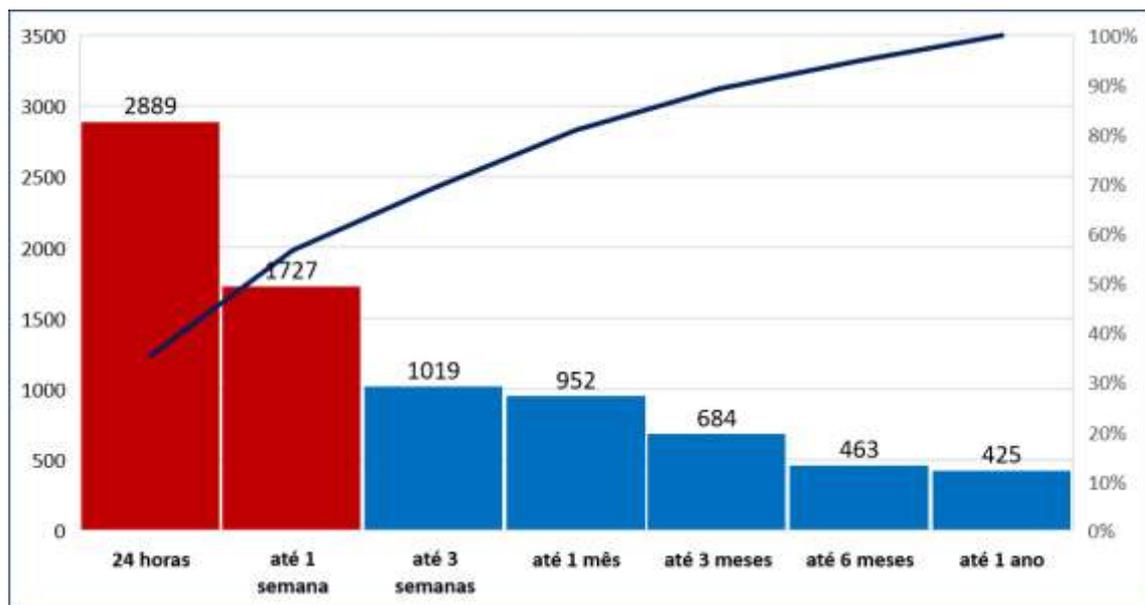
As análises dos gráficos de Mortalidade infantil e Natalidade no estado do Ceará nos apresentam uma proporcionalidade entre os sexos incompatível estatisticamente. Enquanto no Ceará nasceram 2,5% a mais de meninos do que meninas, era de se esperar uma proporcionalidade próxima de 2,5% também na mortalidade infantil, entretanto morreram em média no mesmo período 11,4% a mais de meninos do que de meninas. Essa é uma informação que precisa ser estudada e desvendada em trabalhos futuros.

No mesmo período estudado, quando analisamos o tempo de sobrevivência, observamos que 35,4% dos óbitos ocorreram nas primeiras 24 horas de vida e 56,6%, bem mais que a metade dos óbitos, ocorreram até a primeira semana de vida, confirmando a criticidade dos primeiros dias de vida dos bebês.

Os diagramas de Pareto são gráficos de barras verticais usados na identificação de fontes críticas dos efeitos de um problema. As categorias mostradas no eixo horizontal existem como uma distribuição de probabilidades válidas que representam 100% das possíveis observações. As frequências das ocorrências de cada faixa especificada estão listadas no eixo horizontal e diminuem em grandeza até totalizar a amostra. O diagrama de Pareto é normalmente organizado em categorias para medir frequências ou consequências (PMI, 2012).

O Gráfico 12 apresenta um gráfico de Pareto com os óbitos infantis no Ceará de 2013 a 2017, classificados em faixas por tempo de sobrevivência, são elas: [1] primeiras 24 horas; [2] até uma semana; [3] até três semanas; [4] até um mês; [5] até três meses; [6] até seis meses e [7] até um ano de vida.

Gráfico 12 – Gráfico de Pareto por tempo de sobrevivência.



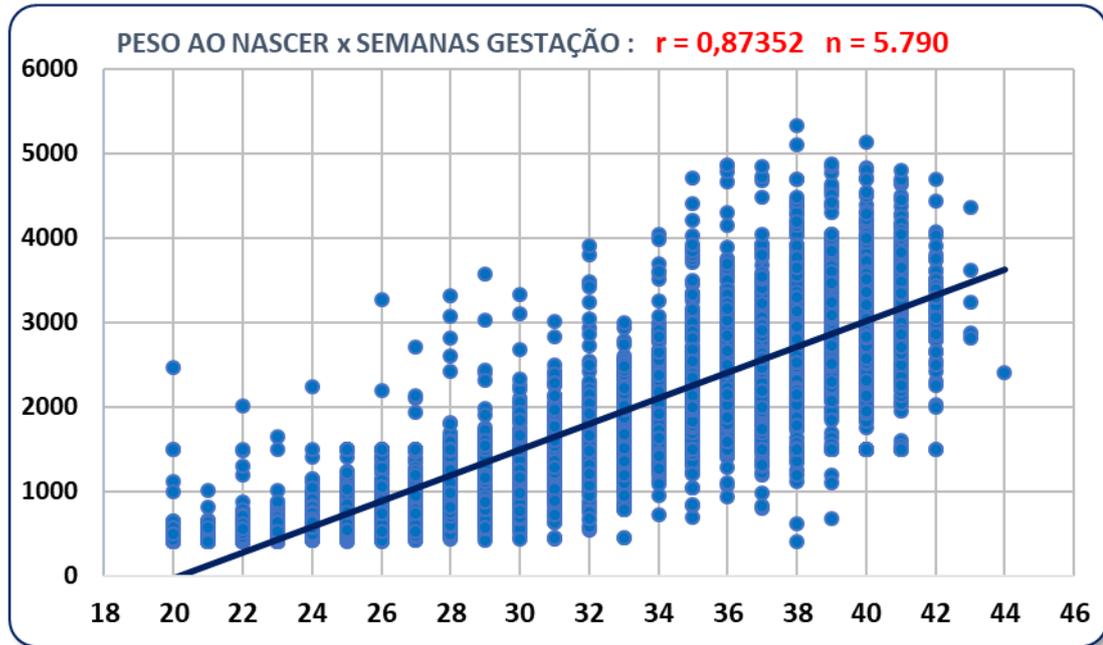
Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM.

No gráfico de Pareto acima, podemos confirmar a criticidade dos primeiros dias de vida dos bebês, onde mais de 80% dos óbitos infantis ocorreram até 30 dias de nascidos, confirmando a necessidade dos cuidados e acolhimentos no início de vida dos bebês.

O Gráfico 13 apresenta a correlação linear entre as variáveis “Peso ao nascer” e “Semanas de Gestação” na base de dados de óbitos infantis. No cálculo dessa correlação foram excluídos os óbitos com valores *outliers* de peso e sem informação da quantidade de semanas de gestação. Foi utilizada uma amostra com 5.790 óbitos infantis ocorridos no estado

do Ceará de 2013 a 2017. Neste gráfico só foram consideradas as gestações com 20 (vinte) semanas ou mais e pesos ao nascer iguais ou superiores a 400 (quatrocentos) gramas.

Gráfico 13 – Gráfico de dispersão - Óbitos: Peso ao nascer x Semanas de Gestação.

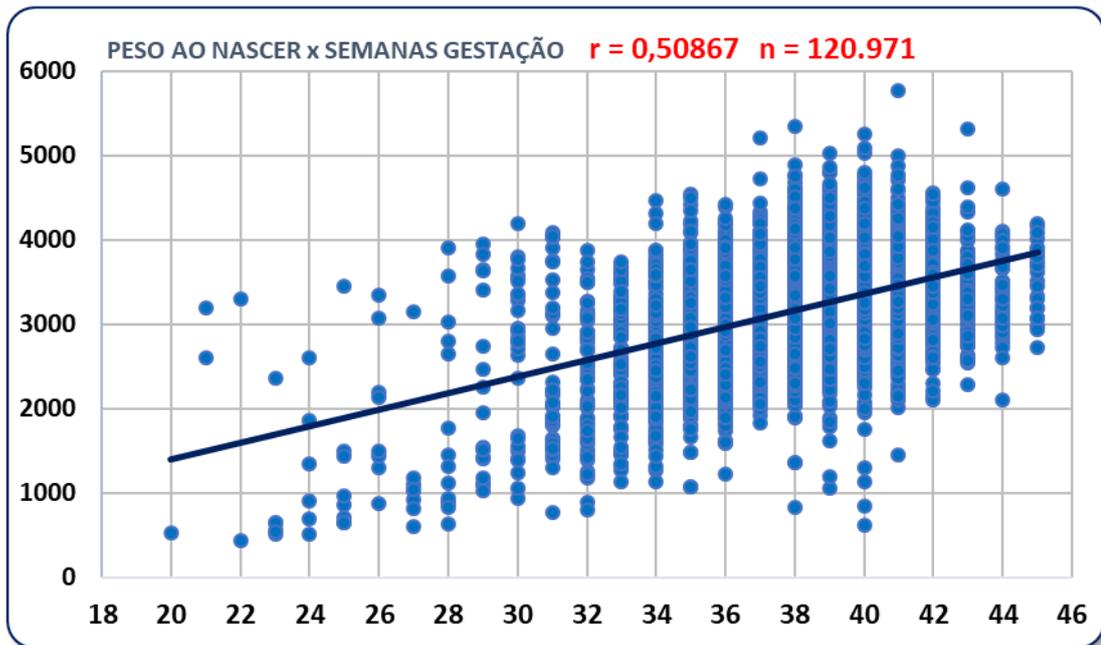


Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM.

O gráfico de dispersão acima, confirma um modelo considerado lógico: Quanto mais semanas de gestação, maior será o peso do bebê ao nascer. Essa informação está confirmada através do coeficiente de correlação ($r = 0,87352$) positivo e forte, isto é, próximo de 1,0 (um).

O Gráfico 14 apresenta a correlação linear entre as variáveis Peso ao nascer e Semanas de Gestação na base de dados de Nascidos vivos no estado do Ceará em 2017. No cálculo dessa correlação foram excluídos os nascidos vivos com valores *outliers* de peso e sem informação da quantidade de semanas de gestação. Foi utilizada uma amostra com 120.971 nascimentos ocorridos no estado do Ceará em 2017. Neste gráfico também só foram considerados nascidos vivos com 20 (vinte) semanas ou mais e pesos ao nascer iguais ou superiores a 400 (quatrocentos) gramas.

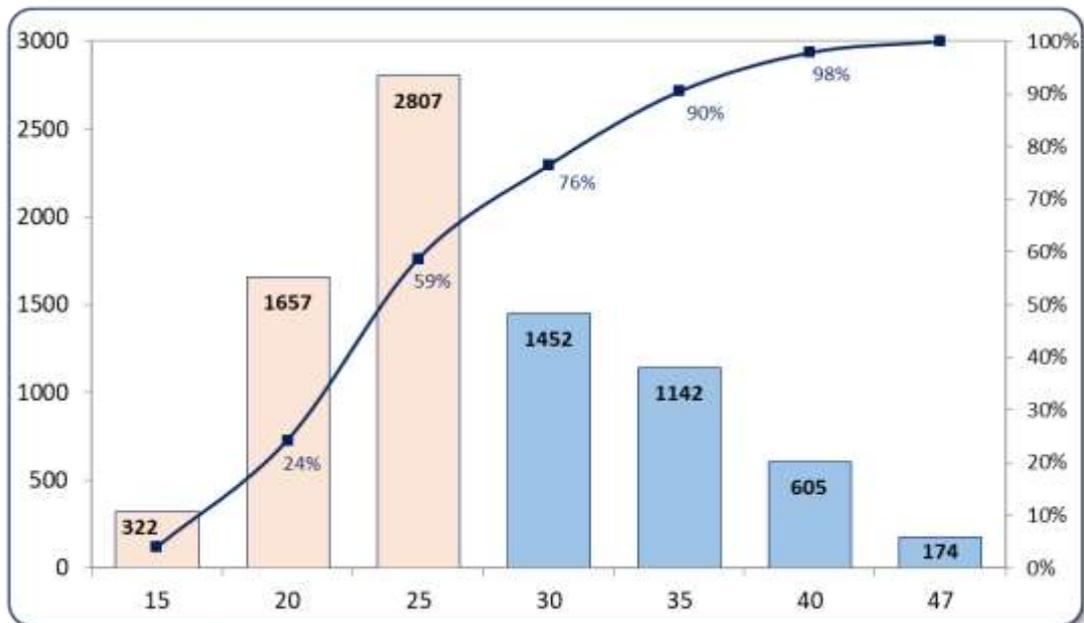
Gráfico 14 – Gráfico de dispersão – NV: Peso ao nascer x Semanas de Gestação.



Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM.

O gráfico de dispersão acima, confirma um modelo considerado lógico: Quanto mais semanas de gestação, maior será o peso do bebê ao nascer, através do coeficiente de correlação ($r = 0,50867$) positivo, entretanto não é tão forte quanto a correlação entre os óbitos infantis exposto anteriormente que atingiu ($r = 0,87352$), ficando próximo de 1,0 (um), isto é, esta correlação é mais forte entre os óbitos infantis.

A variável Idade da Mãe tem valores bem característicos da nossa região, onde a maior concentração de mulheres grávidas ocorre na faixa dos 25 anos. O Gráfico 15 é um histograma da frequência da idade da mãe, entre os 8.159 óbitos infantis ocorridos no estado do Ceará de 2013 a 2017.

Gráfico 15 – Histograma: Idade das mães – óbitos infantis no Ceará de 2013 a 2017.

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM.

O histograma acima nos apresenta um dado preocupante, quase 60% dos óbitos infantis de 2013 a 2017 ocorridos no Ceará as mães tinham até 25 (vinte e cinco) anos de idade, isto é, muito jovem. E praticamente em 1/4 (um quarto) dos óbitos infantis no período estudado, as mães tinha até 20 (vinte) anos de idade.

A Tabela 1 abaixo faz uma comparação do comportamento da variável “Idade da Mãe” entre os óbitos infantis e os nascidos vivos usando estatística descritiva. Foram calculadas 12 (doze) medidas estatísticas para uma amostra de 8.159 óbitos infantis no estado do Ceará de 2013 a 2017 e para a amostra de 127.719 nascidos vivos no estado do Ceará em 2017.

Tabela 1 - Comparativo “Idade da Mãe” entre óbitos infantis x nascidos vivos.

ESTATÍSTICA	Óbitos de 2013 a 2017	Nascidos Vivos em 2017
Média	25,69040324	26,24587571
Erro padrão	0,073992209	0,01862296
Mediana	25	26
Moda	25	22
Desvio padrão	6,683507807	6,65543534
Variância da amostra	44,66927661	44,29481956
Curtose	-0,290782046	-0,634806964
Assimetria	0,45086798	0,313058922
Intervalo	37	41
Mínimo	10	13
Máximo	47	54
Amostra	8159	127719

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM usando Linguagem Python.

Na tabela comparativa acima, podemos observar que a média de idade das mães na coluna óbitos infantis é 2,2% menor que as mães de todos os nascidos vivos em 2017. A mediana é 25 contra 26 dos nascidos vivos. Esses dados não são conclusivos, mas nos permite inferir que “a mortalidade infantil tem alguma relação com a pouca idade das mães”. Novamente é uma pesquisa e debate para um estudo futuro.

A curtose destas amostras é do tipo Platicúrtica, isto é, quando os valores são negativos (-0,29078 e -0,63480 respectivamente) e seu achatamento fica abaixo da distribuição Normal, também chamada de Mesocúrtica (curtose = 0). A curtose da idade das mães de nascidos vivos é menos da metade da curtose dos óbitos infantis, isto é, o achatamento da distribuição destas idades é bem maior do que das idades de mães com óbitos infantis.

Em uma distribuição de probabilidades ou mesmo em um *dataset*, a Curtose (*Kurtosis*) é uma medida importante que nos apresenta o quão “pontuda” ou “achatada” é uma distribuição (MELO JÚNIOR, 2018).

Curtose é uma medida de dispersão que caracteriza o "achatamento" da curva da função de distribuição. A curtose pode ser definida como o grau de achatamento de uma distribuição de frequências (altura do pico da curva) em relação a uma distribuição teórica que geralmente corresponde à distribuição normal (BELFIORE, 2015).

É definida pela fórmula:

$$b_2 = \frac{1}{n} \sum \left[\frac{x_i - \bar{x}}{s} \right]^4 - 3$$

Se $b_2 = 0$, então a função de distribuição tem o mesmo achatamento da distribuição normal, chamamos essas funções de Mesocúrticas. Se $b_2 > 0$, dizemos que a função de distribuição é Leptocúrtica e possui a curva da função de distribuição mais afunilada com um pico mais alto do que a distribuição normal. Se $b_2 < 0$, então a função de distribuição é mais achatada do que a distribuição normal, dizemos que esta curva da função de distribuição é Platicúrtica (BELFIORE, 2015).

Ainda de forma comparativa podemos extrair informações relevantes entres os óbitos infantis e os nascidos vivos. A Tabela 2 apresenta o perfil das gestantes através da distribuição relativa por faixa etária e escolaridade (anos de estudos) da tabela de óbitos infantis no Ceará de 2013 a 2017.

Tabela 2 – Perfil das gestantes (Óbitos infantis) por Faixa etária x Escolaridade.

ÓBITOS INFANTIS no CEARÁ: 2013 a 2017					
FAIXA ETÁRIA	ESCOLARIDADE DA MÃE (em anos de estudos)				
	ZERO	de 01 a 03	de 04 a 07	de 08 a 11	12 ou mais
Até 15	2,2%	6,5%	67,0%	24,3%	0,0%
16 a 20	1,6%	6,7%	34,0%	54,9%	2,8%
21 a 25	5,3%	7,7%	26,5%	52,2%	8,4%
26 a 30	3,1%	8,7%	21,5%	51,7%	15,1%
31 a 35	3,2%	12,9%	20,8%	42,9%	20,2%
36 a 40	6,6%	18,8%	23,5%	32,4%	18,6%
41 acima	4,9%	21,8%	29,6%	21,8%	21,8%

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Na tabela acima podemos observar que a baixa escolaridade é muito forte principalmente entre as mães mais jovens e entre as mais maduras: entre as mães da primeira faixa etária (até 15 anos), mais de 75% não concluíram o Ensino Fundamental I; entre as mães da segunda faixa etária (de 16 a 20 anos), 42,3% também não concluíram o Ensino Fundamental I; já as mães da faixa de 36 a 40 anos, 48,9% não concluíram o Ensino Fundamental I; e na última faixa etária, com 41 anos ou mais, 56,3% não concluíram o Ensino Fundamental I, com um agravante, nas últimas faixas etárias encontramos 6,6% e 4,9% de mães analfabetas, respectivamente. A partir dessas informações, podemos inferir que a baixa escolaridade das mães tem peso importante na mortalidade infantil.

A Tabela 3 apresenta o perfil das gestantes através da distribuição relativa por faixa etária e escolaridade (anos de estudos) da tabela de Nascidos vivos no Ceará em 2017.

Tabela 3 – Perfil das gestantes (Nascidos Vivos) por Faixa etária x Escolaridade.

NASCIDOS VIVOS no CEARÁ em 2017					
FAIXA ETÁRIA	ESCOLARIDADE DA MÃE (em anos de estudos)				
	ZERO	de 01 a 03	de 04 a 07	de 08 a 11	12 ou mais
até 15	0,1%	2,2%	48,3%	49,3%	0,0%
16 a 20	0,1%	1,3%	23,1%	72,6%	2,8%
21 a 25	0,2%	1,6%	17,1%	70,5%	10,6%
26 a 30	0,3%	2,1%	16,7%	60,5%	20,4%
31 a 35	0,7%	4,0%	16,2%	51,9%	27,2%
36 a 40	2,1%	7,4%	19,4%	43,9%	27,2%
41 acima	4,2%	12,2%	25,3%	37,1%	21,2%

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Na tabela acima, analisando os nascidos vivos, podemos observar que o nível de escolaridade entre as mães mais jovens e as mais maduras teve melhoria consistente em relação as mães de óbitos infantis que não concluíram o Ensino Fundamental I: entre as mães da primeira faixa etária (até 15 anos), baixou de mais de 75% para 50,6%; entre as mães da segunda faixa etária (de 16 a 20 anos), baixou de 42,3% para 24,5%; já as mães da faixa etária de 36 a 40 anos, baixou de 48,9% para 28,9%; e na última faixa etária, com 41 anos ou mais, de 56,3% para 41,7%. Esta avaliação comparativa só confirma a importância do nível de escolaridade das mães para a redução da mortalidade infantil.

A variável estado civil aparentemente teria uma influência na mortalidade infantil, entretanto na Tabela 4 abaixo, podemos observar que o comparativo da amostra de 4.816 óbitos infantis no Ceará de 2013 a 2017 e 122.364 Nascidos vivos no Ceará em 2017, a proporcionalidade é muito parecida para os dois casos.

Tabela 4 – Comparativo do estado civil das mães: óbitos infantis x nascidos vivos.

ESTADO CIVIL	ÓBITOS INFANTIS (2013 a 2017)	NASCIDOS VIVOS (2017)
Solteira	38,5%	39,8%
Casada	27,2%	31,7%
Viúva	0,3%	0,2%
Separada judicialmente	0,5%	0,6%
União consensual	33,0%	27,5%
Ignorado	0,4%	0,3%

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Quando consideramos a união estável (Casada + União consensual), os valores são ainda mais próximos: 60,2% para as mães de óbitos infantis e 59,2% para nascidos vivos, confirmado que esta variável não tem grande influência na mortalidade infantil no estado do Ceará no período estudado.

Quando avaliamos a variável “Quantidade de consultas pré-natal”, observamos a importância deste procedimento na mortalidade infantil. Na Tabela 5 abaixo, podemos observar diferenças graves no comparativo proporcional da amostra de 4.816 óbitos infantis no Ceará de 2013 a 2017 e 127.753 Nascidos vivos no Ceará em 2017. Nesta tabela só foram usados os dados que continham a informação de consultas pré-natal.

Tabela 5 – Consultas pré-natal: óbitos infantis x nascidos vivos.

CONSULTAS PRÉ-NATAL	ÓBITOS INFANTIS (2013 a 2017)	NASCIDOS VIVOS (2017)
Nenhuma	7,1%	2,5%
de 01 a 03	22,5%	4,6%
de 04 a 06	35,7%	22,4%
07 ou mais	32,9%	70,2%
Ignorado	1,8%	0,2%

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Na tabela acima, quando comparamos: mães que fizeram até 03 (três) consultas, a diferença é de 29,6% nos óbitos infantis contra 7,1% nos nascidos vivos, essa discrepância vai aparecer quando consideramos as mães que fizeram 04 (quatro) ou mais consultas, pois encontramos 68,6% nos óbitos infantis contra 92,6% nos nascidos vivos, uma diferença de 24% a mais para os nascidos vivos, confirmando a importância destes procedimentos na garantia de vida dos nossos bebês.

Vale ressaltar ainda em relação a tabela acima, que em 7,1% dos óbitos infantis as mães não realizaram nenhuma consulta pré-natal, deixando a gestação e a vida destes bebês totalmente entregues a sorte sem nenhuma orientação profissional.

A Tabela 6 abaixo, faz uma avaliação da frequência das mães que fizeram consultas pré-natal em relação a sua faixa etária. Nesta tabela é possível comparar uma amostra de 4.627 mães da tabela óbitos infantis com uma amostra de 127.435 mães da tabela de nascidos vivos em 2017 de forma proporcional.

Tabela 6 – Comparativo: Faixa etária x Consultas pré-natal.

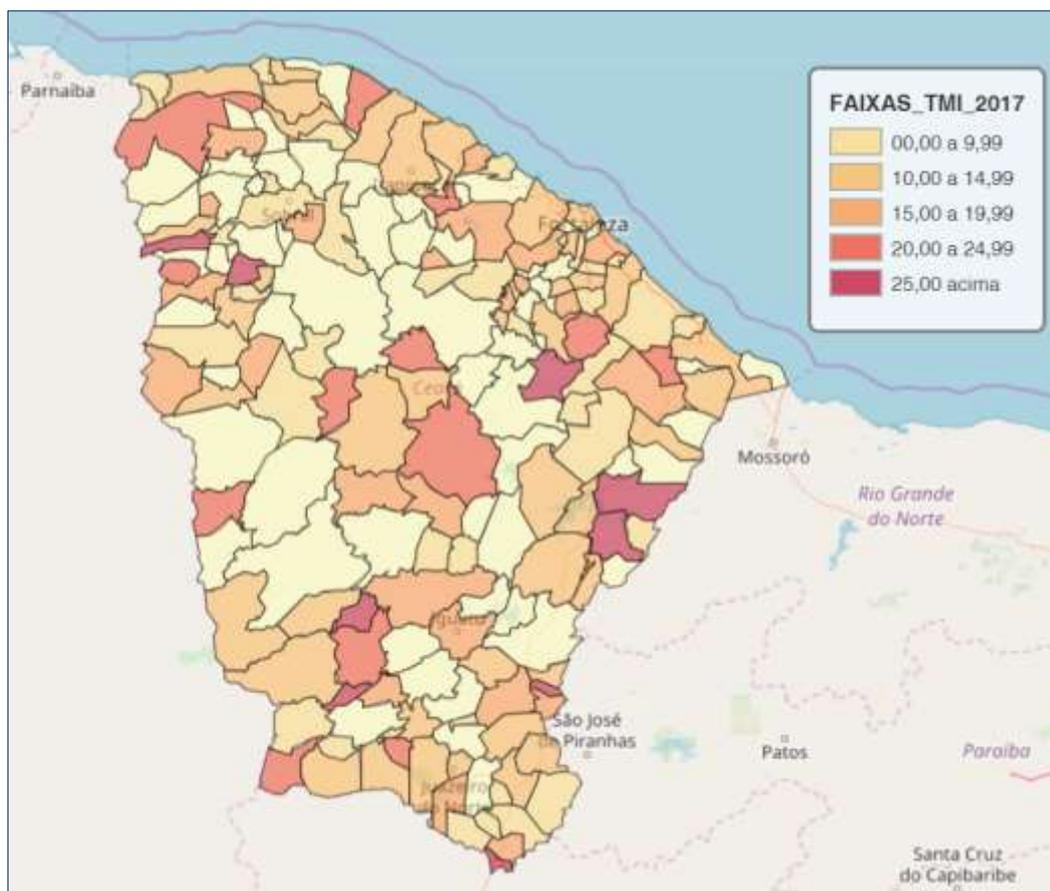
FAIXA ETÁRIA	ÓBITOS INFANTIS no CEARÁ: 2013 a 2017				NASCIDOS VIVOS no CEARÁ em 2017			
	CONSULTAS PRÉ-NATAL				CONSULTAS PRÉ-NATAL			
	Nenhuma	de 01 a 03	de 04 a 06	07 ou mais	Nenhuma	de 01 a 03	de 04 a 06	07 ou mais
Até 15	10,6%	35,2%	35,2%	19,0%	1,9%	7,9%	31,9%	58,3%
16 a 20	8,5%	28,1%	34,9%	28,5%	1,6%	6,3%	27,5%	64,6%
21 a 25	6,8%	22,2%	35,6%	35,3%	1,8%	5,1%	22,9%	70,3%
26 a 30	7,4%	18,1%	37,9%	36,6%	2,4%	3,5%	19,7%	74,5%
31 a 35	5,8%	21,9%	35,9%	36,4%	3,8%	3,4%	19,1%	73,7%
36 a 40	5,3%	17,7%	39,4%	37,6%	4,2%	4,2%	20,5%	71,1%
41 acima	6,7%	23,1%	40,4%	29,8%	3,6%	5,1%	23,9%	67,5%

Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Na tabela acima, é possível observar entre os óbitos infantis que 45,8% das mães mais jovens, até 15 anos, não fizeram o pré-natal ou fizeram só até 03 (três) consultas, contra 9,8% entre os nascidos vivos. Na faixa etária entre 16 a 20 anos também encontramos uma diferença grande entre as mães que fizeram até 03 (três) consultas: 36,6% entre os óbitos infantis contra 7,9% entre os nascidos vivos. Essas diferenças vão emergir na tabela nascidos vivos que em todas as faixas etárias mais de 90% das mães fizeram 04 (quatro) ou mais consultas pré-natal, com destaque para a faixa etária entre 26 a 30 anos que mais de 94% das mães fizeram 04 (quatro) ou mais consultas, confirmando a importância do pré-natal na garantia de vida dos bebês e das mães.

A Figura 30 apresenta o georreferenciamento do mapa do estado do Ceará por faixa de mortalidade infantil em 2017, onde é possível identificar grandes áreas do estado com baixas taxas de mortalidade infantil, mas que algumas regiões ainda dependem de ações de saúde pública para alcançar índices aceitáveis. A relação completa dos municípios do estado do Ceará está no Anexo III.

Figura 30 – Georreferenciamento do estado do Ceará por TMI em 2017.



Fonte: Elaborado pelo autor usando dados do Datasus/SIM e linguagem Python com Matplotlib.

Entre as localidades com maiores índices de Mortalidade Infantil em 2017 no Ceará, destacam-se negativamente as localidades na faixa acima de 25,0 óbitos infantis para cada mil nascimentos, destacadas no mapa do estado do Ceará na cor vermelha mais forte. A Tabela 7 apresenta a lista de localidade desta faixa.

Tabela 7 – Relação de localidades do Ceará com os maiores TMI em 2017.

CD_LOCAL	NOME LOCALIDADE	MACRO REGIÃO	POPULAÇÃO	IDHM	TMI_2017
231126	Quiterianópolis	Sertões	20860	0,594	25,10
230530	Ibiapina	Noroeste	24825	0,608	26,01
230650	Itapiúna	Norte	20014	0,604	26,55
230070	Alto Santo	Jaguaribe	16976	0,601	27,78
230180	Baixio	Centro-Sul	6228	0,627	27,78
231170	Reritaba	Noroeste	18769	0,601	29,79
230080	Antonina do Norte	Centro-Sul	7278	0,599	31,25
230600	Iracema	Jaguaribe	14125	0,652	32,26
231300	Solonópole	Sertões	18158	0,625	33,65
231280	Senador Sá	Noroeste	7513	0,603	33,71
231060	Penaforte	Sul	8956	0,646	36,14
230360	Catarina	Sertões	20451	0,618	37,04
230427	Ererê	Jaguaribe	7163	0,610	43,48
230526	Ibaretama	Sertões	13218	0,577	46,63

Fonte: Elaborado pelo autor usando dados do Datasus/SIM/SINASC.

Na tabela acima podemos observar as 14 (quatorze) localidades do Ceará que figuram entre os piores TMI, ficando na faixa acima de 25,0 (vinte e cinco), sendo que a região dos Sertões contribuiu com quatro localidades: Quiterianópolis, Solonópole, Catarina e Ibaretama. Esta tabela ainda nos mostra que 11 (onze) localidades mesmo com o IDHM considerado médio (entre 0,600 e 0,699) ainda assim figuram entre os piores TMI do estado em 2017.

4.3 MEDIDAS DE APGAR

A escala ou índice conhecido como Apgar, foi criado em 1949 pela anestesista inglesa, Dra. Virgínia Apgar. Esta é uma forma rápida, fácil e eficaz de avaliar as condições de vitalidade do recém-nascido.

O cálculo consiste na avaliação de 05 (cinco) sinais objetivos observados no primeiro minuto de vida (Apgar1) e no quinto minuto após o nascimento (Apgar5), são eles: a frequência cardíaca, o esforço respiratório, o tônus muscular, a cor da pele e a irritabilidade reflexa. Para cada um destes itens é atribuída uma nota de 0 a 2, conforme a gravidade (SÃO PAULO, 2011).

Para a frequência cardíaca, verifica-se a presença ou não dos batimentos cardíacos e se estão acima ou abaixo de 100 batimentos por minuto. Para o esforço respiratório, observa-se se o choro está forte, fraco ou ausente. Quanto ao tônus muscular, observa-se se o bebê tem boa flexão dos membros ou se está flácido. A cor da pele é avaliada observando-se se a pele está rosada ou azulada, parcial ou total. E finalmente quanto a irritabilidade reflexa, verifica-se se o bebê reage ou não aos estímulos com choro forte.

Para calcular o Apgar1 e Apgar5 é utilizada o Quadro 4 que pontua o resultado como zero, um ou dois.

Quadro 4 – Cálculo da pontuação do Apgar.

PONTUAÇÃO	0	1	2
FREQUÊNCIA CARDÍACA	AUSENTE	< 100/minuto	> 100/minuto
ESFORÇO RESPORATÓRIO	AUSENTE	CHORO FRACO	CHORO FORTE
TÔNUS MUSCULAR	FLÁCIDO	FLEXÃO DE PERNAS E BRAÇOS	MOVIMENTO ATIVO/BOA FLEXÃO
COR DA PELE	CIÂNOTICO/PÁLIDO	CIABOSE EXTREMIDADES	ROSADO
IRRITABILIDADE REFLEXA	AUSENTE	ALGUM MOVIMENTO	ESPIRROS/CHORO

Fonte: Elaborada pelo autor a partir de SÃO PAULO (2011).

O somatório da pontuação, poderá ter o mínimo de zero e o valor máximo de dez, que resultará no Índice de Apgar que classifica o recém-nascido conforme o Quadro 5 abaixo.

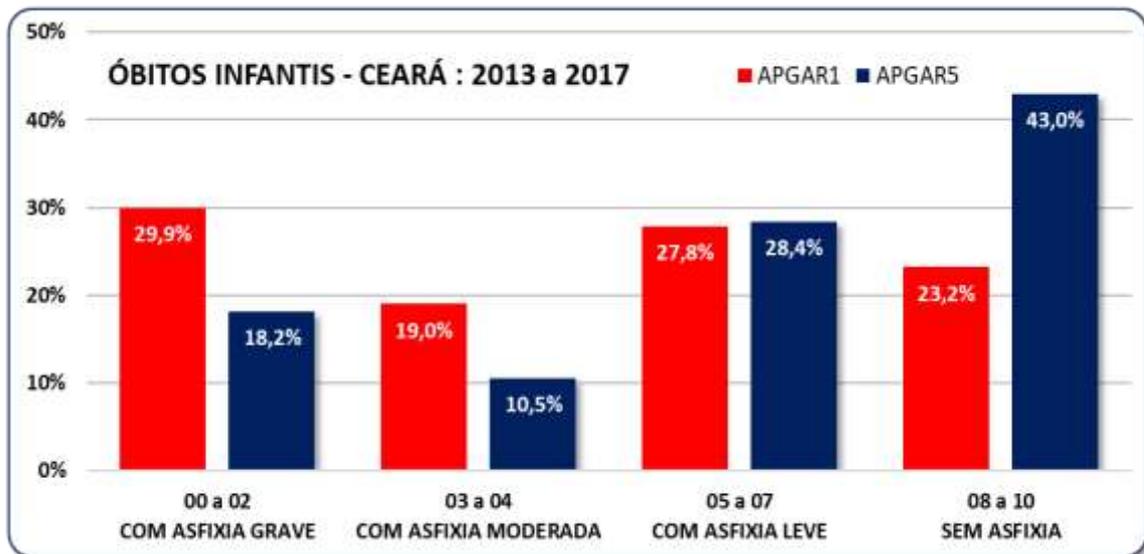
Quadro 5 – Classificação da pontuação do Apgar no 1º e 5º minuto.

CLASSIFICAÇÃO	PONTUAÇÃO
COM ASFIXIA GRAVE	APGAR de 00 a 02
COM ASFIXIA MODERADA	APGAR de 03 a 04
COM ASFIXIA LEVE	APGAR de 05 a 07
SEM ASFIXIA	APGAR de 08 a 10

Fonte: Elaborada pelo autor a partir de SÃO PAULO (2011).

A importância das variáveis Apgar1 e Apgar5 pode ser observada no Gráfico 16 que apresenta a distribuição da classificação do Apgar1 e Apgar5 entre os mais de 8.000 (oito mil) óbitos infantis coletados no estado do Ceará de 2013 a 2017.

Gráfico 16 – Distribuição dos Óbitos Infantis a partir do Apgar1 e Apgar5.



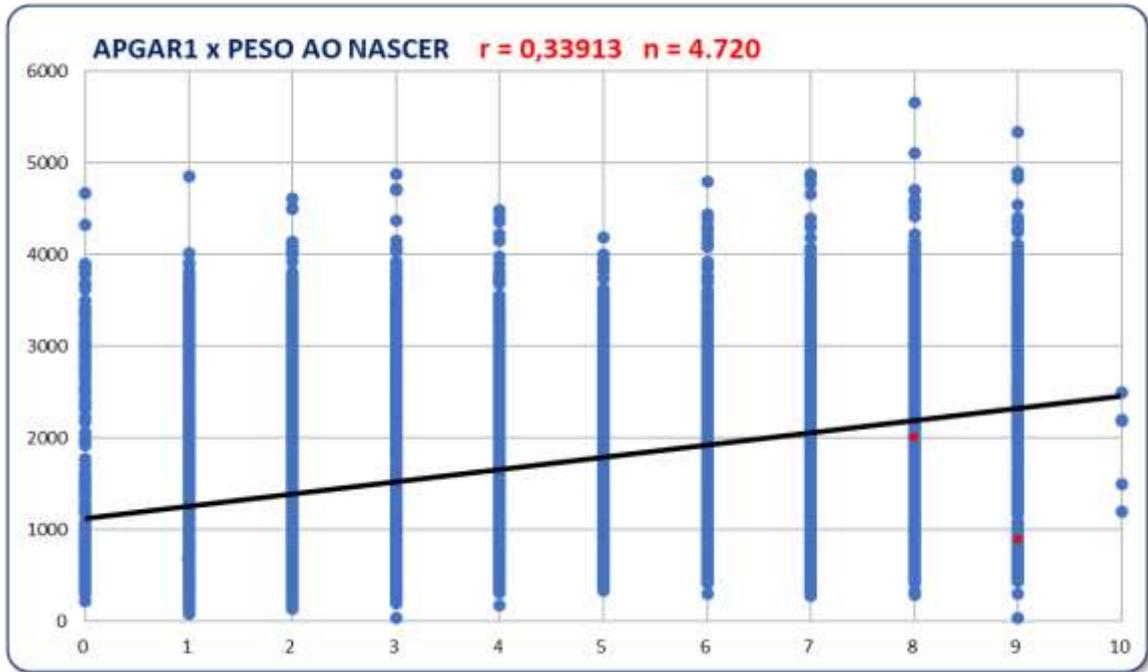
Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

No gráfico acima podemos observar que 48,9%, isto é, quase a metade dos óbitos infantis obtiveram pontuação do Apgar1 abaixo de 05 (cinco). E apenas 23,2%, isto é, menos de 1/4 (um quarto) dos óbitos infantis estavam classificados como “SEM ASFIXIA” com a pontuação entre 08 (oito) e 10 (dez). Esses valores nos mostram da importância da avaliação das medidas de Apgar para agilizar procedimentos urgentes que salvam vidas.

Quando avaliamos as variáveis Apgar1 e Apgar5 nos mais de 631 mil nascidos vivos no Ceará entre 2013 e 2017, observamos que 97,8% dos nascidos vivos obtiveram Apgar1 maior ou igual a 05 (cinco) e 99,5% dos nascidos vivos obtiveram Apgar5 maior ou igual a 05 (cinco), isto é, classificado como “ASFIXIA LEVE” ou “SEM ASFIXIA”, resultados bem superiores aos encontrados entre os óbitos infantis, de 51,0% no Apgar1 e 71,4% no Apgar5.

O Gráfico 17 apresenta informações relevantes quando calculamos a correlação linear entre as variáveis Apgar1 e peso ao nascer da criança. Este gráfico de dispersão apresenta 4.720 óbitos infantis ocorridos no estado do Ceará de 2013 a 2017.

Gráfico 17 – Gráfico de Dispersão – Óbitos infantis: Apgar1 x Peso ao nascer.

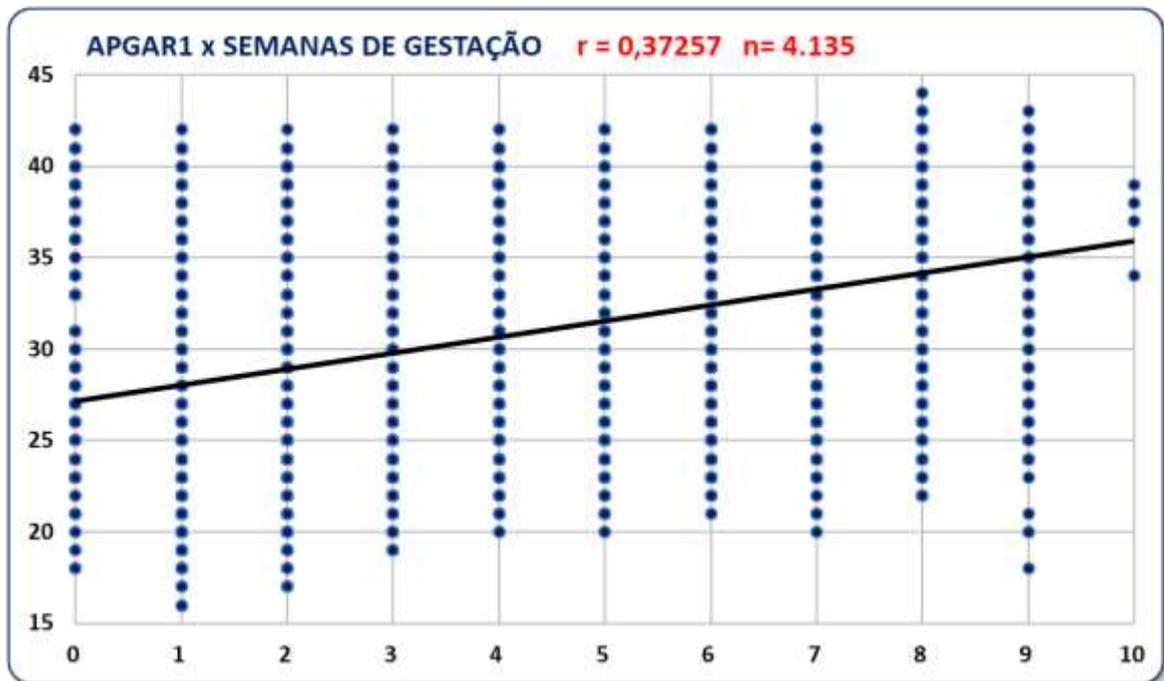


Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Observe no gráfico acima que embora o coeficiente de correlação seja considerado fraco, $r = 0,33913$ próximo de 0,5 (meio), podemos observar uma linha de tendência crescente, isto é, quanto maior o peso ao nascer, maior será o Apgar1. Quando calculamos esse coeficiente linear usando o Apgar5, o valor de “r” fica um pouco mais fraco, caindo para $r = 0,31944$, mas ainda predominando a mesma tendência, quanto maior o peso ao nascer, maior o Apgar.

Complementando a análise das variáveis Apgar1 e Apgar5, o Gráfico 18 apresenta a correlação entre Apgar1 x Semanas de gestação entre os óbitos infantis.

Gráfico 18 – Gráfico de Dispersão – Óbitos infantis: Apgar1 x Semanas de gestação.



Fonte: Elaborado pelo autor a partir dos dados do Datasus/SIM/SINASC.

Observe que novamente o coeficiente de correlação é considerado fraco, $r = 0,37257$ próximo de 0,5 (meio), mas também podemos observar uma linha de tendência crescente, isto é, quanto mais semanas de gestação, maior será o Apgar1. Quando calculamos esse coeficiente linear usando o Apgar5, o valor de “r” praticamente não se altera, $r = 0,37233$, predominando a mesma tendência, quanto mais semanas de gestação, maior o Apgar.

As análises realizadas com as variáveis Apgar1, Apgar5, Peso ao nascer e Semanas de gestação, nos explicam através do formalismo matemático e estatístico o que a lógica e o bom senso já previa, mas agora confirmado por um modelo de correlação que a tendência existe, mas não é tão forte quanto se imagina informalmente.

4.4 PROCESSO DE GERAÇÃO DO MODELO DE PREDIÇÃO

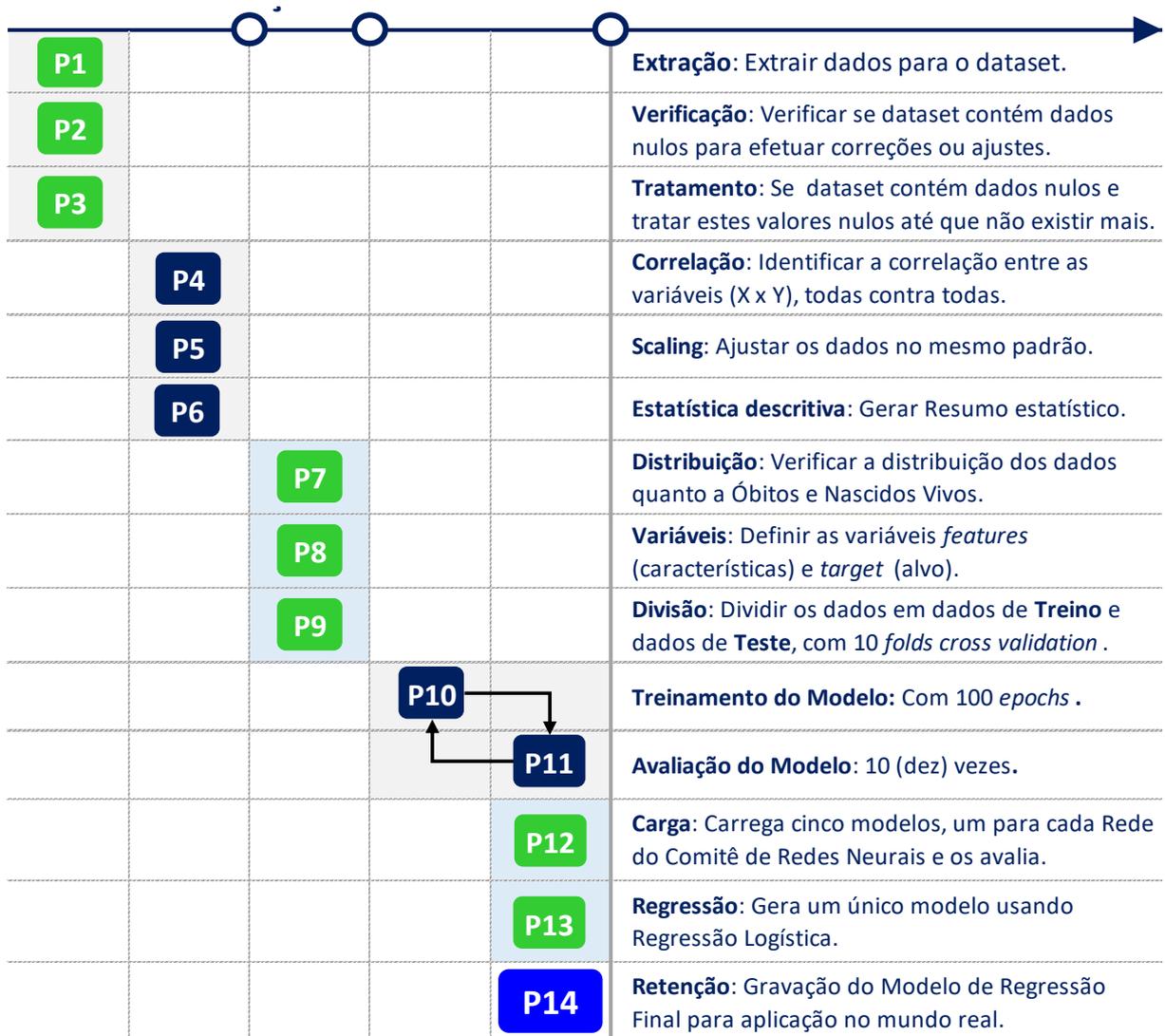
Para gerar os modelos preditivos foram usados: dados históricos da mortalidade infantil e dos nascidos vivos, estatísticas e algoritmos de Redes Neurais, dessa forma foi possível inferir ou predizer o que ocorreria no futuro, assim poderemos usar esses conhecimentos para evitar ou reduzir a mortalidade infantil no Ceará e no Brasil.

Ao se desejar concluir com evidências incompletas podemos usar a estatística através das deduções probabilísticas. Quando deduzimos a partir da probabilidade, estamos usando a inferência estatística. Para recorrer à inferência estatística, usamos o raciocínio indutivo e a teoria das probabilidades (FEIJOO, 2017).

V - ANÁLISE DOS DADOS

Para geração dos modelos preditivos foi utilizado um processo de preparação, tratamento, processamento e avaliação dos dados, conforme as fases descritas na Figura 31.

Figura 31 – Fases para geração dos modelos preditivos.



Fonte – Adaptação do autor.

A partir do processo desenhado acima, realizamos vários processamentos dos dados coletados e armazenados no nosso banco de dados Mysql em busca dos ajustes e acertos ideais para geração do modelo. Cada processamento precisou de quase 02 (duas) horas em uma CPU I7 de 8ª geração com 16 GB de Memória RAM e 128 GB de Memória SSD, gravando os resultados em um disco rígido de 1 TB de espaço.

O Apêndice V - *SCRIPT PYTHON PARA EXECUÇÃO* apresenta a codificação do *script* em linguagem Python que foi executado na geração do modelo, destacando cada passo do processo.

O Apêndice VI - *EXECUÇÃO DO SCRIPT PYTHON* apresenta de forma resumida os resultados obtidos na execução do modelo a partir do *Script Python*. Neste Apêndice é possível visualizar os resultados obtidos em cada etapa do processamento final para geração da matriz de pesos sinápticos.

A execução do processo utilizou uma estrutura de Comitê com 05 (cinco) Redes Neurais em *Stacked* (empilhamento) com Rede Neural *Deep Learning* com 04 (quatro) camadas e estratégia de *split* dos dados com *K-Folds* (K=10) e modelo final ajustado com *Logistic Regression*, obtendo os seguintes resultados:

Fase I: Foram extraídos os dados para um *DataSet* com 18.394 (dezoito mil, trezentos e noventa e quatro) fatos com 23 (vinte e três) colunas (variáveis ou atributos), sendo 8.159 óbitos infantis e 10.235 nascidos vivos que sobreviveram ao primeiro ano de vida.

Fase II: Foi executado um algoritmo para verificar se o *DataSet* continha valores nulos em sua formação. Este processo retornou o resultado TRUE, necessitando de uma ação do *script* para eliminar essa questão.

Fase III: Os dados nulos foram tratados usando técnicas de preenchimento baseados na média do atributo, restando 18.394 (dezoito mil, trezentos e noventa e quatro) fatos com 20 (vinte) colunas (variáveis ou atributos).

Fase IV: Foi executado um algoritmo para identificar a correlação entre as variáveis do *DataSet*. Os resultados mais significativos foram: APGAR1 e APGAR5 com 0,73602; APGAR1 e HORAS_VIDA com 0,62382; PESO e HORAS_VIDA com 0,58139 e PESO e APGAR1 com 0,54601. A relação com todas as correlações calculadas está no Apêndice VII – MATRIZ DE CORRELAÇÃO DAS VARIÁVEIS *FEATURES*.

Fase V: Os dados foram ajustados em *Scaling* para que todos os dados estivessem na mesma proporção de valores, isto é, nenhuma variável tinha um *Range* muito diferente das demais.

Fase VI: Resumo estatístico de todas as variáveis utilizadas, detalhando a Média, Desvio Padrão, Valor Mínimo, Valor Máximo, Primeiro Quartil (25%), Segundo Quartil (50%) e o Terceiro Quartil (75%).

Fase VII: Verificação da distribuição dos dados: 8.159 Casos de óbitos infantis e 10.235 Casos de Nascidos vivos que não vieram a óbito no primeiro ano de vida, com Percentual de 44.356855496357504% de óbitos e 55.643144503642496% de nascidos vivos.

Fase VIII: Nesta fase são identificadas as variáveis *features* (explicativas), ou seja, as características dos eventos que explicam as variáveis *targets* (alvos). No nosso processo a variável *target* utilizada é binária (1 – Óbito Infantil e 2 – Não houve óbito infantil). O Quadro 6 apresenta as variáveis explicativas e a variável Alvo utilizadas no processamento dos dados.

Quadro 6 – Variáveis explicativas e variável Alvo.

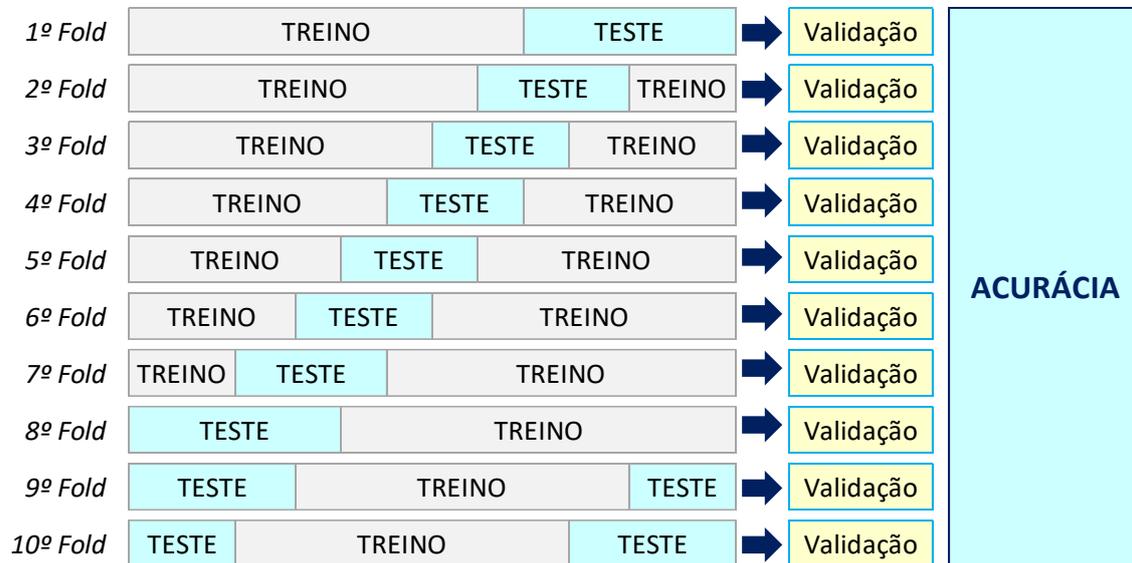
NM_VARIÁVEL	DESCRIÇÃO
NU_DO	Número do Documento de óbito
NU_DN	Número do Documento de Nascido Vivo
MM_OBITO	Mês da ocorrência do Óbito
HORAS_VIDA	Quantidade de Horas de vida no caso de Óbito
SEXO	1 - Masculino e 2 - Feminino
RAÇA	1 – Branca; 2 – Preta; 3 – Amarela; 4 – Parda e 5 – Indígena
IDH_MUN_RES	Faixa do IDH (Índice do Desenvolvimento Humano) do Local de residência
CD_REGIÃO	Código da Região no estado do Ceará
LOC_OCOR	Código do Local de ocorrência do parto
IDADEMAE	Faixa etária da idade da mãe
ESMAE	Código da Faixa de escolaridade da mãe
QT_FILVIVO	Quantidade de filhos vivos da mãe
QT_FILMORT	Quantidade de Filhos mortos da mãe
GRAVIDEZ	1 – única; 2 – dupla; 3 – tripla e mais
GESTACAO	Código da Faixa de Semanas de Gestação
PARTO	1 – vaginal ou 2 – cesáreo
PESO	Código da Faixa do Peso ao Nascer
ASSISTMED	Houve assistência médica no parto (1 - Sim e 2 - Não)
APGAR1	Índice Apgar com 1 (um) minuto de vida
APGAR5	Índice Apgar com 5 (cinco) minutos de vida
ESTCIVIL	Código do estado civil da mãe
QTD_CONSULTAS	Quantidade de consultas pré-natal até o parto
CLASSIFICAÇÃO	Variável ALVO, informa se ocorreu "Óbito" ou "Não Óbito"

Fonte – Adaptação do autor.

Fase IX: Nesta fase os dados foram divididos em dados para o treinamento e dados para testes de 10 (dez) formas diferentes e mutuamente exclusivos, na proporção de 70% por 30%. Essa técnica utiliza 10 (dez) *folds* em *cross validation* (Validação cruzada), isto é,

misturando os dados de treino e teste em grupos diferentes e cruzando-os, conforme mostra a Figura 32, gerando uma única matriz de pesos sinápticos.

Figura 32 – Técnica de *K-folds* e *Cross Validation*.



Fonte – Adaptação do autor.

A validação cruzada (*cross validation*) é uma ferramenta estatística conhecida por dividir aleatoriamente o conjunto de dados em um conjunto de treinamento e um conjunto de testes. A motivação aqui é validar o modelo com um conjunto de dados diferente daquela usado para definir os parâmetros. O uso da validação cruzada é indicado quando é necessário projetar uma rede neural grande e o objetivo é uma boa generalização. Dessa forma se espera que a rede se torne bem treinada de modo que aprenda o suficiente sobre o passado para generalizar o futuro (HAYKIN, 2001).

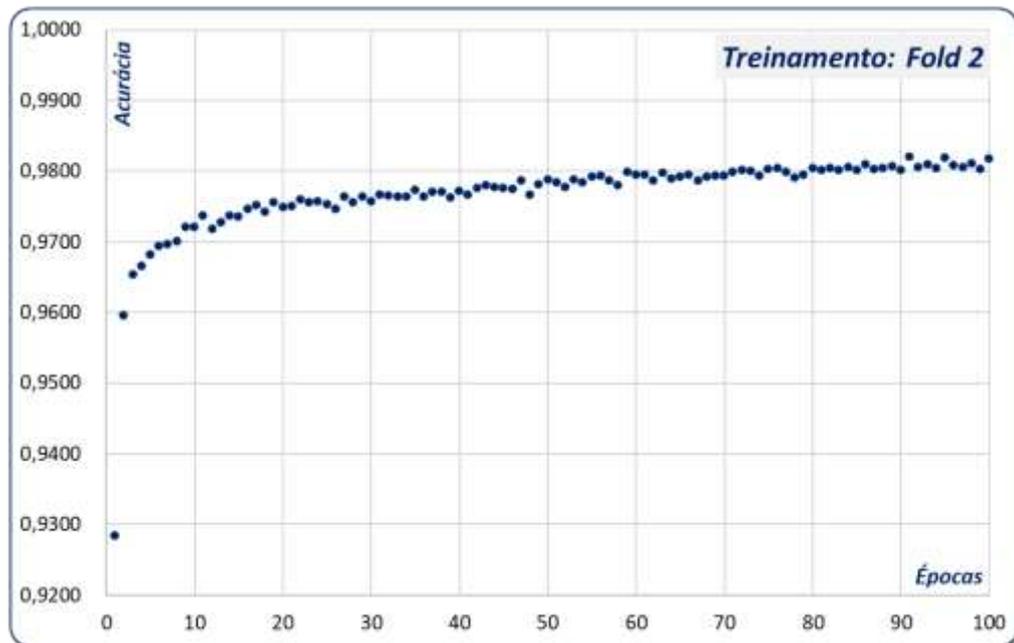
O método de validação cruzada que foi utilizado é chamado de *K-folds*, que consiste em dividir os dados em 10 (dez) subconjuntos mutuamente exclusivos, que de forma circular, seleciona sete partes para treinamento e três partes para os testes. Após *k*-iterações calcula-se a acurácia para obter uma medida mais confiável.

Acurácia é medida a partir da quantidade de objetos ou fatos classificados corretamente pelo algoritmo.

Fase X: Nesta fase iniciamos o processo de treinamento com 100 (cem) épocas, isto é, o modelo de aprendizado foi executado cem vezes para a base de dados de treinamento, e esse processo repetido por 10 (dez) vezes ou *folds* com bases de treinamentos diferentes.

O Gráfico 19 apresenta o treinamento ($K\text{-fold} = 2$) com as informações das acurácias obtidas nas 100 (cem) épocas executadas. Esse treinamento obteve os melhores resultados quando aplicado no *dataset* de teste com acurácia de 98,04%. Nesse gráfico é possível observar uma curva crescente do aprendizado.

Gráfico 19 – Gráfico de Acurácia do melhor treinamento ($K\text{-fold} = 2$).



Fonte – Adaptação do autor.

O Apêndice VIII – DESEMPENHO DOS TREINAMENTOS ($K = 10$), apresenta os gráficos dos dez treinamentos com suas 100 (cem) épocas cada, possibilitando de forma visual comparar o desempenho de cada treinamento.

Fase XI: Nesta fase são feitas as avaliações do modelo. Os dados obtidos na avaliação do modelo usando os dados retidos para testes, isto é, que não foram utilizados no treinamento do modelo. Estes testes apresentaram resultados satisfatórios, com o menor valor de acuraria de 96,90% no sétimo *fold* e o melhor valor de acurácia de 98,04% no segundo *fold*, conforme a Tabela 8.

Tabela 8 – Resultados obtidos em 10 treinamentos com 100 épocas cada.

ORDEM	ERRO MÉDIO	ACURÁCIA MÉDIA	ACURÁCIA DO TESTE
1º <i>fold</i>	6,950%	97,772%	97,61%
2º <i>fold</i>	7,115%	97,673%	98,04% 
3º <i>fold</i>	6,959%	97,700%	97,17%
4º <i>fold</i>	6,869%	97,763%	96,96%
5º <i>fold</i>	6,993%	97,707%	97,99%
6º <i>fold</i>	6,947%	97,769%	97,77%
7º <i>fold</i>	6,875%	97,801%	96,90% 
8º <i>fold</i>	6,988%	97,782%	97,39%
9º <i>fold</i>	6,762%	97,832%	97,01%
10º <i>fold</i>	6,766%	97,804%	97,28%

Fonte – Adaptação do autor.

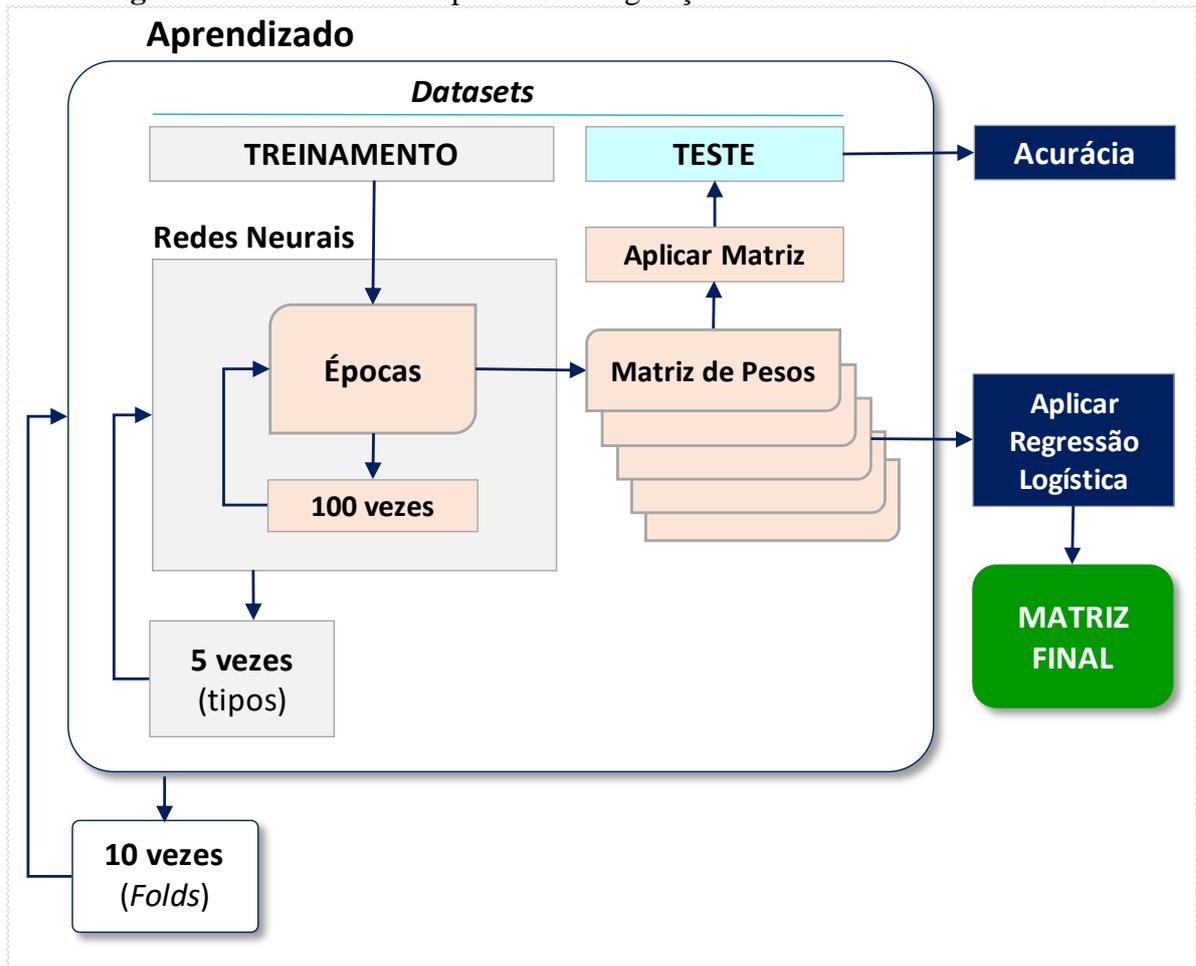
Fase XII: Carrega os 05 (cinco) modelos e os avalia separadamente cada modelo, um para cada Rede do Comitê de Redes Neurais. Nestes testes os resultados obtiveram as acurácias variando de 98,3% a 98,5%, perfazendo 98,6 no teste.

Fase XIII: Geração de um único modelo usando Regressão Logística a partir das cinco matrizes do comitê de redes neurais.

Fase XIV: Depois que todos os modelos foram treinados e testados, os pesos finais dos treinamentos realizados foram armazenados para posterior aplicação no mundo real. Ao final de todos os treinamentos e testes, chegamos ao valor médio da acurácia de 97,41% com erro de mais ou menos 0,40%.

O processo descrito nas fases acima também pode ser analisado e visualizado na Figura 33 que mostra o fluxo do processo de aprendizado utilizado.

Figura 33 – Processo de Aprendizado e geração da Matriz Final.



Fonte – Adaptação do autor.

O diagrama acima resume todo o processamento do modelo de aprendizado, onde cada *dataset* foi dividido em treinamento e teste, executado pelo comitê de cinco redes neurais com épocas cada. Cada matriz gerada é testada com o *dataset* de teste e calculado a sua acurácia. Esse processo se repete 10 *folds* (vezes) com *datasets* diferentes de treinamento e teste. Ao final é aplicada a regressão logística para geração da Matriz Final de pesos a ser usada no mundo real.

O uso contínuo desse modelo e processo requer o reaprendizado constante para que a matriz de pesos esteja sempre atualizada com a realidade mais recente, isto é, se faz necessário a coleta de dados recentes para novos treinamento e testes, gerando novos aprendizados, aproximando a realidade do modelo proposto de aprendizado. O Apêndice IX apresenta um exemplo dos pesos sinápticos obtidos durante a execução dos modelos.

VI – CONCLUSÕES

No percurso para a construção de uma tese uma percepção fica logo evidenciada no início, “Uma tese não se faz só”. Fazer uma tese não é um trabalho individual, é um trabalho em grupo que tem a frente o doutorando, mas que depende dele e de vários colaboradores, como: orientadores, professores, colegas de turma e amigos (pesquisadores ou não), cada um contribuindo de sua forma e em seu tempo, com sua gentil doação dos conhecimentos, conselhos e informações.

Essa é uma pesquisa multidisciplinar que envolveu a computação e a saúde para auxiliar a gestão pública. São três áreas interligadas com um objetivo comum: “Eliminar ou reduzir um problema de saúde pública que afeta o nosso estado e o nosso país desde o Brasil colônia: A Mortalidade Infantil”, e mais uma vez a tese pediu a ajuda de todos, dos profissionais da computação, da saúde e da gestão pública. Um trabalho de pesquisa que envolveu vários profissionais de áreas distintas sob a coordenação e responsabilidade do doutorando e seu orientador.

A estatística descritiva aplicada nessa pesquisa nos permitiu chegar a conclusões importantes, algumas já conhecidas ou imaginadas, mas que o rigor estatístico não deixa dúvidas. Por exemplo:

- A importância das consultas de pré-natal para a redução da mortalidade infantil fica clara quando avaliamos a utilização desse recurso pelas gestantes: em quase 30% dos óbitos infantis as mães só fizeram até três consultas e apenas 68,6% fizeram quatro consultas ou mais contra os 92,6% das mães de nascidos vivos até um ano, isto é, 24% a menos.
- O baixo grau de instrução da população brasileira, também tem reflexos quando analisamos os cuidados com a gravidez e o primeiro ano de vida dos bebês. Analisando os dados encontramos que 50% dos óbitos infantis ocorreram entre as mães de até 25 anos que não concluíram o Ensino Fundamental I, esse número é ainda pior quando filtramos só as mães de até 15 anos, onde a proporção é de 75% que não concluíram o Ensino Fundamental I. Entre as mães de nascidos vivos até um ano a situação é inversa, 75% das mães até 25 anos e 50% das mães até 15 anos concluíram o Ensino Fundamental I, confirmando o peso do grau de instrução nos cuidados e atenção à gravidez e ao primeiro ano das crianças.
- As conclusões acima vão refletir no grande volume de óbitos que ocorrem na primeira semana de vida, isto é, 60% dos óbitos infantis ocorrem nesse breve período, sendo que 35%,

ou seja, mais da metade ocorrem nas primeiras 24 horas de vida, confirmando a falta de cuidados com a gravidez, com a higiene, com o pré-natal, causados pelo baixo grau de instrução e algumas vezes pela falta de acesso à informação e aos cuidados necessários.

Para atingir o objetivo central dessa pesquisa, esse trabalho desenvolveu um método usando algoritmos de IA (Inteligência Artificial) para predição de eventos de mortalidade infantil, embora esse aprendizado tenha permitido a aplicação desse método em outras questões da saúde, segurança e educação. Os resultados obtidos a partir dos milhares de testes e processamentos realizados com muitas horas de máquina (CPU) nos permitiu inferir como de grande valia a aplicação desses modelos na predição da mortalidade infantil, visto que os testes já nos mostraram acurácia média de 97,41% de acertos com margem de erro de +/- 0,41%, isto é, certificando que é viável e valiosa a utilização desses modelos aqui aplicados para predizer o futuro de nossas crianças.

O grande volume de dados em registros e informações das características dos óbitos e dos nascidos vivos possibilitou que os algoritmos atingissem acurácia de alto padrão, considerada “muito boa” e com baixa margem de erro, permitindo assim, a aplicação desses pesos sinápticos no mundo real com grande margem de segurança.

Outra conclusão que preocupa é quando avaliamos a qualidade dos dados dos sistemas SIM e SINASC no Datasus que ainda são insuficientes quanto a qualidade das informações e a garantia da coleta completa das informações. Alguns estudos já mostraram as dificuldades de coletar 100% dos óbitos, bem como alguns óbitos são lançados com informações incompletas prejudicando as análises.

Analisando os dados coletados e as informações estatísticas levantadas, nos permite supor que o acesso à informação pelas mães é fator crucial para redução dos riscos de mortalidade infantil. Esse acesso à informação pode ser através da educação formal, pois é muito baixo o nível de escolaridade das mães que figuram na mortalidade infantil, bem como a falta de acesso ou ausência delas nas consultas pré-natal, fonte principal de informações e acompanhamento da gestação e orientação às mães. O acesso aos serviços de saúde também impacta diretamente nos indicadores da mortalidade infantil, carecendo de ações públicas eficazes para facilitar o acesso às consultas pré-natal, inclusive executando busca ativa às mães faltosas.

VII – PERSPECTIVAS E PROPOSTAS DE NOVOS TRABALHOS

A pesquisa desenvolvida nessa tese não se encerra aqui, pelo contrário, ela abre várias possibilidades de aplicações no mundo real para combater a mortalidade infantil e com possibilidades de aplicações em outras áreas da saúde, segurança e educação.

Os dados da estatística descritiva levantados, independentes da aplicação preditiva da IA, já se apresentam como uma fonte rica em informações para a tomada de decisão do poder público no direcionamento de suas ações preventivas da mortalidade infantil.

No combate à mortalidade infantil o modelo de predição proposto poderá ser aplicado como ferramenta dos profissionais de saúde na detecção e identificação precoce das gestações de baixo, médio e alto risco, possibilitando assim que o poder público se antecipe em suas ações de combate à mortalidade infantil. O Apêndice III – Ferramenta de predição e o Apêndice IV – Simulador para educação na área da Saúde, são propostas de exemplos reais de aplicações utilizando os conhecimentos adquiridos nessa pesquisa a serem aplicados no mundo real com um considerável retorno para a saúde pública da nossa comunidade.

Como já foi descrito na metodologia, os dados utilizados nessa pesquisa são dados públicos disponibilizados pelo DATASUS (Ministério da Saúde). Como proposta de pesquisa futura, pretendo buscar incansavelmente o acesso às bases de dados do CIDACS (Centro de Integração de Dados e Conhecimentos para a Saúde), mantido pela Fundação Oswaldo Cruz – Fiocruz, a Fundação Universidade de Brasília – UnB, a Universidade Federal da Bahia – UFBA e o Ministério do Desenvolvimento Social e Agrário – MDSA em Salvador na Bahia. Eles possuem hoje, a principal base de dados de saúde da população brasileira, com uma coorte de 100 milhões de brasileiros, envolvendo entre outras a pesquisa: “Determinação social e impacto de políticas públicas na saúde materno-infantil”, que muito poderia nos ajudar, mas que até o momento está restrita aos seus pesquisadores internos.

A clusterização é uma técnica que IA que pode ser aplicada também em trabalhos futuros para classificar os grupos de Mortalidade Infantil usando as características dos registros de óbitos. Esta técnica usa os algoritmos do tipo K-Means que agrupa os objetos de características semelhantes e próximas ao centroide do grupo, possibilitando ao poder público conhecer os agrupamentos de casos de óbitos, não pelas características determinísticas, mas pelas características semelhantes descobertas pelos algoritmos e assim tomar as providência direcionadas para cada grupo identificado.

Outro estudo importante a ser feito é o uso do mesmo modelo com os mesmos algoritmos só que usando menos informações (eventos e características), com informações mais fáceis de coletar no mundo real, para avaliar se a acurácia se mantém no alto nível, pois assim, tornaria o modelo mais fácil de se aplicar e obter resultados.

O modelo proposto está disponível para ser usado em qualquer localidade, município, região metropolitana ou estado do Brasil de forma imediata, pois o código usado está disponibilizado no Apêndice V – SCRIPT PYTHON PARA EXECUÇÃO. Dessa forma espero ter contribuído e continuar contribuindo para a saúde pública brasileira. Com essa pesquisa “EU NÃO QUERO PREDIZER ÓBITOS INFANTIS, EU QUERO PODER AJUDAR A EVITÁ-LOS”.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Wanessa da Silva; SZWARCOWALD, Célia L. Adequação das informações de mortalidade e correção dos óbitos informados a partir da Pesquisa de Busca Ativa. *Revista Ciência & Saúde Coletiva*, vol. 22, n. 10. Rio de Janeiro, out/2017.

AMARAL F. Introdução à ciência de dados: Mineração de dados e Bigdata. Rio de Janeiro: Alta Books; 2016.

ARMSTRONG, D. The invention of infant mortality. *Sociology of Health Illness*, 1986.

ASCENCIO, Ana Fernanda G.; ARAÚJO, Graziela S. Estrutura de dados: algoritmos, análise da complexidade e implementações em JAVA e C/C++. São Paulo: Ed. Pearson Prentice Hall, 2010.

ATLAS DO DESENVOLVIMENTO HUMANO NO BRASIL: base de dados. Disponível em: <<http://www.atlasbrasil.org.br>>. Acesso em: 30 de outubro de 2018.

BARBOSA, A. P. Terapia intensiva neonatal e pediátrica no Brasil: o ideal, o real e o possível. *J. Pediatr.*, Rio de Janeiro, v. 80, n. 6, p. 437-438, 2004.

BARR A; FEIGENBAUM E. A. The Handbook of Artificial Intelligence, volume I-II. Willian Kaufmann Inc., Los Altos, California, 1981.

BELFIORE, Patrícia. Estatística aplicada à administração, contabilidade e economia com Excel e SPSS. 1ª ed. Rio de Janeiro: Elsevier, 2015.

BOUZADA, Marco A.; CORSO, Jansen M.; PEIXE, Julinês B.; FARIAS, Carlos A. Métodos quantitativos aplicados a casos reais. 1ª ed. Rio de Janeiro: Elsevier, 2013.

BRAGA, Antônio de P.; CARVALHO, André P. L.; LUDEMIR, Teresa B. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro: Ed. LTC, 2007.

BRAGA L. Introdução à Mineração de Dados. Rio de Janeiro: E-Papers; 2005.

BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. Brasília-DF, Senado, 1988.

BRASIL, Ministério da Saúde. Uma análise da situação de saúde. Brasília; maio, 2004.

BRASIL, Ministério da Saúde. Secretaria de Atenção à Saúde. Manual prático para implementação da Rede Cegonha. Brasília: 2011.

CAMPOS André L. N. A modelagem de processos com BPMN. Rio de Janeiro: ed. Brasport, 2013.

CARVALHO, João Vinicius F.; CHIANN, Chang. Redes Bayesianas: um método para avaliação de interdependência e contágio em séries temporais multivariadas. *Revista Brasileira de Economia*, vol. 67, n. 2, Rio de Janeiro, Abr/Jun 2013.

CARVALHO, Maria Raquel R.; GURGEL, Luilma A.; BRILHANTE, Ana Paula C.; MARTINS, Marley C.; JORGE, Maria Salete B. Manual de Saúde Pública.: Saúde da mulher e rede materno-infantil: a política de atenção e seus desafios. Salvador: SANAR, 2016.

CASS, Mark J. A teoria da prova em Leibniz. *Science e Studia - Revista Latino-Americana de Filosofia e História da Ciência – USP*, São Paulo, v. 11, n. 2, p. 267-279, 2013.

CASTRO, Leandro N.; FERRARI, Daniel G. Introdução à Mineração de dados: conceitos básicos, algoritmos e aplicações. 1ª Ed. São Paulo: Saraiva, 2016.

CAUCHICK, Paulo A.; CAMPOS, Lucila N.; JABBOUR, Charbel J.; JABBOUR, Ana Beatriz L. Elaboração de Artigos Acadêmicos: Estruturas, métodos e técnicas. 1ª Ed. Rio de Janeiro: Ed. Elsevier, 2017.

CEARÁ, Secretaria de Saúde do estado do Ceará. Informe epidemiológico: Mortalidade Materna, 2014.

CEARÁ, Secretaria de Saúde do estado do Ceará. Informe epidemiológico: Mortalidade Materna, 2015.

CERVO, Amado L.; SILVA, Roberto da; BERVIAN, Pedro A. Metodologia Científica. 6ª ed. São Paulo: Ed. Pearson Prentice Hall, 2007.

CHARNIAK, Eugene; McDERMOTT, Drew. Introduction to artificial intelligence. Addison-Wesley Logman Publishing Co., Boston, USA, 1987.

CHILD, Wiliam. Wittgenstein. 1ª ed. Porto Alegre: Penso, 2013.

CRESWELL, John W. Projeto de pesquisa: Métodos qualitativo, quantitativo e misto. 3ª ed. Porto Alegre: Artmed, 2010.

DENG, L.; YU, D. Deep Learning: Methods and applications. NOW Publishers, 2014.

DESCARTES, René. Discurso do Método. 4ª edição, Lisboa: Ed. Babel, 2010.

DIAS, Bárbara A.; NETO, Edson T.; ANDRADE, Mária A. Classificações de evitabilidade dos óbitos infantis: diferentes métodos, diferentes repercussões? Caderno de Saúde Pública, 2017.

DONIEC, Katarzyna; DALL'ALBA, Rafael; KING, Lawrence. Brazil's health catastrophe in the making. *The Lancet*: Elsevier. July, 2018.

DRAELOS, T. J.; MINER, N. E.; LAMB, C. C.; COX, J. A.; VINEYARD, C. M.; CARLSON, K. D.; SEVERA, W. M.; JAMES, C. D.; AIMONE, J. B. Neurogenesis deep learning: Extending deep networks to accommodate new classes. In *Neural Networks (IJCNN)*, 2017 International Joint Conference on, pages 526–533. IEEE, 2017.

ECO, Umberto. Como se faz uma tese. 20ª edição, São Paulo: Perspectiva, 2005.

ELLENBERG, Jordan. O poder do pensamento matemático: a ciência de como não estar errado. 1ª edição. Rio de Janeiro: Editora Zahar, 2015.

FAYYAD U.; PIATETSKY-SHAPIRO G.; SMYTH P. From Data mining to Knowledge Discovery: An Overview, Knowledge Discovery and Data mining, Menlo Park: AAAI Press; 1996.

FEIJOO, AMLC. Objetivos da inferência estatística. In: A pesquisa e a estatística na psicologia e na educação [online]. Rio de Janeiro: Centro Edelstein de Pesquisas Sociais, 2010, pp. 31-38. ISBN: 978-85-7982-048-9. Available from SciELO Books <<http://books.scielo.org>>.

FERREIRA, Emanuella Faheina Chaves. Governança corporativa na saúde suplementar: Estudo de caso em operadora de plano de saúde. Dissertação (Mestrado em Administração de Empresas) – Universidade de Fortaleza – UNIFOR, CMA, Fortaleza, 2011, 93 p.

FLETCHER, Robert H.; FLETCHER, Suzanne W.; FLETCHER, Grant S. Epidemiologia clínica: Elementos essenciais. 5ª edição, Porto Alegre: Artmed, 2014.

FOLHA DE SÃO PAULO: Com Zika e crise no país, mortalidade infantil sobe pela 1ª vez em 26 anos. Edição on line de 16/07/2018, acessada em 16/07/2018. Link: <https://www1.folha.uol.com.br/cotidiano/2018/07/com-zika-e-criseno-pais-mortalidade-infantil-sobe-pela-1a-vez-em-26-anos.shtml>, 2018.

FOREMAN J. Data Smart: Usando Data Science para transformar informação em insight. Rio de Janeiro: Alta Books; 2016.

FOUCAULT, M. Microfísica do poder. 11ª ed. Rio de Janeiro: Graal, 1993.

FURLAN, Marco A.; GOMES, Marcelo M.; SOARES, Marcio V.; CONCILIO, Ricardo. Algoritmos e lógica de programação. 3ª ed. São Paulo: Editora Cengage, 2019.

GIL, Antônio Carlos. Como elaborar projetos de pesquisa. 4ª ed. São Paulo: Editora Atlas, 2007.

GOLDSCHMIDT R, PASSOS E, BEZERRA E. Data Ming: conceitos, técnicas, algoritmos, orientações e aplicações. Rio de Janeiro: Elsevier, 2015.

GONÇALVES, Mara; MORAES, Danielle R. O QUIMO – Política Pública de Saúde - Legislação – SUS. Rio de Janeiro: Editora Águia Dourada, 2007

GURGEL, C. Doenças e curas: o Brasil nos primeiros séculos. São Paulo: Contexto, 2010.

HAN, J.; KAMBER, M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco, CA, 2000.

HAYKIN, Simon. Redes neurais: princípios e prática. 2ª edição. Porto Alegre: Ed. Bookman, 2001.

HESSEN J. Teoria do Conhecimento. 3ª edição. São Paulo: WMF Martins Fontes; 2012.

HEUSER, Carlos Alberto. Projeto de Banco de Dados. 6ª edição. Porto Alegre: Ed. Bookman, 2009.

HUME, David. Tratado da natureza humana: uma tentativa de introduzir o método experimental de raciocínio nos assuntos morais. Tradução de Débora Danowski. 2ª ed. rev. e ampliada. São Paulo: Editora UNESP, 2009.

KIMBALL, Ralph; ROSS, Margy; THORNTHWAITE, Warren; MUNDY, Joy; BACKER, Bob. The Data Warehouse: Lifecycle Toolkit. 2ª ed. Editora John Wiley & Sons, 2008.

KOONTZ, A. M.; KATHLEEN, A. B.; RUDERMAN, M. The Evolution of Fetal and Infant Mortality Review as Public Health Strategy. USA Maternal and Child Health Journal, vol. 8, n. 4. Dec. 2004.

KOVÁCS, Zsolt L. Redes Neurais Artificiais: Fundamentos e Aplicações. 4ª ed. São Paulo: Ed. Livraria da Física, 2006.

KURZWEIL, Raymond. The age of Intelligent Machines. MIT Press, USA, 1990.

LARSON, Ron; FARBER, Betsy. Estatística aplicada. 4ª ed. São Paulo: Pearson Prentice Hall, 2010.

LEAL, Maria do C.; SZWARCOWALD, Célia L.; ALMEIDA, Paulo V.; AQUINO, Estela M.; BARRETO, Maurício L.; BARROS, Fernando, VICTORA, César. Reproductive, maternal, neonatal and child health in the 30 years since the creation of the Unified Health System (SUS). Ciência & Saúde Coletiva, Rio de Janeiro, v. 23, n. 6, 2018.

LEITE, A. J.; CUNHA, A. J.; VICTORIA, C. G. Epidemiologia da saúde da criança. 7ª ed. Rio de Janeiro: Hucitec, 2013.

LIMA, Isaías; PINHEIRO, Carlos A. M.; SANTOS, Flávia A. O. Inteligência Artificial. Rio de Janeiro: Ed. Elsevier, 2014.

LUGER, George F. Inteligência Artificial. 6ª ed. São Paulo: Pearson Education do Brasil, 2013.

MALTA, D.; DUARTE, E.; ALMEIDA, M.; DIAS, M.; MORAIS, O.; MOURA, L.; FERRAZ, W.; SOUZA, M.; Lista de causas de mortes evitáveis por intervenções do Sistema Único de Saúde do Brasil. Epidemiologia Serviços de Saúde. 2007.

MALTA, D.; DUARTE, E.; ESCALANTE, J.; ALMEIDA, M.; SARDINHA, L.; MACÁRIO, E.; MONTEIRO, R.; MORAIS, Neto O. Mortes evitáveis em menores de um ano, Brasil, 1997 a 2006: contribuições para a avaliação de desempenho do Sistema Único de Saúde. Caderno de Saúde Pública, 2010.

MARTIN, James; MCLURE, Carma. Técnicas estruturadas e case. São Paulo: Ed. Makron Books, 1991.

MASCARENHAS, Sidnei A. Metodologia Científica. São Paulo: Pearson Education do Brasil, 2012.

MELO JÚNIOR, Cleuton Sampaio de. Data Science para programadores – Um guia completo utilizando a linguagem Python. Rio de Janeiro: Editora Ciência Moderna, 2018.

MINAYO, Maria Cecília de Souza. O desafio do conhecimento. 6. ed. São Paulo: HUCITEC, 2000. 269 p.

MITCHEL, T. M. Machine Learning. New York: Ed. McGraw-Hill Science, 1997.

MONTE, Alana S.; MENDES, Igor C.; ORÍÁ, Mônica O. B.; CARVALHO, Francisco Herlânio C.; BROWN, Helen; DAMASCENO, Ana Kelve de C. Near miss materno: fatores influenciadores e direcionamentos para redução da morbidade e mortalidade materna. Revista Rene - UFC;19:e3182, 2018.

MORSE, Marcia Lait et. al. Mortalidade materna no Brasil: o que mostra a produção científica nos últimos 30 anos? Cad. Saúde Pública, Rio de Janeiro, v. 27, n. 4, Apr. 2011.

MOTA, Hugo. "O que é IDH?"; Brasil Escola. Disponível em <<https://brasilecola.uol.com.br/o-que-e/geografia/o-que-e-idh.htm>>. Acesso em 30 de outubro de 2018.

MS, Ministério da Saúde. Inquérito domiciliar sobre comportamento de risco e morbidade referida de doenças e agravos não transmissíveis. Rio de Janeiro, 2004.

MS, Ministério da Saúde. Nota técnica da secretaria de Vigilância em Saúde: Vigilância epidemiológica de óbitos infantis e fetais. Brasília, 2009.

NUNES, Everardo D. Saúde Coletiva: Uma história recente de um passado remoto. Tratado de Saúde Coletiva. 2ª ed. São Paulo: Hucitec, 2012.

OLSEN, Wendy. Coleta de dados: debates e métodos fundamentais em pesquisa social. Porto Alegre: Ed. Penso, 2015.

OMS – Organização Mundial da Saúde: USP – Universidade de São Paulo, Centro colaborador da OMS para classificação de doenças em português. Classificação estatística internacional de doenças e problemas relacionados à saúde. 10ª revisão, São Paulo: Universidade de São Paulo, 1993.

OPAS. Organização Mundial da Saúde. 47º Conselho Diretor da Organização Pan-Americana da Saúde. Saúde neonatal no contexto da saúde materna, neonatal e da criança para o cumprimento das metas de desenvolvimento do milênio da declaração do milênio das Nações Unidas, Washington, 2006. 28p.

PAIM, J. S. Políticas de saúde no Brasil: Epidemiologia e saúde. 6ª edição. Editora MEDSI/Guanabara Koogan, 2003.

PIMENTEL, Pablo F. C. O raciocínio indutivo em David Hume e Thomas Reis. *Revista de Filosofia Cognitio*, São Paulo, v. 18, n. 2, p. 243-258, jul/dez 2017.

PLLANA S.; JANCIAK I.; BREZANY P.; WOHRER A. A survey of the state of art in Data Mining and Integration Query Languages. 14th International Conference on Network-Based Information Systems NBiS, 2011.

PMI – PROJECT MANAGEMENT INSTITUTE, Guia PMBOK: Um guia do conjunto de conhecimentos em gerenciamento de projetos. 5ª edição, Pensilvânia-USA, 2012, 616 p.

PORTO, Otávio Maia. A Análise do óbito infantil evitável no Brasil: estado da arte. Dissertação de Mestrado em Saúde Coletiva – Universidade Federal Fluminense, Faculdade de Medicina. Niterói: 2012. 93 p.

RASINGHANI, Maheshi S. Business Intelligence in the Digital Economy: Opportunities, Limitations and Risks. Idea Group Pub, 2004.

RASELLA, D.; BASU, S.; HONE, T.; OCKÉ-REIS, C. O.; MILLET, C. Child morbidity and mortality associated with alternative policy responses to the economic crisis in Brazil: a nationwide microsimulation study. *PLOS Medicine* 15(5): e1002570, 2018.

RIBEIRO, V. S.; SILVA, A. A. M. Tendências da mortalidade neonatal em São Luís, Maranhão, Brasil, de 1979 a 1996. *Cad. Saúde Pública*, Rio de Janeiro, v. 16, n. 2, p. 429-438, 2000.

RIBEIRO, Adolfo M.; GUIMARÃES, Maria J.; LIMA, Marília C.; SARINHO, Sílvia W.; COUTINHO, Sonia B. Fatores de Risco para a mortalidade neonatal em crianças com baixo peso ao nascer. *Revista de Saúde Pública*, vol. 42, n. 2. São Paulo, abr/2009.

RICH, Elaine; KNIGHT, Kavin. Artificial Intelligence. McGraw-Hill Higher Education, USA, 1990.

REALE, Giovanni; ANTISERI, Dario. História da Filosofia: Filosofia pagã antiga. Volume I, São Paulo: Ed. Paulus, 2003.

ROUQUAYROL, Maria Z. *Epidemiologia & Saúde*, 3ª ed. Rio de Janeiro: Editora Médica & Científica, 1989.

ROUQUAYROL, Maria Z. *Contribuição da Epidemiologia. Tratado de Saúde Coletiva*. 2ª ed. São Paulo: Hucitec, 2012.

RUSSEL, Stuart J.; NORVIG, Peter. *Inteligência Artificial*. 3ª ed. Rio de Janeiro: Ed. Elsevier, 2013

RUSSELL, Bertrand. *Fundamentos da Filosofia*. 1ª ed. Barcelona (Espanha): Debolsillo, 2009a.

RUSSELL, Bertrand. *Religião e Ciência*. 1ª ed. Ribeirão Preto: FUNPEC, 2009b.

RUSSELL, Bertrand. Os Problemas da Filosofia. 1ª ed. Lisboa (Portugal): Edições70, 2016.

SANTANA, Danielly S.; GUIDA, José Paulo S.; PACAGNELLA, Rodolfo C.; CECATTI, José G. Near miss materno - entendendo e aplicando o conceito. Rev Med: mar-abr; 97(2): 187-94, São Paulo, 2018.

SAMPAIO, José J. C.; FILHO, Carlos G.; FARIAS, Mariana R.; GOMES, Valéria B. Manual de Saúde Pública.: História das Políticas Públicas de Saúde no Brasil e No Mundo. Salvador: SANAR, 2016.

SAMPIERI, Roberto H.; CALLADO, Carlos F.; LUCIO, Maria P. Metodologia de Pesquisa. 5ª ed. Porto Alegre: Penso, 2013.

SÃO PAULO (cidade). Secretaria Municipal da Saúde. Coordenação de Epidemiologia e Informação – CEInfo. Declaração de Nascido Vivo. Manual de preenchimento da Declaração de Nascido Vivo. São Paulo: Secretaria Municipal da Saúde, 2011. 24p.

SCHRAMM, J. M. de A.; SZWARCOWALD, C. L. Diferenciais nas taxas de mortalidade neonatal e natimortalidade hospitalares no Brasil: um estudo com base no Sistema de Informações Hospitalares do Sistema único de Saúde (SIH/SUS). Cad. Saúde Pública, Rio de Janeiro, v. 16, n. 4, p. 1031-1040, 2000.

SIEGERIST, H. E. Civilização e Doença. São Paulo: Hucitec, 2011.

SILVA L.; PERES S.; BOSCARIOLI C. Introdução à Mineração de dados: Com aplicações em R. 1ª edição. Rio de Janeiro: Elsevier; 2016.

SILVA, Ana Valeska S.; OLIVEIRA, Rejane M. C.; MOTA, Zélia G.; CHAVES, Edna Maria mortalidade infantil na atenção à saúde da criança e da mulher. Salvador: SANAR, 2016.

SILVA, Vinicius Tolentino O.; ZAIDAN, Fernando H. CASE STUDY BUSINESS INTELLIGENCE: CURRICULUM DATA EXTRACTION LATTES FOR THE INSTITUTION OF HIGHER EDUCATION. 12th INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS & TECHNOLOGY MANAGEMENT – CONTECSI. São Paulo-SP, 2015.

SIMÕES, C.; MONTEIRO, C. Tendência secular e diferenças regionais da mortalidade infantil no Brasil. São Paulo: Hucitec, 1995.

SOUSA, M. H.; CECATTI, J. G.; HARDY, E. E.; SERRUYA, S. J. Severe maternal morbidity (near miss) as a sentinel event of maternal death. An attempt to use routine data for surveillance. Reprod Health 2008; 28:5-6.

SZWARCFITER, Jayme L. Teoria computacional dos Grafos. 1ª edição. Rio de Janeiro: Elsevier, 2018.

TAN, P. N.; STEINBACH, M.; KUMAR, V. Introdução ao Data Mining. 1ª ed. Rio de Janeiro: Ed. Ciência Moderna, 2009.

TURBAN, Efraim; LEIDNER Dorothy.; MCLEAN Ephraim.; WETHERBE James. et al. Tecnologia da informação para a gestão, 6ª ed. Porto Alegre: Ed. Bookman, 2010.

TURBAN, Efraim; VOLONINO, Linda. Tecnologia da informação para a gestão: em busca do melhor desempenho estratégico e operacional. 8ª ed. Porto Alegre: Ed. Bookman, 2013.

UK, HOUSE OF LORDS. AI in the UK: ready, willing and able? Select Committee on Artificial Intelligence - Report of Session 2017–19. April, 2018.

VIELLAS, Elaine F.; DOMINGUES, Rosa Maria S.M.; DIAS, Marcos A. B.; GAMA, Silvana G. N; FILHA, Mariza M. T.; COSTA, Janaina V.; BASTOS, Maria H.; LEAL, Maria do C. Assistência Pré-natal no Brasil. Cad. Saúde Pública, Rio de Janeiro, 30 Sup:S85-S100, 2014.

WALLER, M. A.; FAWCETT, S. E. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. Journal of Business Logistics, 34(2), 77–84. 2013.

WEISS S, INDURKHYA N. Predictive Data mining: A Practical guide. São Francisco: Morgan Kaufmann Publishers; 1998.

WHEELAN, Charles. Estatística: o que é, para que serve e como funciona. 1ª edição. Rio de Janeiro: Editora Zahar, 2016.

ZAKI M. Parallel and Distributed Data Mining: An Introduction. Large-Scale Parallel Data Mining. Springer-Verlag Berlin Heidelberg; 2000.

ZHOU L.; PAN S.; WANG J.; VASILAKOS A. Machine Learning on Big Data – Opportunities and Challenges. Science Direct, Vol. 237, P. 350-361; 2017.

ANEXO I – DECLARAÇÃO DE NASCIDO VIVO.

 República Federativa do Brasil Ministério da Saúde 2ª VIA - CARTÓRIO		Declaração de Nascido Vivo 30-52287570-1				
I	Cartório	11 Cartório	Código	12 Registro	13 Data	
		4 Município			5 UF	
II	Local da Ocorrência	6 Local da Ocorrência		7 Estabelecimento		
		1 <input type="checkbox"/> Hospital 2 <input type="checkbox"/> Outros Estab. Saúde 3 <input type="checkbox"/> Domicílio 4 <input type="checkbox"/> Outros 9 <input type="checkbox"/> Ignorado		Código		
		8 Endereço da ocorrência, se fora do estab. ou da resid. da mãe (Rua, praça, avenida, etc)		Número	Complemento	9 CEP
		10 Bairro/Distrito	Código	11 Município de ocorrência	Código	
					12 UF	
III	Mãe	13 Nome da Mãe		14 Cartão SUS		
		15 Idade (anos)	16 Estado Civil	17 Escolaridade (Em anos de estudo concluídos)	18 Ocupação habitual e ramo de atividade	19 Núm. de filhos tidos em gestações anteriores (obs.: utilizar 99 se ignorados)
		1 <input type="checkbox"/> Solteira 2 <input type="checkbox"/> Casada 3 <input type="checkbox"/> Viúva 4 <input type="checkbox"/> Separada judicialmente/divorciada 9 <input type="checkbox"/> Ignorado	1 <input type="checkbox"/> Nenhuma 2 <input type="checkbox"/> De 1 a 3 3 <input type="checkbox"/> De 4 a 7 4 <input type="checkbox"/> De 8 a 11 5 <input type="checkbox"/> 12 e mais 9 <input type="checkbox"/> Ignorado	Código	Nascidos vivos Nascidos mortos	
		20 Residência da mãe		Número	Complemento	21 CEP
		22 Bairro/Distrito	Código	23 Município	Código	
					24 UF	
IV	Gestação e Parto	25 Duração da gestação (em semanas)		26 Tipo de gravidez		
		1 <input type="checkbox"/> Menos de 22 2 <input type="checkbox"/> De 22 a 27 3 <input type="checkbox"/> De 28 a 31 4 <input type="checkbox"/> De 32 a 36 5 <input type="checkbox"/> De 37 a 41 6 <input type="checkbox"/> 42 e mais 9 <input type="checkbox"/> Ignorado	1 <input type="checkbox"/> Única 2 <input type="checkbox"/> Dupla 3 <input type="checkbox"/> Tripla e mais 9 <input type="checkbox"/> Ignorado	27 Tipo de parto		
		28 Nascimento		29 Sexo		
		Data		M - Masculino F - Feminino 1 - Ignorado		
V	Recém-Nascido	30 Raça/cor		31 Índice de Apgar		
		1 <input type="checkbox"/> Branca 2 <input type="checkbox"/> Preta 3 <input type="checkbox"/> Amarela 4 <input type="checkbox"/> Parda 5 <input type="checkbox"/> Indígena	32 Peso ao nascer		1º minuto 5º minuto	
		33 Detectada alguma malformação congênita e/ou anomalia cromossômica?		Código		
		1 <input type="checkbox"/> Sim 2 <input type="checkbox"/> Não 9 <input type="checkbox"/> Ignorado Qual?				
VI	Identificação	34 Polegar direito da mãe		35 Pé direito da criança		
VII	Preench.	36 Responsável pelo preenchimento		37 Função		
		Nome		Identidade		
		38 Órgão Emissor		39 Data		
ATENÇÃO: ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO O Registro de Nascimento é obrigatório por lei. Para registrar esta criança, o pai ou responsável deverá levar este documento ao cartório de registro civil.						
Versão 12/08 - 1ª Impressão 12/2008						

ANEXO II – DECLARAÇÃO DE ÓBITO.

República Federativa do Brasil
Ministério da Saúde
1ª VIA - SECRETARIA DE SAÚDE

Declaração de Óbito

I	Cidade	1) Cidade		Código		2) Registro	3) Data				
		4) Município		5) UF	6) Cidades						
II	Identificação	7) Tipo de Óbito		8) Óbito Data		9) IBC		10) Nacionalidade			
		11) Nome do falecido									
		12) Nome do pai				13) Nome da mãe					
		14) Data de nascimento		15) Idade		16) Sexo		17) Raça/cor		18) Ocupação habitual e ramo de atividade	
III	Residência	19) Estado civil		20) Escolaridade		21) Ocupação habitual e ramo de atividade					
		22) Logradouro (Rua, praça, avenida etc.)		Código		Número		Complemento			
		23) Bairro/Distrito		Código		24) Município de residência		Código		25) UF	
IV	Ocorrência	26) Local de ocorrência do óbito		27) Estabelecimento		28) Endereço da ocorrência, se fora do estabelecimento ou da residência					
		29) Bairro/Distrito		Código		30) Município de ocorrência		Código		31) UF	
		32) Salvo/Diário		Código		33) Município de ocorrência		Código		34) UF	
V	Fórmula menor que 1 ano	PREENCHIMENTO EXCLUSIVO PARA ÓBITOS FETAIS E DE MENORES DE 1 ANO									
		35) Idade		36) Escolaridade		37) Ocupação habitual e ramo de atividade da mãe		38) Número de filhos vivos			
		39) Direção da gestação		40) Tipo de Gravidez		41) Tipo de parto		42) Morte em relação ao parto			
		43) Tempo de gestação		44) Tipo de parto		45) Tempo de parto		46) Tempo de parto			
VI	Condições e causas do óbito	47) A morte ocorreu durante a gravidez, parto ou parto?		48) A morte ocorreu durante o parto?		49) ASSISTÊNCIA MÉDICA					
		50) Diagnóstico confirmado por		51) Cirurgia?		52) Necropsia?					
		CAUSAS DA MORTE									
		PARTE I									
VII	Médico	53) Nome do médico		54) CRM		55) O médico que assina atendeu ao falecido?					
		56) Meio de contato (Telefone, fax, e-mail etc.)		57) Data do assento		58) Assinatura					
		PROVÁVEIS CIRCUNSTÂNCIAS DE MORTE NÃO NATURAL, informações de caráter estritamente epidemiológico:									
VIII	Causas externas	59) Descrição sumária do evento, incluindo o tipo de local de ocorrência		60) Descrição sumária do evento, incluindo o tipo de local de ocorrência		61) Fonte de informação					
		62) Logradouro (Rua, praça, avenida etc.)		Código		63) Localidade e Município					
IX	Localidade e Município	64) Localidade		65) Município							

Versão 2008-01

ANEXO III – RELAÇÃO DOS MUNICÍPIOS DO ESTADO DO CEARÁ

UF	CD_MUN	NOME MUNICÍPIO	POPULAÇÃO	MACRO_REGIÃO	MICRO_REGIÃO
CE	230010	Abaiara	11,605	Sul	Brejo Santo
CE	230015	Acarape	16,543	Norte	Baturité
CE	230020	Acaraú	62,199	Noroeste	Litoral de Camocim e Acaraú
CE	230030	Acopiara	53,572	Sertões	Sertão de Senador Pompeu
CE	230040	Aiuaba	17,194	Sertões	Sertão de Inhamuns
CE	230050	Alcântaras	11,459	Noroeste	Meruoca
CE	230060	Altaneira	7,479	Sul	Caririaçu
CE	230070	Alto Santo	16,976	Jaguaribe	Baixo Jaguaribe
CE	230075	Amontada	42,901	Norte	Itapipoca
CE	230080	Antonina do Norte	7,278	Centro-Sul	Iguatu
CE	230090	Apuiarés	14,719	Norte	Médio Curu
CE	230100	Aquiraz	79,128	RMF	Fortaleza
CE	230110	Aracati	73,629	Jaguaribe	Litoral de Aracati
CE	230120	Aracoiaba	26,269	Norte	Baturité
CE	230125	Ararendá	10,823	Sertões	Sertão de Crateús
CE	230130	Araripe	21,398	Sul	Chapada do Araripe
CE	230140	Aratuba	11,244	Norte	Baturité
CE	230150	Arneiroz	7,777	Sertões	Sertão de Inhamuns
CE	230160	Assaré	23,254	Sul	Chapada do Araripe
CE	230170	Aurora	24,496	Sul	Barro
CE	230180	Baixio	6,228	Centro-Sul	Iguatu
CE	230185	Banabuiú	18,027	Sertões	Sertão de Quixeramobim
CE	230190	Barbalha	59,811	Sul	Cariiri
CE	230195	Barreira	20,978	Norte	Chorozinho
CE	230200	Barro	22,440	Sul	Barro
CE	230205	Barroquinha	14,880	Noroeste	Litoral de Camocim e Acaraú
CE	230210	Baturité	35,351	Norte	Baturité
CE	230220	Beberibe	53,110	Norte	Cascavel
CE	230230	Bela Cruz	32,378	Noroeste	Litoral de Camocim e Acaraú
CE	230240	Boa Viagem	54,049	Sertões	Sertão de Quixeramobim
CE	230250	Brejo Santo	48,830	Sul	Brejo Santo
CE	230260	Camocim	62,985	Noroeste	Litoral de Camocim e Acaraú
CE	230270	Campos Sales	27,209	Sul	Chapada do Araripe
CE	230280	Canindé	77,514	Norte	Canindé
CE	230290	Capistrano	17,668	Norte	Baturité
CE	230300	Caridade	22,320	Norte	Canindé
CE	230310	Cariré	18,660	Noroeste	Sobral
CE	230320	Caririaçu	26,892	Sul	Caririaçu
CE	230330	Cariús	18,804	Centro-Sul	Iguatu
CE	230340	Carnaubal	17,631	Noroeste	Ibiapaba
CE	230350	Cascavel	71,079	Norte	Cascavel
CE	230360	Catarina	20,451	Sertões	Sertão de Inhamuns
CE	230365	Catunda	10,365	Noroeste	Santa Quitéria
CE	230370	Caucaia	362,223	RMF	Fortaleza
CE	230380	Cedro	25,063	Centro-Sul	Iguatu
CE	230390	Chaval	12,952	Noroeste	Litoral de Camocim e Acaraú
CE	230393	Choró	13,384	Sertões	Sertão de Quixeramobim
CE	230395	Chorozinho	19,197	Norte	Chorozinho
CE	230400	Coreaú	23,107	Noroeste	Coreaú
CE	230410	Crateús	74,426	Sertões	Sertão de Crateús

ANEXO III (Continuação).

UF	CD_MUN	NOME MUNICÍPIO	POPULAÇÃO	MACRO_REGIÃO	MICRO_REGIÃO
CE	230420	Crato	130,604	Sul	Cariri
CE	230423	Croatá	17,874	Noroeste	Ibiapaba
CE	230425	Cruz	23,983	Noroeste	Litoral de Camocim e Acaraú
CE	230426	Deputado Irapuan Pinheiro	9,521	Sertões	Sertão de Senador Pompeu
CE	230427	Ererê	7,163	Jaguaribe	Serra do Pereiro
CE	230428	Eusébio	52,667	RMF	Fortaleza
CE	230430	Farias Brito	18,720	Sul	Caririáçu
CE	230435	Forquilha	24,047	Noroeste	Sobral
CE	230440	Fortaleza	2,627,482	RMF	Fortaleza
CE	230445	Fortim	16,272	Jaguaribe	Litoral de Aracati
CE	230450	Frecheirinha	13,669	Noroeste	Coreaú
CE	230460	General Sampaio	6,922	Norte	Médio Curu
CE	230465	Graça	15,307	Noroeste	Sobral
CE	230470	Granja	54,365	Noroeste	Litoral de Camocim e Acaraú
CE	230480	Granjeiro	4,425	Sul	Caririáçu
CE	230490	Groaíras	11,012	Noroeste	Sobral
CE	230495	Guaiúba	26,331	RMF	Fortaleza
CE	230500	Guaraciaba do Norte	39,445	Noroeste	Ibiapaba
CE	230510	Guaramiranga	3,547	Norte	Baturité
CE	230520	Hidrolândia	20,215	Noroeste	Santa Quitéria
CE	230523	Horizonte	65,928	RMF	Pacajus
CE	230526	Ibaretama	13,218	Sertões	Sertão de Quixeramobim
CE	230530	Ibiapina	24,825	Noroeste	Ibiapaba
CE	230533	Ibicuitinga	12,350	Jaguaribe	Baixo Jaguaribe
CE	230535	Icapuí	19,685	Jaguaribe	Litoral de Aracati
CE	230540	Icó	67,486	Centro-Sul	Iguatu
CE	230550	Iguatu	102,614	Centro-Sul	Iguatu
CE	230560	Independência	25,967	Sertões	Sertão de Crateús
CE	230565	Ipaporanga	11,499	Sertões	Sertão de Crateús
CE	230570	Ipaumirim	12,349	Centro-Sul	Iguatu
CE	230580	Ipu	41,576	Noroeste	Ipu
CE	230590	Ipueiras	37,896	Noroeste	Ipu
CE	230600	Iracema	14,125	Jaguaribe	Serra do Pereiro
CE	230610	Irauçuba	23,858	Noroeste	Sobral
CE	230620	Itaiçaba	7,738	Jaguaribe	Litoral de Aracati
CE	230625	Itaitinga	39,310	RMF	Fortaleza
CE	230630	Itapajé	51,945	Norte	Uruburetama
CE	230640	Itapipoca	127,465	Norte	Itapipoca
CE	230650	Itapiúna	20,014	Norte	Baturité
CE	230655	Itarema	41,230	Noroeste	Litoral de Camocim e Acaraú
CE	230660	Itatira	20,675	Norte	Canindé
CE	230670	Jagaretama	17,958	Jaguaribe	Médio Jaguaribe
CE	230680	Jaguaribara	11,295	Jaguaribe	Médio Jaguaribe
CE	230690	Jaguaribe	34,448	Jaguaribe	Médio Jaguaribe
CE	230700	Jaguaruana	33,740	Jaguaribe	Baixo Jaguaribe
CE	230710	Jardim	27,076	Sul	Cariri
CE	230720	Jati	7,847	Sul	Brejo Santo
CE	230725	Jijoca de Jericoacoara	19,510	Noroeste	Litoral de Camocim e Acaraú
CE	230730	Juazeiro do Norte	270,383	Sul	Cariri
CE	230740	Jucás	24,597	Centro-Sul	Iguatu

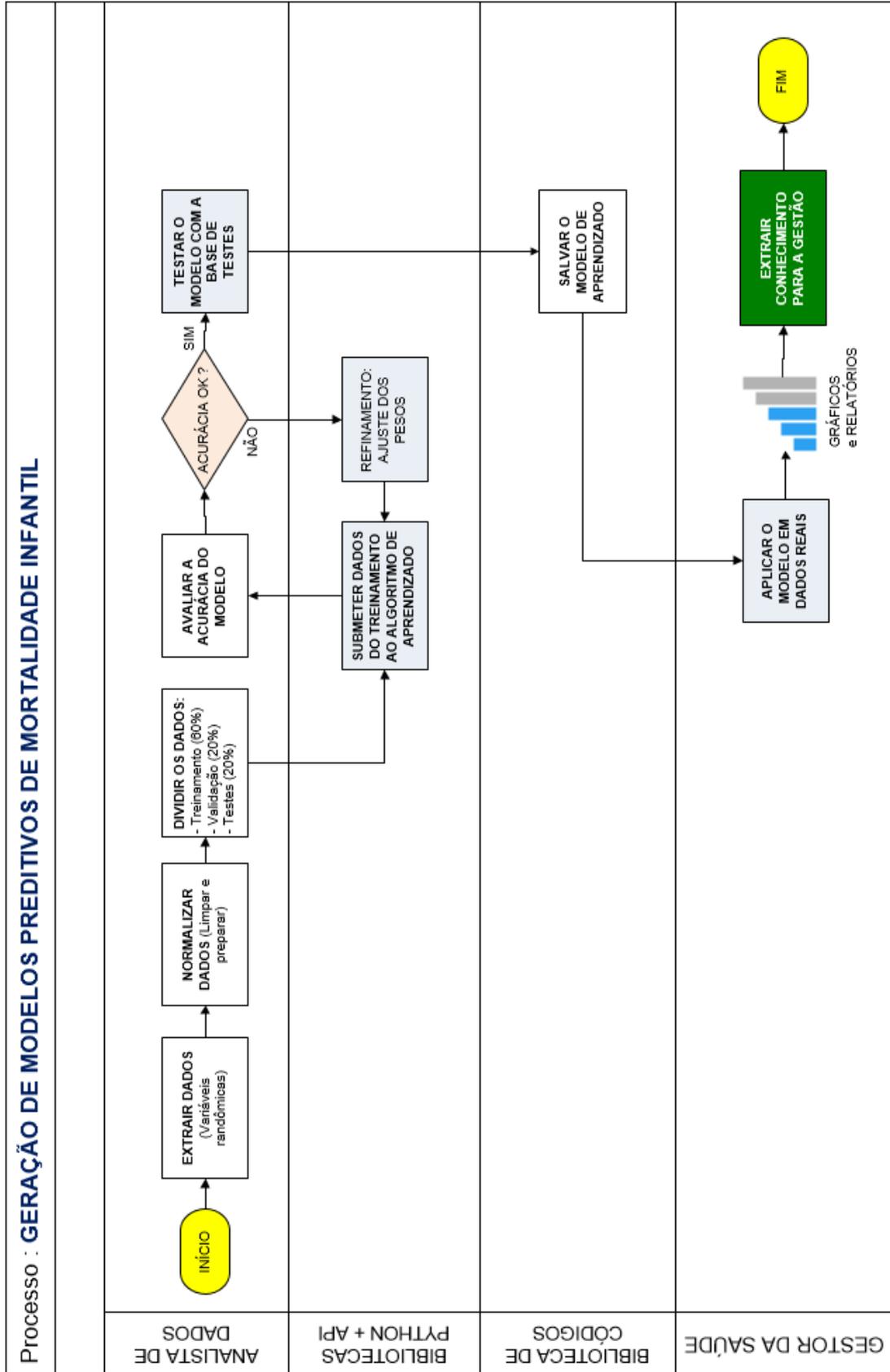
ANEXO III (Continuação).

UF	CD_MUN	NOME MUNICÍPIO	POPULAÇÃO	MACRO_REGIÃO	MICRO_REGIÃO
CE	230750	Lavras da Mangabeira	31,335	Centro-Sul	Iguatu
CE	230760	Limoeiro do Norte	58,915	Jaguaribe	Baixo Jaguaribe
CE	230763	Madalena	19,800	Sertões	Sertão de Quixeramobim
CE	230765	Maracanaú	224,804	RMF	Fortaleza
CE	230770	Maranguape	126,486	RMF	Fortaleza
CE	230780	Marco	26,981	Noroeste	Litoral de Camocim e Acaraú
CE	230790	Martinópolis	11,082	Noroeste	Litoral de Camocim e Acaraú
CE	230800	Massapê	38,210	Noroeste	Sobral
CE	230810	Mauriti	46,548	Sul	Barro
CE	230820	Meruoca	14,948	Noroeste	Meruoca
CE	230830	Milagres	28,231	Sul	Brejo Santo
CE	230835	Milhã	13,136	Sertões	Sertão de Senador Pompeu
CE	230837	Miraíma	13,583	Noroeste	Sobral
CE	230840	Missão Velha	35,409	Sul	Cariri
CE	230850	Mombaça	43,735	Sertões	Sertão de Senador Pompeu
CE	230860	Monsenhor Tabosa	17,038	Sertões	Sertão de Crateús
CE	230870	Morada Nova	61,548	Jaguaribe	Baixo Jaguaribe
CE	230880	Moraújo	8,636	Noroeste	Coreaú
CE	230890	Morrinhos	22,222	Noroeste	Litoral de Camocim e Acaraú
CE	230900	Mucambo	14,377	Noroeste	Sobral
CE	230910	Mulungu	12,831	Norte	Baturité
CE	230920	Nova Olinda	15,433	Sul	Cariri
CE	230930	Nova Russas	32,035	Sertões	Sertão de Crateús
CE	230940	Novo Oriente	28,353	Sertões	Sertão de Crateús
CE	230945	Ocara	25,394	Norte	Chorozinho
CE	230950	Orós	21,292	Centro-Sul	Iguatu
CE	230960	Pacajus	70,911	RMF	Pacajus
CE	230970	Pacatuba	82,824	RMF	Fortaleza
CE	230980	Pacoti	11,960	Norte	Baturité
CE	230990	Pacujá	6,202	Noroeste	Sobral
CE	231000	Palhano	9,285	Jaguaribe	Baixo Jaguaribe
CE	231010	Palmácia	13,145	Norte	Baturité
CE	231020	Paracuru	33,894	Norte	Baixo Curu
CE	231025	Paraipaba	32,515	Norte	Baixo Curu
CE	231030	Parambu	31,137	Sertões	Sertão de Inhamuns
CE	231040	Paramoti	11,578	Norte	Canindé
CE	231050	Pedra Branca	42,841	Sertões	Sertão de Senador Pompeu
CE	231060	Penaforte	8,956	Sul	Brejo Santo
CE	231070	Pentecoste	37,077	Norte	Médio Curu
CE	231080	Pereiro	16,163	Jaguaribe	Serra do Pereiro
CE	231085	Pindoretama	20,644	Norte	Cascavel
CE	231090	Piquet Carneiro	16,731	Sertões	Sertão de Senador Pompeu
CE	231095	Pires Ferreira	10,784	Noroeste	Ipu
CE	231100	Poranga	12,243	Noroeste	Ipu
CE	231110	Porteiras	14,921	Sul	Cariri
CE	231120	Potengi	10,918	Sul	Chapada do Araripe
CE	231123	Potiretama	6,356	Jaguaribe	Serra do Pereiro
CE	231126	Quiterianópolis	20,860	Sertões	Sertão de Crateús
CE	231130	Quixadá	86,605	Sertões	Sertão de Quixeramobim
CE	231135	Quixelô	14,860	Centro-Sul	Iguatu

ANEXO III (Continuação).

UF	CD_MUN	NOME MUNICÍPIO	POPULAÇÃO	MACRO_REGIÃO	MICRO_REGIÃO
CE	231140	Quixeramobim	78,658	Sertões	Sertão de Quixeramobim
CE	231150	Quixeré	21,876	Jaguaribe	Baixo Jaguaribe
CE	231160	Redenção	27,441	Norte	Baturité
CE	231170	Reriutaba	18,769	Noroeste	Ipu
CE	231180	Russas	76,475	Jaguaribe	Baixo Jaguaribe
CE	231190	Saboeiro	15,678	Sertões	Sertão de Inhamuns
CE	231195	Salitre	16,331	Sul	Chapada do Araripe
CE	231200	Santana do Acaraú	32,023	Noroeste	Sobral
CE	231210	Santana do Cariri	17,489	Sul	Cariri
CE	231220	Santa Quitéria	43,360	Noroeste	Santa Quitéria
CE	231230	São Benedito	46,648	Noroeste	Ibiapaba
CE	231240	São Gonçalo do Amarante	48,265	Norte	Baixo Curu
CE	231250	São João do Jaguaribe	7,621	Jaguaribe	Baixo Jaguaribe
CE	231260	São Luís do Curu	12,849	Norte	Médio Curu
CE	231270	Senador Pompeu	26,447	Sertões	Sertão de Senador Pompeu
CE	231280	Senador Sá	7,513	Noroeste	Sobral
CE	231290	Sobral	205,529	Noroeste	Sobral
CE	231300	Solonópole	18,158	Sertões	Sertão de Senador Pompeu
CE	231310	Tabuleiro do Norte	30,489	Jaguaribe	Baixo Jaguaribe
CE	231320	Tamboril	25,525	Sertões	Sertão de Crateús
CE	231325	Tarrafas	8,852	Centro-Sul	Iguatu
CE	231330	Tauá	58,119	Sertões	Sertão de Inhamuns
CE	231335	Tejuçuoca	18,902	Norte	Médio Curu
CE	231340	Tianguá	74,719	Noroeste	Ibiapaba
CE	231350	Trairi	55,207	Norte	Itapipoca
CE	231355	Tururu	15,935	Norte	Uruburetama
CE	231360	Ubajara	34,332	Noroeste	Ibiapaba
CE	231370	Umari	7,671	Centro-Sul	Iguatu
CE	231375	Umirim	19,679	Norte	Uruburetama
CE	231380	Uruburetama	21,609	Norte	Uruburetama
CE	231390	Uruoca	13,677	Noroeste	Coreaú
CE	231395	Varjota	18,239	Noroeste	Ipu
CE	231400	Várzea Alegre	40,440	Sul	Várzea Alegre
CE	231410	Viçosa do Ceará	60,030	Noroeste	Ibiapaba

APÊNDICE I – PROCESSO PARA GERAÇÃO DE MODELOS PREDITOS



APÊNDICE II – DICIONÁRIO DE DADOS DO ARQUIVO DE ÓBITOS INFANTIS.

NM_CAMPO	TAMANHO	DESCRIÇÃO
TIPOBITO	1	1 - FETAL ; 2 - NÃO FETAL
DTOBITO	8	DDMMAAAA
HORAOBITO	4	HHMM
CODMUNNATU	7	CÓDIGO DO MUNICÍPIO DE NATURALIDADE DO FALECIDO
DTNASC	8	DDMMAAAA
IDADE	3	Idade: composto de dois subcampos. O primeiro de 1 dígito, indica a unidade da idade, conforme a tabela a seguir. O segundo, de dois dígitos, indica a quantidade de unidades: 0 – Idade menor de 1 hora, o subcampo varia de 01 e 59; 1 – Hora, o subcampo varia de 01 a 23; 2 – Dias, o subcampo varia de 01 a 29; 3 – Meses, o subcampo varia de 01 a 11; 4 – Anos, o subcampo varia de 00 a 99; 5 – Anos (mais de 100 anos), o segundo subcampo varia de 0 a 99. Exemplos: 020 – 20 minutos; 103 – 3 horas; 204 – 4 dias; 305 – 5 meses; 400 – menor de 1 ano, mas não se sabe o número de horas, dias ou meses; 410 – 10 anos; 505 – 105 anos.
SEXO	1	1- MASCULINO; 2 - FEMININO; 9 - IGNORADO
RACACOR	1	1 - BRANCA; 2 - PRETA; 3 - AMARELA; 4 - PARDA; 5 - INDIGENA
CODMUNRES	7	CÓDIGO DO MUNICÍPIO DE RESIDENCIA
LOCOCOR	1	Local de ocorrência do óbito: 1 – hospital; 2 – outros estabelecimentos de saúde; 3 – domicílio; 4 – via pública; 5 – outros; 9 – ignorado.
CODESTAB	8	CÓDIGO DO ESTABELECIMENTO
CODMUNOCOR	8	CÓDIGO DO MUNICÍPIO DE OCORRENCIA
IDADEMAE	2	IDADE DA MÃE
ESCMAE	1	Escolaridade em anos. Valores: 1 – Nenhuma; 2 – de 1 a 3 anos; 3 – de 4 a 7 anos; 4 – de 8 a 11 anos; 5 – 12 anos e mais; 9 – Ignorado.
OCUPMAE	6	Informar a ocupação da mãe (informar a anterior, se a mesma for aposentada ou desempregada).
QTDFILVIVO	2	Quantidade de filhos vivos
QTDFILMORT	2	Quantidade de filhos mortos
GRAVIDEZ	1	Informar o tipo de gravidez: 1 – única; 2 – dupla; 3 – tripla e mais; 9 – ignorada.
SEMAGESTAC	3	Informar o número de semanas de gestação. 99 – ignorado
GESTACAO	1	Faixa em semanas de gestação: 1 – Menos 22 semanas; 2 – 2 a 27 semanas; 3 – 8 a 31 semanas; 4 – 2 a 36 semanas; 5 – 37 a 41 semanas; 6 – 42 e + semanas.
PARTO	1	Informar o tipo de parto: 1 – vaginal; 2 – cesáreo; 9 – ignorado.
OBITOPARTO	1	Informar como foi a morte em relação ao parto: 1 – antes; 2 – durante; 3 – depois; 9 – Ignorado.
PESO	4	Peso ao nascer em gramas
CAUSABAS	4	CAUSA BÁSICA DO ÓBITO

APÊNDICE III – FERRAMENTA DE PREDIÇÃO

Essa proposta consiste na criação de um APP (Aplicativo Mobile) que consuma como API (*Application Programming Interface*) a inteligência do modelo preditivo gerado na pesquisa, permitindo que os profissionais de saúde que acompanham gestantes ou crianças recém-nascidas, possam inferir sobre o futuro dessa criança, baseado nas informações disponíveis naquela momento para prever o risco futuro, orientando os responsáveis ou tomando as ações cabíveis como agente promotor da saúde.

Essa aplicação deverá ser usada por profissionais de saúde que acompanham gestantes, agentes comunitários de saúde e os profissionais que acompanham a saúde do recém-nascido.

A Diagrama de Contexto simplificado dessa solução consiste no modelo abaixo:



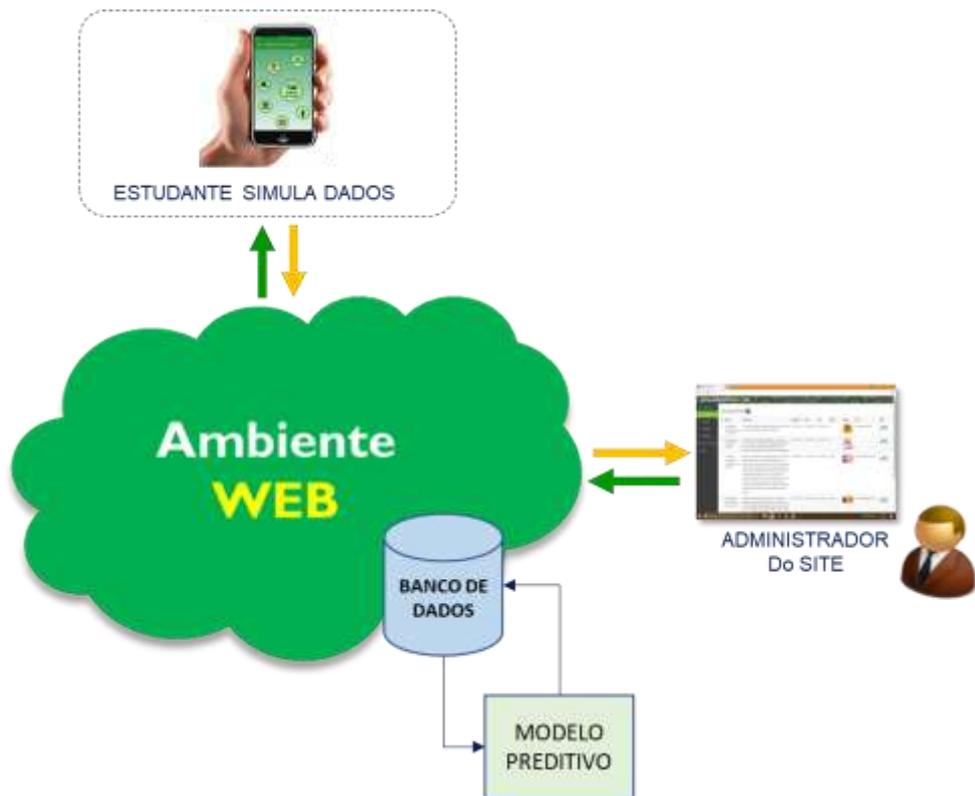
O profissional de saúde durante o atendimento coleta os dados da mãe ou da criança recém-nascida e aplica o modelo preditivo que será atualizado constantemente no banco de dados da aplicação.

APÊNDICE IV – SIMULADOR PARA A EDUCAÇÃO NA ÁREA DA SAÚDE

Essa proposta também consiste na criação de um APP (Aplicativo Mobile) que consuma como API (*Application Programming Interface*) a inteligência do modelo preditivo gerado na pesquisa, mas com o objetivo de aplicar na educação e formação dos profissionais de saúde, orientando os profissionais de saúde que acompanham gestantes ou crianças recém-nascidas, através da simulação em ambiente controlado das possíveis situações de campo que eles vão vivenciar. Na simulação as informações de entrada são fictícias, mas usando modelos preditivos reais gerados pelo modelo de forma rotineira. Esta prática vai permitir a simulação, orientando a tomada de decisão dos profissionais em treinamento.

Essa aplicação poderá ser usada por profissionais em formação ou reciclagem, como: universitários, técnicos em saúde, agentes comunitários entre outros.

A Diagrama de Contexto simplificado dessa solução consiste no modelo abaixo:



Ambiente destinado a simulação e testes para a formação dos profissionais de saúde.

APÊNDICE V – SCRIPT PYTHON PARA EXECUÇÃO

```

# -----*
#
# Spyder Editor
#
# -*- coding: utf-8 -*-
#
# scatter plot of blobs dataset
from keras.models import Sequential
from keras.layers import Dense, Activation
import keras
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.datasets.samples_generator import make_blobs
from sklearn.preprocessing import MinMaxScaler
from keras.utils import np_utils
from matplotlib import pyplot
from pandas import DataFrame
import numpy as np
import os
import sys
from keras.models import model_from_json
import pandas as pd
from sklearn.model_selection import StratifiedKFold
import warnings
warnings.simplefilter('ignore')
#
# gravado no GIT
#
#*****
# PREPARAÇÃO DOS DADOS PARA MACHINE LEARNING
#*****
#
# Primeiro Passo: Extrair os dados para um DataSet
#
# Segundo Passo: Verificar se o DataSet tem valores nulos
# Apos extrair os dados do banco de dados para o dataset
# Verificar se o dataset tem valores nulos com:
# df.isnull().values.any()
# retorna True ou False
#
# Terceiro Passo: Corrigir ou excluir os valores nulos
# excluir com df.dropna(how='all', inplace=True)
# corrigir com a média do atributo: df.fillna(df['<nome da variavel>'].mean())
#
# Quarto Passo: Identificar a correlação entre as variáveis
# Identificar a correlação entre as variáveis
# lembrando que correlação não implica causalidade
# define uma função para exibir a correlação
# def plot_corr(df, size=(quantidade de atributos)):
#     corr = df.corr()
#     fig, ax = plt.subplots(figsize = (size, size))
#     ax.matshow(corr)
#     plt.xticks(range(len(corr.columns)), corr.columns)
#     plt.yticks(range(len(corr.columns)), corr.columns)
#
# plot_corr(df) cria nesse ponto o gráfico
#
# Quinto Passo: Visualizar a correlação em Tabela
# Coeficiente de Correlação:

```

```

# +1 = forte correlação positiva (quando uma cresce a outra também cresce)
# 0 = Não há correlação (uma não interfere na outra)
# -1 = forte correlação negativa (quando uma cresce a outra diminui)
#
# df.corr()
#
# Sexto Passo: Ajustes nos dados
# exemplo: se o target for categórico transformá-lo em numérico
# define as classes (com dicionário): por exemplo:
#         obitoX_map = {True: 1, False: 0}
#
# aplica o mapeamento ao dataset:
# df['obitoX'] = df['obitoX'].map(obitoX_map)
#
# Sétimo Passo: Verificar a distribuição dos dados
# nessa fase verificamos quantas linhas são de cada classe
# exemplo:
# num_true = len(df['obitoX'] == True)
# num_false = len(df['obitoX'] == False)
# print("Numeros de casos positivos: {0} (1:2.2f)%".format(num_true / (num_true + num_false)*100))
# print("Numeros de casos negativos: {0} (1:2.2f)%".format(num_false / (num_true + num_false)*100))
# Lembrando que o ideal é balancear o número de registros de cada classe.
# Todas com números próximos, senão iguais.
#
# Oitavo Passo: Definir as variáveis features e target
# Definir as variáveis preditoras (Features Selection)
# Deve-se escolher aquelas que são mais relevantes para a
# construção do modelo de predição:
# exemplo:
# atributos = ['sexo', 'raca', 'idhm_mun_res', 'cd_regiao', 'loc_ocor', 'idademae', 'escmae', 'qt_filvivo']
#
# Definir a variável a ser prevista (target)
# exemplo:
# atrib_prev = ['obitoX']
#
# Nono Passo: Criar os objetos
# exemplo:
# X = df[atributos].values
# Y = df[atrib_prev].values
#
# Décimo Passo: Dividir os Dados em Dados de Treino e Dados de Teste
# Geralmente 70% para Treino e 30% para teste
# Essa divisão pode ser feita importando o pacote:
# from sklearn.model_selection import train_test_split
#
# definir a taxa de split, exemplo:
# split_test_size = 0.30
#
# criar os dados de Treino e de Teste, exemplo:
# X_Treino, X_Testes, Y_Treino, Y_Testes = train_test_split(X, Y, test_size = split_test_size, random_state
= 42)
#
# Décimo Primeiro Passo: Verificar o resultado do split
# print("original True : {0} ({1:0.2f}%)".format(len(df.loc[df['diabetes'] == 1]),
#         (len(df.loc[df['diabetes'] == 1])/len(df.index)* 100)))
#
# print("original False : {0} ({1:0.2f}%)".format(len(df.loc[df['diabetes'] == 0]),
#         (len(df.loc[df['diabetes'] == 0])/len(df.index)* 100)))
# print("")
# print("Training True : {0} ({1:0.2f}%)".format(len(Y_Treino[Y_Treino[:] == 1]),

```

```

#             (len(Y_Treino[Y_Treino[:]] == 1)/len(Y_Treino)* 100)))
#
# print("Training False : {0} ({1:0.2f}%)" .format(len(Y_Treino[Y_Treino[:]] == 0)),
#       (len(Y_Treino[Y_Treino[:]] == 0)/len(Y_Treino)* 100)))
#
# print("Test True : {0} ({1:0.2f}%)" .format(len(Y_Test[Y_Test[:]] == 1)),
#       (len(Y_Treino[Y_Treino[:]] == 1)/len(Y_Test)* 100)))
#
# print("Test False : {0} ({1:0.2f}%)" .format(len(Y_Test[Y_Test[:]] == 0)),
#       (len(Y_Test[Y_Test[:]] == 0)/len(Y_Test)* 100)))
#
# OBS: Deve-se manter a proporção dos valores True e False do DataSet Original,
#      nos DataSet de Treino e de Teste
#
# Décimo Segundo Passo: Identificar os valores missing ocultos
#                       um valor missing é um valor faltante
#                       os valores zeros, o que significam?
#
#                       verificando os valores zeros nas colunas
#                       print("# Linhas missing na <primeira coluna> : {0}" .format(len(df.loc[df['nome da coluna'] == 0])))
#                       repetir para todas as colunas
#
# Décimo Terceiro Passo: Tratar os valores missing encontrados no Passo anterior
#                       Uma das formas de se corrigir é substituir o zero pelo valor médio
#                       da coluna.
#                       para isso importar o pacote
#                       from sklearn.preprocessing import Imputer
#                       depois cria-se o objeto:
#                       preenche_0 = Imputer(missing_values = 0, strategy = "mean", axis =0)
#
#                       Substitui os dados zerados pela média dos dados
#                       X_Treino = preenche_0.fit_transform(X_Treino)
#                       X_Test = preenche_0.fit_transform(X_Test)
#
# fit model on dataset
def fit_model(trainX, trainy):
    # define model
    model = Sequential()
    model.add(Dense(100, input_shape=(19,), activation='relu'))
    model.add(Dense(units=10, activation='relu'))
    model.add(Dense(units=5, activation='relu'))
    model.add(Dense(units=2, activation='softmax'))
    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    # fit model
    model.fit(trainX, trainy, epochs=100, batch_size=10, verbose=0)
    return model

print("\n*****")
print('**** PREPARAÇÃO DOS DADOS PARA MACHINE LEARNING ****')
print('*****')

print("\n\nCARACTERÍSTICAS DO MODELO\n")
print(' Modelo Keras Sequential')
print(' Estrutura de Comitê com 5 Redes Neurais em Stacked (empilhamento) Model')
print(' Rede Neural Deep Learning com 4 camadas')
print(' estratégia de split dos dados com K FOLDS (K=10)')
print(' modelo final ajustado pela média dos 5 modelos com Logistic Regression\n\n')
print("\n PRIMEIRO PASSO: Extrair os dados para um DataSet")

# generate 2d classification dataset

```

```

dfexcel= pd.read_excel(r'C:\Users\suporte\Documents\Projeto_PYTHON\lePROFV2_dados.xlsx')
df = dfexcel.copy()
print(df)

lista1 =[]
lista2 =[]
for i in df['CLASSIFICAÇÃO']:
    if i == 1: #Obito
        lista1.append(1)
    elif i == 2: #Vida
        lista2.append(0)
num_obito = len(lista1)
num_vida = len(lista2)
print('Dados Extraídos')

Y1 = df.iloc[:,22].values # coluna 22 : Classificação
base = df.drop('CLASSIFICAÇÃO', axis=1)
base = base.drop('NU_DO', axis=1)
base = base.drop('NU_DN', axis=1)

print("\n SEGUNDO PASSO: Verificar se o DataSet tem valores nulos')
print('Tem Nulos: ',base.isnull().values.any())

#Substitui todos os valores null pela média das devidas colunas
print("\n TERCEIRO PASSO: Tratar os valores nulos')
media = (base['APGAR1'].mean() + base["APGAR5"].mean()) / 2
base = base.fillna(media)

print('Ainda Tem Nulos?: ',base.isnull().values.any())

print('Dados limpos sem nulos')
print(base)
X2 = base.iloc[:,1:20].values # colunas do 0 até a coluna 20

print("\n QUARTO PASSO: Identificar a correlação entre as variáveis')
print('Correlação...:', base.corr())

#print('y...',Y1)
#print('X...',X1)

print("\n QUINTO PASSO: Ajustar os dados - Scaling')
# Feature Scaling
sc = MinMaxScaler(feature_range = (0, 1))
X1 = sc.fit_transform(X2)
print('Dados Ajustados..Re-Scaling',X1)

print("\n SEXTO PASSO: Resumo Estatístico')
# Resumo estatístico
print(base.describe())

print("\n SÉTIMO PASSO: Verificação da Distribuição dos Dados')
print("Total casos de óbitos:",num_obito)
print("Total casos de vida:",num_vida)
print("Percentual de casos óbitos: {0} (1:2.2f)%".format(num_obito/ (num_obito + num_vida)*100))
print("Percentual de casos vida: {0} (1:2.2f)%".format(num_vida/ (num_obito + num_vida)*100))

# transforma os outputs em integer on-hot-encoder
onehot_encoder = OneHotEncoder(sparse=False) # gerar a matriz de vetores binarios
inteiros = Y1.reshape(len(Y1),1)
Y = inteiros

```

```

#Yi = onehot_encoder.fit_transform(inteiros)
X = X1

print('Targets...:', Y1)

print("\n OITAVO PASSO: Definir as variáveis features e target')
print('Variáveis definidas')

print("\n NONO PASSO: Dividir os Dados em Dados de Treino e Dados de Teste')
print('Definido com 10 folds cross validation')
# Define 10 folds cross validation testes harness
# fix random seed for reproducibility
seed = 7
kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=seed)
#cvscores = Y.reshape(len(Y),4)
cvscores = []

print("\n DÉCIMO PASSO: Iniciar o processo de Treinamento e Teste do Modelo')
for train, test in kfold.split(X, Y):
    # create model
    model = Sequential()
    model.add(Dense(100, input_shape=(19,), activation='relu'))
    model.add(Dense(units=10, activation='relu'))
    model.add(Dense(units=5, activation='relu'))
    model.add(Dense(units=2, activation='softmax'))

    print("***** TREINANDO O MODELO *****")
    # Divisão do kfold em X e Y treinamento
    Xtrain = X[train]
    # onehot_encoder
    Ytrain = onehot_encoder.fit_transform(Y[train])

    # compile model
    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

    # Fit model
    model.fit(Xtrain, Ytrain, epochs=100, batch_size=10, verbose=2)

    print("***** AVALIANDO O MODELO *****")
    # evaluate the model
    # Divisão do kfold em X e Y teste
    Xtest = X[test]
    # onehot_encoder
    Ytest = onehot_encoder.fit_transform(Y[test])

    scores = model.evaluate(Xtest, Ytest, verbose=0)
    print("Teste...", "%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
    cvscores.append(scores[1] * 100)

print("%.2f%% (+/- %.2f%%)" % (np.mean(cvscores), np.std(cvscores)))

# fit and save models (submodleos)
n_members = 5
for i in range(n_members):
    # fit model
    model = fit_model(Xtrain, Ytrain)
    # save model
    filename = r'C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof' +
str(i + 1) + '.h5'
    model.save(filename)

```

```

print('>Salvo %s' % filename)

#####
#####  COMBINE COMMITTEE MACHINE  #####
#####

#
# Carrega apartir dos arquivos .h5
from keras.utils import to_categorical
from numpy import dstack
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import pickle
from datetime import datetime
from warnings import simplefilter
simplefilter('ignore')

inicioTempo = datetime.now() #Início Execução

all_models = []
def load_all_models(n_models):
    for i in range(n_models):
        # define model
        model = Sequential()
        model.add(Dense(100, input_shape=(19,), activation='relu'))
        model.add(Dense(units=10, activation='relu'))
        model.add(Dense(units=5, activation='relu'))
        model.add(Dense(units=2, activation='softmax'))
        model.compile(loss='categorical_crossentropy', optimizer='nadam', metrics=['accuracy'])

        # define filename for this ensemble
        filename = r'C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\ + 'model_le_prof' +
str(i + 1) + '.h5'
        model.load_weights(filename)
        print("Carrega o modelo")
        # add to list of members
        all_models.append(model)
        print('>Modelo carregado %s' % filename)
    return all_models

# cria conjuntos de dados de entrada de modelo empilhados como saídas do conjunto
def stacked_dataset(members, inputX):
    stackX = None
    for model in members:
        # make prediction
        yhat = model.predict(inputX, verbose=0)
        # stack predictions into [rows, members, probabilities]
        if stackX is None:
            stackX = yhat
        else:
            stackX = dstack((stackX, yhat))
    # flatten predictions to [rows, members x probabilities]
    stackX = stackX.reshape((stackX.shape[0], stackX.shape[1]*stackX.shape[2]))
    return stackX

# Treina o modelo baseado nas saídas dos membros do conjunto
def fit_stacked_model(members, inputX, inputy):
    # create dataset using ensemble
    stackedX = stacked_dataset(members, inputX)
    # fit standalone model
    model = LogisticRegression()

```

```

    model.fit(stackedX, inputy)
    return model

# Faz predição com o stacked model
def stacked_prediction(members, model, inputX):
    # Cria dataset usando o esamble
    stackedX = stacked_dataset(members, inputX)
    # Faz predição
    yhat = model.predict(stackedX)
    return yhat

print("*****Iniciando o modelo com um novo dataset*****\n")
df = pd.read_excel('lePROFV2_dados.xlsx') #Carrega dados da planilha

Y1 = df.iloc[:,22].values # coluna 22 : Classificação
base = df.drop('CLASSIFICAÇÃO', axis=1)
base = base.drop('NU_DO', axis=1)
base = base.drop('NU_DN', axis=1)

media = (base['APGAR1'].mean() + base['APGAR5'].mean()) / 2 #Substitui o valores nulos pela média das
colunas (APAGR1 e APAGR5)
base = base.fillna(media)

X2 = base.iloc[:,1:20].values # colunas do 1 até a coluna 20, excluindo a coluna MM_obito

# Feature Scaling
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler(feature_range = (0, 1))
X1 = sc.fit_transform(X2)

X = X1 # Array final dos dados X
y = Y1 # Array final dos dados Y

# Divide os dados em treino e teste
n_train = 0
trainX, testX = X[:n_train, :], X[n_train:, :]
trainy, testy = y[:n_train], y[n_train:]

# Carrega todos os modelos
n_members = 5
members = load_all_models(n_members)
print("Total modelos carregados...: %d ' % n_members+"\n")

# Avalia os modelos separadamente
print("*****Avaliando a acurácia de cada modelo separadamente*****")
for model in members:
    testy_enc = to_categorical(testy)
    _, acc = model.evaluate(testX, testy_enc[:, 1:3], verbose=0)
    print('Acurácia do modelo: %.3f % acc)

# fit stacked model using the ensemble
model = fit_stacked_model(members, testX, testy)

# Avalia o conjunto de dados de teste
yPred = stacked_prediction(members, model, testX[:,0:20])
acc = accuracy_score(testy, yPred)
print('Accuracia do teste: %.3f % acc+\n')

# Salva o modelo no disk
filename = r'C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\RegressionModel.sav'
pickle.dump(model, open(filename, 'wb'))

```

```

print('Modelo de regressão salvo \n')
#print("\n***** Fim de Treinamento do Modelo *****")

***** Predição do modelo *****

#NovoDf = df.drop('CLASSIFICAÇÃO', axis=1)
#MontaDataFrame = []
listPred = []
for i in range(len(yPred)):
    #df1 = pd.DataFrame(data)
    if yPred[i] == 1:
        listPred.append('Óbito')
    elif yPred[i] == 2:
        listPred.append('Vida')
    #MontaDataFrame.append({"NU_DO":df["NU_DO"], "MM_OBITO":df["MM_OBITO"],
"NU_DN":df["NU_DN"][i], "Valor Predito":listPred[i]})
#NovoDf['Valor Predito'] = listPred
df['Class Predita'] = listPred

#df1 = pd.DataFrame(MontaDataFrame)

#Salva o dataframe em excel
df.to_excel(r'C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\SaidaToExcel.xlsx',
            sheet_name='class_final')

fimTempo = datetime.now() #Fim execução
print('Duração: {}'.format(fimTempo - inicioTempo))

#*-----*

```

APÊNDICE VI – EXECUÇÃO DO SCRIPT PYTHON

```
*****
*****  PREPARAÇÃO DOS DADOS PARA MACHINE LEARNING  *****
*****
```

CARACTERÍSTICAS DO MODELO e EXECUÇÃO DO MODELO

Modelo Keras Sequential

Estrutura de Comitê com 5 Redes Neurais em Stacked (empilhamento) Model

Rede Neural Deep Learning com 4 camadas

estratégia de split dos dados com K FOLDS (K=10)

modelo final ajustado pela média dos 5 modelos com Logistic Regression

)

PRIMEIRO PASSO: Extrair os dados para um DataSet

	NU_DO	NU_DN	MM_OBITO	HORAS_VIDA	SEXO	RACA	IDHM_MUN_RES	\
0	18605719	18605719	4	6	1	4	3	
1	22740151	70346072	1	3	1	4	4	
2	22576293	64498949	2	8	1	4	4	
3	20833889	20833889	7	2	1	4	4	
4	23330406	71751540	4	1	1	4	4	
5	25864962	74398859	6	1	2	4	3	
6	24711008	71820124	11	1	2	4	3	
7	24711743	74329931	11	2	1	4	3	
8	22552950	64504932	12	1	1	1	4	
9	25863424	75690990	12	1	2	4	4	
10	18597901	18597901	5	1	1	4	4	

...
18380	0	75751043	0	8	1	4	3	
18381	0	69636535	0	8	1	4	3	
18382	0	75674819	0	8	1	4	4	
18383	0	75648765	0	8	1	4	3	
18384	0	74366430	0	8	1	4	3	
18385	0	71794778	0	8	1	4	3	
18386	0	75723348	0	8	1	4	2	
18387	0	75758139	0	8	1	4	3	
18388	0	75706478	0	8	1	4	3	
18389	0	75725460	0	8	2	4	4	
18390	0	75675036	0	8	2	4	4	
18391	0	75708643	0	8	1	4	3	
18392	0	75736961	0	8	2	1	3	
18393	0	74400245	0	8	1	4	4	

	CD_REGIÃO	LOC_OCOR	IDADEMAE	...	GRAVIDEZ	GESTACAO	PARTO	PESO	\
0	6	1	3	...	1	9	2	1	
1	7	1	5	...	1	2	1	1	
2	7	1	4	...	1	9	2	1	
3	5	1	4	...	1	3	2	1	
4	7	1	2	...	1	1	1	1	
5	7	1	3	...	1	1	1	1	
6	4	1	2	...	1	1	1	1	
7	4	1	4	...	1	1	1	1	
8	7	2	1	...	1	1	1	1	
9	7	1	3	...	2	1	1	1	
10	7	1	4	...	1	1	1	1	
...	
18380	6	1	4	...	1	5	2	7	

18381	5	1	5 ...	1	5	2	7
18382	7	1	4 ...	1	5	2	7
18383	4	1	5 ...	1	5	2	7
18384	6	1	4 ...	1	5	2	7
18385	2	1	6 ...	1	5	1	7
18386	4	1	6 ...	1	4	2	7
18387	2	1	5 ...	1	5	2	7
18388	2	1	2 ...	1	6	2	7
18389	7	1	4 ...	1	5	1	7
18390	7	1	4 ...	1	5	2	7
18391	3	1	3 ...	1	5	1	7
18392	5	1	6 ...	1	5	2	7
18393	7	1	3 ...	1	5	2	7

	ASSISTMED	APGAR1	APGAR5	ESTCIVIL	QTD_CONSULTAS	CLASSIFICAÇÃO
0	1	6.0	6.0	1	2	1
1	1	3.0	6.0	1	2	1
2	1	9.0	6.0	5	4	1
3	1	6.0	6.0	1	2	1
4	1	1.0	6.0	1	1	1
5	1	6.0	6.0	1	1	1
6	1	1.0	6.0	2	2	1
7	1	2.0	6.0	2	3	1
8	1	2.0	6.0	1	1	1
9	1	2.0	6.0	1	1	1
10	1	6.0	6.0	1	2	1
...
18380	1	9.0	10.0	2	4	2
18381	1	8.0	8.0	1	3	2
18382	1	9.0	9.0	1	4	2
18383	1	8.0	9.0	2	4	2
18384	1	9.0	10.0	1	4	2
18385	1	6.0	9.0	1	3	2
18386	1	8.0	9.0	2	4	2
18387	1	9.0	9.0	5	3	2
18388	1	8.0	9.0	1	4	2
18389	1	7.0	8.0	2	4	2
18390	1	8.0	8.0	1	4	2
18391	1	8.0	9.0	5	2	2
18392	1	6.0	7.0	2	2	2
18393	1	8.0	9.0	2	4	2

[18394 rows x 23 columns]

Dados Extraídos

SEGUNDO PASSO: Verificar se o DataSet tem valores nulos

Tem Nulos: True

TERCEIRO PASSO: Tratar os valores nulos

Ainda Tem Nulos?: False

Dados limpos sem nulos

	MM_OBITO	HORAS_VIDA	SEXO	RACA	IDHM_MUN_RES	CD_REGIÃO	LOC_OCOR \
0	4	6	1	4	3	6	1
1	1	3	1	4	4	7	1
2	2	8	1	4	4	7	1
3	7	2	1	4	4	5	1
4	4	1	1	4	4	7	1
5	6	1	2	4	3	7	1
6	11	1	2	4	3	4	1
7	11	2	1	4	3	4	1

8	12	1	1	1	4	7	2
9	12	1	2	4	4	7	1
10	5	1	1	4	4	7	1
...
18380	0	8	1	4	3	6	1
18381	0	8	1	4	3	5	1
18382	0	8	1	4	4	7	1
18383	0	8	1	4	3	4	1
18384	0	8	1	4	3	6	1
18385	0	8	1	4	3	2	1
18386	0	8	1	4	2	4	1
18387	0	8	1	4	3	2	1
18388	0	8	1	4	3	2	1
18389	0	8	2	4	4	7	1
18390	0	8	2	4	4	7	1
18391	0	8	1	4	3	3	1
18392	0	8	2	1	3	5	1
18393	0	8	1	4	4	7	1

	IDADEMAE	ESMAE	QT_FILVIVO	QT_FILMORT	GRAVIDEZ	GESTACAO	PARTO \
0	3	5	1	0	1	9	2
1	5	4	2	0	1	2	1
2	4	9	2	0	1	9	2
3	4	4	2	1	1	3	2
4	2	2	3	1	1	1	1
5	3	4	0	3	1	1	1
6	2	4	2	0	1	1	1
7	4	4	1	0	1	1	1
8	1	3	0	0	1	1	1
9	3	9	0	1	2	1	1
10	4	4	0	1	1	1	1

...
18380	4	4	2	0	1	5	2
18381	5	4	1	0	1	5	2
18382	4	4	2	0	1	5	2
18383	5	4	1	1	1	5	2
18384	4	4	1	1	1	5	2
18385	6	5	1	0	1	5	1
18386	6	1	4	0	1	4	2
18387	5	4	1	0	1	5	2
18388	2	4	0	0	1	6	2
18389	4	5	2	0	1	5	1
18390	4	4	1	0	1	5	2
18391	3	4	3	0	1	5	1
18392	6	4	3	1	1	5	2
18393	3	4	1	0	1	5	2

	PESO	ASSISTMED	APGAR1	APGAR5	ESTCIVIL	QTD_CONSULTAS
0	1	1	6.000000	6.000000	1	2
1	1	1	3.000000	6.000000	1	2
2	1	1	9.000000	6.000000	5	4
3	1	1	6.000000	6.000000	1	2
4	1	1	1.000000	6.000000	1	1
5	1	1	6.000000	6.000000	1	1
6	1	1	1.000000	6.000000	2	2
7	1	1	2.000000	6.000000	2	3
8	1	1	2.000000	6.000000	1	1
9	1	1	2.000000	6.000000	1	1
10	1	1	6.000000	6.000000	1	2
...

18380	7	1	9.000000	10.000000	2	4
18381	7	1	8.000000	8.000000	1	3
18382	7	1	9.000000	9.000000	1	4
18383	7	1	8.000000	9.000000	2	4
18384	7	1	9.000000	10.000000	1	4
18385	7	1	6.000000	9.000000	1	3
18386	7	1	8.000000	9.000000	2	4
18387	7	1	9.000000	9.000000	5	3
18388	7	1	8.000000	9.000000	1	4
18389	7	1	7.000000	8.000000	2	4
18390	7	1	8.000000	8.000000	1	4
18391	7	1	8.000000	9.000000	5	2
18392	7	1	6.000000	7.000000	2	2
18393	7	1	8.000000	9.000000	2	4

[18394 rows x 20 columns]

QUARTO PASSO: Identificar a correlação entre as variáveis

Correlação...:	MM_OBITO	HORAS_VIDA	SEXO	RACA	IDHM_MUN_RES \
MM_OBITO	1.000000	-0.569572	-0.036370	-0.147684	-0.006025
HORAS_VIDA	-0.569572	1.000000	0.028156	0.079573	0.019063
SEXO	-0.036370	0.028156	1.000000	-0.004781	-0.009760
RACA	-0.147684	0.079573	-0.004781	1.000000	0.009917
IDHM_MUN_RES	-0.006025	0.019063	-0.009760	0.009917	1.000000
CD_REGIÃO	-0.015166	0.028858	-0.001842	0.017286	0.515123
LOC_OCOR	0.124115	-0.048103	0.000651	-0.059731	-0.059751
IDADEMAE	-0.053892	0.048163	0.014016	-0.018825	0.057658
ESCMAE	0.209464	-0.057876	-0.022122	-0.080042	0.072676
QT_FILVIVO	0.090856	-0.056184	0.005341	-0.012295	-0.091078
QT_FILMORT	0.038093	-0.058294	0.017596	0.006269	0.005482
GRAVIDEZ	0.130001	-0.152125	-0.002435	-0.019969	0.021172
GESTACAO	-0.071295	0.305420	-0.006893	-0.028921	-0.053135
PARTO	0.192113	0.021096	-0.015541	-0.071300	0.017069
PESO	-0.575384	0.581386	0.002084	0.097847	-0.046137
ASSISTMED	0.107319	-0.053496	-0.007646	-0.038671	-0.030258
APGAR1	-0.454210	0.623822	0.023105	0.057118	-0.006505
APGAR5	-0.483599	0.533185	0.026983	0.084031	-0.009371
ESTCIVIL	0.050612	-0.031416	0.004860	-0.028380	-0.083784
QTD_CONSULTAS	-0.243676	0.279686	0.003949	0.032082	-0.061718

	CD_REGIÃO	LOC_OCOR	IDADEMAE	ESCMAE	QT_FILVIVO \
MM_OBITO	-0.015166	0.124115	-0.053892	0.209464	0.090856
HORAS_VIDA	0.028858	-0.048103	0.048163	-0.057876	-0.056184
SEXO	-0.001842	0.000651	0.014016	-0.022122	0.005341
RACA	0.017286	-0.059731	-0.018825	-0.080042	-0.012295
IDHM_MUN_RES	0.515123	-0.059751	0.057658	0.072676	-0.091078
CD_REGIÃO	1.000000	-0.050441	0.030498	0.020595	-0.036005
LOC_OCOR	-0.050441	1.000000	-0.028540	0.032274	0.046320
IDADEMAE	0.030498	-0.028540	1.000000	-0.048441	0.391219
ESCMAE	0.020595	0.032274	-0.048441	1.000000	-0.233590
QT_FILVIVO	-0.036005	0.046320	0.391219	-0.233590	1.000000
QT_FILMORT	-0.015015	-0.007264	0.215062	-0.088337	0.216705
GRAVIDEZ	0.018030	-0.012471	0.031490	-0.018621	0.084854
GESTACAO	-0.050427	0.095474	-0.045832	0.315084	-0.095986
PARTO	0.002930	0.094281	-0.042544	0.537037	-0.159080
PESO	-0.038545	-0.058914	0.065793	-0.228525	0.001066
ASSISTMED	-0.040945	0.316124	-0.030447	0.022507	0.027038
APGAR1	-0.001373	-0.004704	0.039135	-0.030076	-0.038904
APGAR5	0.000568	-0.043935	0.048971	-0.166183	-0.009387
ESTCIVIL	0.019812	0.018661	0.007323	-0.052311	0.110707

QTD_CONSULTAS	-0.069490	-0.043232	0.068583	-0.088241	-0.047956	
MM_OBITO	0.038093	0.130001	-0.071295	0.192113	-0.575384	0.107319
HORAS_VIDA	-0.058294	-0.152125	0.305420	0.021096	0.581386	-0.053496
SEXO	0.017596	-0.002435	-0.006893	-0.015541	0.002084	-0.007646
RACA	0.006269	-0.019969	-0.028921	-0.071300	0.097847	-0.038671
IDHM_MUN_RES	0.005482	0.021172	-0.053135	0.017069	-0.046137	-0.030258
CD_REGIÃO	-0.015015	0.018030	-0.050427	0.002930	-0.038545	-0.040945
LOC_OCOR	-0.007264	-0.012471	0.095474	0.094281	-0.058914	0.316124
IDADEMAE	0.215062	0.031490	-0.045832	-0.042544	0.065793	-0.030447
ESMAE	-0.088337	-0.018621	0.315084	0.537037	-0.228525	0.022507
QT_FILVIVO	0.216705	0.084854	-0.095986	-0.159080	0.001066	0.027038
QT_FILMORT	1.000000	0.036268	-0.105256	-0.075985	-0.042883	-0.001200
GRAVIDEZ	0.036268	1.000000	-0.183853	-0.038232	-0.261869	-0.002743
GESTACAO	-0.105256	-0.183853	1.000000	0.508343	0.295195	0.071853
PARTO	-0.075985	-0.038232	0.508343	1.000000	-0.223854	0.072029
PESO	-0.042883	-0.261869	0.295195	-0.223854	1.000000	-0.057854
ASSISTMED	-0.001200	-0.002743	0.071853	0.072029	-0.057854	1.000000
APGAR1	-0.059516	-0.134477	0.298542	0.017971	0.546005	-0.016481
APGAR5	-0.033325	-0.126921	0.116567	-0.160976	0.499361	-0.037387
ESTCIVIL	0.031188	-0.012662	0.021688	-0.072598	0.085400	0.008499
QTD_CONSULTAS	0.000236	-0.093173	0.191545	-0.058804	0.462060	-0.017305

	APGAR1	APGAR5	ESTCIVIL	QTD_CONSULTAS
MM_OBITO	-0.454210	-0.483599	0.050612	-0.243676
HORAS_VIDA	0.623822	0.533185	-0.031416	0.279686
SEXO	0.023105	0.026983	0.004860	0.003949
RACA	0.057118	0.084031	-0.028380	0.032082
IDHM_MUN_RES	-0.006505	-0.009371	-0.083784	-0.061718
CD_REGIÃO	-0.001373	0.000568	0.019812	-0.069490
LOC_OCOR	-0.004704	-0.043935	0.018661	-0.043232
IDADEMAE	0.039135	0.048971	0.007323	0.068583
ESMAE	-0.030076	-0.166183	-0.052311	-0.088241
QT_FILVIVO	-0.038904	-0.009387	0.110707	-0.047956
QT_FILMORT	-0.059516	-0.033325	0.031188	0.000236
GRAVIDEZ	-0.134477	-0.126921	-0.012662	-0.093173
GESTACAO	0.298542	0.116567	0.021688	0.191545
PARTO	0.017971	-0.160976	-0.072598	-0.058804
PESO	0.546005	0.499361	0.085400	0.462060
ASSISTMED	-0.016481	-0.037387	0.008499	-0.017305
APGAR1	1.000000	0.736020	0.000373	0.279619
APGAR5	0.736020	1.000000	0.005794	0.250981
ESTCIVIL	0.000373	0.005794	1.000000	0.143884
QTD_CONSULTAS	0.279619	0.250981	0.143884	1.000000

QUINTO PASSO: Ajustar os dados - Scaling

```
Dados Ajustados..Re-Scaling [[0.71428571 0.    0.75    ... 0.06060606 0.    0.125  ]
[0.28571429 0.    0.75    ... 0.06060606 0.    0.125  ]
[1.    0.    0.75    ... 0.06060606 1.    0.375  ]
...
[1.    0.    0.75    ... 0.09090909 1.    0.125  ]
[1.    1.    0.    ... 0.07070707 0.25  0.125  ]
[1.    0.    0.75    ... 0.09090909 0.25  0.375  ]]
```

SEXTO PASSO: Resumo Estatístico

	MM_OBITO	HORAS_VIDA	SEXO	RACA	IDHM_MUN_RES \
count	18394.000000	18394.000000	18394.000000	18394.000000	18394.000000
mean	2.833043	6.508644	1.466620	3.598347	3.216919
std	3.907220	2.427966	0.498898	1.016971	0.623159

min	0.000000	1.000000	1.000000	1.000000	2.000000
25%	0.000000	6.000000	1.000000	4.000000	3.000000
50%	0.000000	8.000000	1.000000	4.000000	3.000000
75%	5.000000	8.000000	2.000000	4.000000	4.000000
max	12.000000	8.000000	2.000000	5.000000	4.000000

	CD_REGIÃO	LOC_OCOR	IDADEMAE	ESCMAE	QT_FILVIVO \
count	18394.000000	18394.000000	18394.000000	18394.000000	18394.000000
mean	5.370882	1.081113	3.580896	4.322877	1.068555
std	1.790954	0.469651	1.369768	1.794405	1.358857
min	1.000000	1.000000	1.000000	1.000000	0.000000
25%	4.000000	1.000000	3.000000	3.000000	0.000000
50%	6.000000	1.000000	3.000000	4.000000	1.000000
75%	7.000000	1.000000	5.000000	4.000000	2.000000
max	7.000000	5.000000	7.000000	9.000000	9.000000

	QT_FILMORT	GRAVIDEZ	GESTACAO	PARTO	PESO \
count	18394.000000	18394.000000	18394.000000	18394.000000	18394.000000
mean	0.230564	1.053278	4.796673	1.978743	5.370773
std	0.557818	0.236618	1.815582	1.834071	1.969281
min	0.000000	1.000000	1.000000	1.000000	1.000000
25%	0.000000	1.000000	4.000000	1.000000	3.000000
50%	0.000000	1.000000	5.000000	2.000000	6.000000
75%	0.000000	1.000000	5.000000	2.000000	7.000000
max	5.000000	3.000000	9.000000	9.000000	7.000000

	ASSISTMED	APGAR1	APGAR5	ESTCIVIL	QTD_CONSULTAS
count	18394.000000	18394.000000	18394.000000	18394.000000	18394.000000
mean	1.014733	7.069039	8.231106	2.529086	3.380505
std	0.120486	2.319332	1.528372	1.678681	0.941785
min	1.000000	0.000000	0.000000	1.000000	1.000000
25%	1.000000	6.000000	7.650024	1.000000	3.000000
50%	1.000000	8.000000	9.000000	2.000000	4.000000
75%	1.000000	9.000000	9.000000	5.000000	4.000000
max	2.000000	99.000000	99.000000	5.000000	9.000000

SÉTIMO PASSO: Verificação da Distribuição dos Dados

Total casos de óbitos: 8159

Total casos de vida: 10235

Percentual de casos óbitos: 44.356855496357504 (1:2.2f)%

Percentual de casos vida: 55.643144503642496 (1:2.2f)%

Targets....: [1 1 1 ... 2 2 2]

OITAVO PASSO: Definir as variáveis features e target

Variáveis definidas

NONO PASSO: Dividir os Dados em Dados de Treino e Dados de Teste
Definido com 10 folds cross validation

DÉCIMO PASSO: Iniciar o processo de Treinamento e Teste do Modelo

***** TREINANDO O MODELO (1) *****

Epoch 1/100	- 20s - loss: 0.2113 - acc: 0.9166
Epoch 2/100	- 10s - loss: 0.1203 - acc: 0.9580
Epoch 3/100	- 10s - loss: 0.1077 - acc: 0.9634
Epoch 4/100	- 10s - loss: 0.1001 - acc: 0.9668
Epoch 5/100	- 10s - loss: 0.0945 - acc: 0.9688
Epoch 6/100	- 10s - loss: 0.0909 - acc: 0.9709
Epoch 7/100	- 11s - loss: 0.0900 - acc: 0.9694
Epoch 8/100	- 11s - loss: 0.0865 - acc: 0.9718
Epoch 9/100	- 10s - loss: 0.0869 - acc: 0.9711

```

Epoch 10/100 - 10s - loss: 0.0854 - acc: 0.9718
Epoch 11/100 - 10s - loss: 0.0843 - acc: 0.9723
...
Epoch 89/100 - 10s - loss: 0.0557 - acc: 0.9831
Epoch 90/100 - 10s - loss: 0.0559 - acc: 0.9821
Epoch 91/100 - 10s - loss: 0.0557 - acc: 0.9833
Epoch 92/100 - 10s - loss: 0.0547 - acc: 0.9828
Epoch 93/100 - 10s - loss: 0.0556 - acc: 0.9828
Epoch 94/100 - 10s - loss: 0.0550 - acc: 0.9831
Epoch 95/100 - 10s - loss: 0.0559 - acc: 0.9828
Epoch 96/100 - 10s - loss: 0.0562 - acc: 0.9830
Epoch 97/100 - 10s - loss: 0.0544 - acc: 0.9836
Epoch 98/100 - 10s - loss: 0.0547 - acc: 0.9830
Epoch 99/100 - 10s - loss: 0.0539 - acc: 0.9836
Epoch 100/100 - 10s - loss: 0.0539 - acc: 0.9829
***** AVALIANDO O MODELO (1) *****
Teste.... acc: 97.61%

```

```

***** TREINANDO O MODELO (2) *****
Epoch 1/100 - 18s - loss: 0.1944 - acc: 0.9284
Epoch 2/100 - 10s - loss: 0.1194 - acc: 0.9596
Epoch 3/100 - 10s - loss: 0.1076 - acc: 0.9654
Epoch 4/100 - 10s - loss: 0.1002 - acc: 0.9665
Epoch 5/100 - 10s - loss: 0.0969 - acc: 0.9682
Epoch 6/100 - 10s - loss: 0.0928 - acc: 0.9694
Epoch 7/100 - 10s - loss: 0.0914 - acc: 0.9697
Epoch 8/100 - 10s - loss: 0.0899 - acc: 0.9701
Epoch 9/100 - 10s - loss: 0.0876 - acc: 0.9721
Epoch 10/100 - 10s - loss: 0.0855 - acc: 0.9721
Epoch 11/100 - 10s - loss: 0.0856 - acc: 0.9737
...
Epoch 89/100 - 10s - loss: 0.0601 - acc: 0.9807
Epoch 90/100 - 10s - loss: 0.0611 - acc: 0.9801
Epoch 91/100 - 10s - loss: 0.0589 - acc: 0.9820
Epoch 92/100 - 10s - loss: 0.0599 - acc: 0.9805
Epoch 93/100 - 10s - loss: 0.0578 - acc: 0.9809
Epoch 94/100 - 10s - loss: 0.0591 - acc: 0.9804
Epoch 95/100 - 10s - loss: 0.0576 - acc: 0.9819
Epoch 96/100 - 10s - loss: 0.0593 - acc: 0.9808
Epoch 97/100 - 10s - loss: 0.0585 - acc: 0.9805
Epoch 98/100 - 10s - loss: 0.0586 - acc: 0.9811
Epoch 99/100 - 10s - loss: 0.0583 - acc: 0.9802
Epoch 100/100 - 10s - loss: 0.0570 - acc: 0.9818
***** AVALIANDO O MODELO (2) *****
Teste.... acc: 98.04%

```

```

***** TREINANDO O MODELO (3) *****
Epoch 1/100 - 19s - loss: 0.1824 - acc: 0.9354
Epoch 2/100 - 10s - loss: 0.1146 - acc: 0.9613
Epoch 3/100 - 10s - loss: 0.1049 - acc: 0.9651
Epoch 4/100 - 11s - loss: 0.0983 - acc: 0.9677
Epoch 5/100 - 11s - loss: 0.0953 - acc: 0.9683
Epoch 6/100 - 11s - loss: 0.0912 - acc: 0.9705
Epoch 7/100 - 11s - loss: 0.0884 - acc: 0.9713
Epoch 8/100 - 11s - loss: 0.0862 - acc: 0.9728
Epoch 9/100 - 10s - loss: 0.0863 - acc: 0.9711
Epoch 10/100 - 10s - loss: 0.0833 - acc: 0.9723
Epoch 11/100 - 11s - loss: 0.0841 - acc: 0.9726
...
Epoch 89/100 - 10s - loss: 0.0600 - acc: 0.9802

```

Epoch 90/100 - 10s - loss: 0.0581 - acc: 0.9813
 Epoch 91/100 - 10s - loss: 0.0600 - acc: 0.9810
 Epoch 92/100 - 10s - loss: 0.0574 - acc: 0.9814
 Epoch 93/100 - 10s - loss: 0.0581 - acc: 0.9808
 Epoch 94/100 - 10s - loss: 0.0573 - acc: 0.9817
 Epoch 95/100 - 10s - loss: 0.0572 - acc: 0.9816
 Epoch 96/100 - 10s - loss: 0.0561 - acc: 0.9818
 Epoch 97/100 - 10s - loss: 0.0560 - acc: 0.9820
 Epoch 98/100 - 10s - loss: 0.0554 - acc: 0.9821
 Epoch 99/100 - 10s - loss: 0.0566 - acc: 0.9814
 Epoch 100/100 - 10s - loss: 0.0548 - acc: 0.9822
 ***** AVALIANDO O MODELO (3) *****
 Teste.... acc: 97.17%

***** TREINANDO O MODELO (4) *****
 Epoch 1/100 - 18s - loss: 0.1810 - acc: 0.9358
 Epoch 2/100 - 10s - loss: 0.1183 - acc: 0.9606
 Epoch 3/100 - 13s - loss: 0.1041 - acc: 0.9645
 Epoch 4/100 - 15s - loss: 0.0963 - acc: 0.9680
 Epoch 5/100 - 11s - loss: 0.0949 - acc: 0.9691
 Epoch 6/100 - 11s - loss: 0.0924 - acc: 0.9702
 Epoch 7/100 - 10s - loss: 0.0890 - acc: 0.9703
 Epoch 8/100 - 10s - loss: 0.0872 - acc: 0.9715
 Epoch 9/100 - 10s - loss: 0.0864 - acc: 0.9728
 Epoch 10/100 - 10s - loss: 0.0870 - acc: 0.9715
 Epoch 11/100 - 10s - loss: 0.0841 - acc: 0.9718

 Epoch 89/100 - 11s - loss: 0.0563 - acc: 0.9819
 Epoch 90/100 - 11s - loss: 0.0586 - acc: 0.9805
 Epoch 91/100 - 11s - loss: 0.0559 - acc: 0.9821
 Epoch 92/100 - 11s - loss: 0.0558 - acc: 0.9809
 Epoch 93/100 - 11s - loss: 0.0557 - acc: 0.9828
 Epoch 94/100 - 12s - loss: 0.0547 - acc: 0.9819
 Epoch 95/100 - 11s - loss: 0.0553 - acc: 0.9821
 Epoch 96/100 - 11s - loss: 0.0539 - acc: 0.9827
 Epoch 97/100 - 11s - loss: 0.0528 - acc: 0.9825
 Epoch 98/100 - 11s - loss: 0.0534 - acc: 0.9816
 Epoch 99/100 - 11s - loss: 0.0538 - acc: 0.9828
 Epoch 100/100 - 11s - loss: 0.0560 - acc: 0.9823
 ***** AVALIANDO O MODELO (4) *****
 Teste.... acc: 96.96%

***** TREINANDO O MODELO (5) *****
 Epoch 1/100 - 20s - loss: 0.2056 - acc: 0.9291
 Epoch 2/100 - 11s - loss: 0.1194 - acc: 0.9579
 Epoch 3/100 - 11s - loss: 0.1082 - acc: 0.9633
 Epoch 4/100 - 11s - loss: 0.1014 - acc: 0.9661
 Epoch 5/100 - 10s - loss: 0.0946 - acc: 0.9684
 Epoch 6/100 - 11s - loss: 0.0926 - acc: 0.9693
 Epoch 7/100 - 11s - loss: 0.0916 - acc: 0.9697
 Epoch 8/100 - 11s - loss: 0.0894 - acc: 0.9711
 Epoch 9/100 - 11s - loss: 0.0868 - acc: 0.9715
 Epoch 10/100 - 11s - loss: 0.0863 - acc: 0.9721
 Epoch 11/100 - 11s - loss: 0.0861 - acc: 0.9705

 Epoch 89/100 - 10s - loss: 0.0571 - acc: 0.9816
 Epoch 90/100 - 10s - loss: 0.0582 - acc: 0.9812
 Epoch 91/100 - 10s - loss: 0.0554 - acc: 0.9820
 Epoch 92/100 - 10s - loss: 0.0583 - acc: 0.9810
 Epoch 93/100 - 10s - loss: 0.0557 - acc: 0.9818

Epoch 94/100 - 10s - loss: 0.0560 - acc: 0.9819
 Epoch 95/100 - 10s - loss: 0.0567 - acc: 0.9818
 Epoch 96/100 - 10s - loss: 0.0547 - acc: 0.9822
 Epoch 97/100 - 10s - loss: 0.0582 - acc: 0.9824
 Epoch 98/100 - 10s - loss: 0.0558 - acc: 0.9822
 Epoch 99/100 - 10s - loss: 0.0550 - acc: 0.9822
 Epoch 100/100 - 10s - loss: 0.0552 - acc: 0.9825
 ***** AVALIANDO O MODELO (5) *****
 Teste.... acc: 97.99%

***** TREINANDO O MODELO (6) *****
 Epoch 1/100 - 18s - loss: 0.1829 - acc: 0.9337
 Epoch 2/100 - 10s - loss: 0.1171 - acc: 0.9604
 Epoch 3/100 - 10s - loss: 0.1079 - acc: 0.9648
 Epoch 4/100 - 10s - loss: 0.0989 - acc: 0.9672
 Epoch 5/100 - 10s - loss: 0.0972 - acc: 0.9691
 Epoch 6/100 - 10s - loss: 0.0952 - acc: 0.9688
 Epoch 7/100 - 10s - loss: 0.0917 - acc: 0.9687
 Epoch 8/100 - 10s - loss: 0.0876 - acc: 0.9717
 Epoch 9/100 - 10s - loss: 0.0876 - acc: 0.9715
 Epoch 10/100 - 10s - loss: 0.0860 - acc: 0.9719
 Epoch 11/100 - 10s - loss: 0.0853 - acc: 0.9720

 Epoch 89/100 - 11s - loss: 0.0576 - acc: 0.9814
 Epoch 90/100 - 13s - loss: 0.0570 - acc: 0.9820
 Epoch 91/100 - 12s - loss: 0.0567 - acc: 0.9818
 Epoch 92/100 - 11s - loss: 0.0563 - acc: 0.9824
 Epoch 93/100 - 12s - loss: 0.0586 - acc: 0.9817
 Epoch 94/100 - 11s - loss: 0.0572 - acc: 0.9824
 Epoch 95/100 - 10s - loss: 0.0558 - acc: 0.9828
 Epoch 96/100 - 10s - loss: 0.0563 - acc: 0.9822
 Epoch 97/100 - 10s - loss: 0.0551 - acc: 0.9831
 Epoch 98/100 - 10s - loss: 0.0569 - acc: 0.9822
 Epoch 99/100 - 10s - loss: 0.0543 - acc: 0.9834
 Epoch 100/100 - 10s - loss: 0.0556 - acc: 0.9829
 ***** AVALIANDO O MODELO (6) *****
 Teste.... acc: 97.77%

***** TREINANDO O MODELO (7) *****
 Epoch 1/100 - 18s - loss: 0.1827 - acc: 0.9384
 Epoch 2/100 - 11s - loss: 0.1153 - acc: 0.9618
 Epoch 3/100 - 10s - loss: 0.1034 - acc: 0.9662
 Epoch 4/100 - 10s - loss: 0.0962 - acc: 0.9676
 Epoch 5/100 - 10s - loss: 0.0954 - acc: 0.9694
 Epoch 6/100 - 10s - loss: 0.0903 - acc: 0.9706
 Epoch 7/100 - 10s - loss: 0.0904 - acc: 0.9699
 Epoch 8/100 - 10s - loss: 0.0850 - acc: 0.9731
 Epoch 9/100 - 10s - loss: 0.0833 - acc: 0.9731
 Epoch 10/100 - 10s - loss: 0.0844 - acc: 0.9715
 Epoch 11/100 - 10s - loss: 0.0821 - acc: 0.9729

 Epoch 89/100 - 10s - loss: 0.0562 - acc: 0.9824
 Epoch 90/100 - 10s - loss: 0.0572 - acc: 0.9816
 Epoch 91/100 - 10s - loss: 0.0560 - acc: 0.9830
 Epoch 92/100 - 10s - loss: 0.0566 - acc: 0.9820
 Epoch 93/100 - 9s - loss: 0.0559 - acc: 0.9826
 Epoch 94/100 - 10s - loss: 0.0557 - acc: 0.9827
 Epoch 95/100 - 10s - loss: 0.0544 - acc: 0.9828
 Epoch 96/100 - 10s - loss: 0.0582 - acc: 0.9810
 Epoch 97/100 - 10s - loss: 0.0543 - acc: 0.9835

Epoch 98/100 - 10s - loss: 0.0571 - acc: 0.9822
 Epoch 99/100 - 10s - loss: 0.0541 - acc: 0.9834
 Epoch 100/100 - 10s - loss: 0.0551 - acc: 0.9827
 ***** AVALIANDO O MODELO (7) *****
 Teste.... acc: 96.90%

***** TREINANDO O MODELO (8) *****
 Epoch 1/100 - 18s - loss: 0.2641 - acc: 0.9363
 Epoch 2/100 - 10s - loss: 0.1372 - acc: 0.9606
 Epoch 3/100 - 10s - loss: 0.1149 - acc: 0.9641
 Epoch 4/100 - 10s - loss: 0.1051 - acc: 0.9673
 Epoch 5/100 - 10s - loss: 0.1009 - acc: 0.9685
 Epoch 6/100 - 10s - loss: 0.0966 - acc: 0.9694
 Epoch 7/100 - 10s - loss: 0.0937 - acc: 0.9697
 Epoch 8/100 - 10s - loss: 0.0922 - acc: 0.9702
 Epoch 9/100 - 10s - loss: 0.0883 - acc: 0.9720
 Epoch 10/100 - 10s - loss: 0.0875 - acc: 0.9719
 Epoch 11/100 - 10s - loss: 0.0832 - acc: 0.9732

 Epoch 89/100 - 10s - loss: 0.0559 - acc: 0.9822
 Epoch 90/100 - 10s - loss: 0.0563 - acc: 0.9818
 Epoch 91/100 - 10s - loss: 0.0549 - acc: 0.9833
 Epoch 92/100 - 10s - loss: 0.0541 - acc: 0.9830
 Epoch 93/100 - 10s - loss: 0.0553 - acc: 0.9816
 Epoch 94/100 - 10s - loss: 0.0530 - acc: 0.9828
 Epoch 95/100 - 10s - loss: 0.0546 - acc: 0.9833
 Epoch 96/100 - 10s - loss: 0.0569 - acc: 0.9816
 Epoch 97/100 - 10s - loss: 0.0541 - acc: 0.9818
 Epoch 98/100 - 10s - loss: 0.0543 - acc: 0.9835
 Epoch 99/100 - 10s - loss: 0.0551 - acc: 0.9818
 Epoch 100/100 - 10s - loss: 0.0524 - acc: 0.9837
 ***** AVALIANDO O MODELO (8) *****
 Teste.... acc: 97.39%

***** TREINANDO O MODELO (9) *****
 Epoch 1/100 - 18s - loss: 0.1809 - acc: 0.9322
 Epoch 2/100 - 10s - loss: 0.1161 - acc: 0.9612
 Epoch 3/100 - 10s - loss: 0.1017 - acc: 0.9668
 Epoch 4/100 - 10s - loss: 0.0957 - acc: 0.9683
 Epoch 5/100 - 10s - loss: 0.0917 - acc: 0.9706
 Epoch 6/100 - 10s - loss: 0.0898 - acc: 0.9708
 Epoch 7/100 - 10s - loss: 0.0878 - acc: 0.9712
 Epoch 8/100 - 10s - loss: 0.0864 - acc: 0.9728
 Epoch 9/100 - 10s - loss: 0.0845 - acc: 0.9731
 Epoch 10/100 - 10s - loss: 0.0836 - acc: 0.9729
 Epoch 11/100 - 10s - loss: 0.0818 - acc: 0.9733

 Epoch 89/100 - 10s - loss: 0.0553 - acc: 0.9821
 Epoch 90/100 - 10s - loss: 0.0552 - acc: 0.9827
 Epoch 91/100 - 10s - loss: 0.0562 - acc: 0.9823
 Epoch 92/100 - 10s - loss: 0.0550 - acc: 0.9830
 Epoch 93/100 - 10s - loss: 0.0560 - acc: 0.9825
 Epoch 94/100 - 10s - loss: 0.0534 - acc: 0.9829
 Epoch 95/100 - 10s - loss: 0.0556 - acc: 0.9825
 Epoch 96/100 - 10s - loss: 0.0558 - acc: 0.9834
 Epoch 97/100 - 10s - loss: 0.0523 - acc: 0.9838
 Epoch 98/100 - 10s - loss: 0.0531 - acc: 0.9838
 Epoch 99/100 - 11s - loss: 0.0536 - acc: 0.9836
 Epoch 100/100 - 11s - loss: 0.0547 - acc: 0.9827
 ***** AVALIANDO O MODELO (9) *****

Teste.... acc: 97.01%

***** TREINANDO O MODELO (10) *****

Epoch 1/100 - 20s - loss: 0.1935 - acc: 0.9244
 Epoch 2/100 - 10s - loss: 0.1197 - acc: 0.9580
 Epoch 3/100 - 10s - loss: 0.1037 - acc: 0.9662
 Epoch 4/100 - 10s - loss: 0.0987 - acc: 0.9681
 Epoch 5/100 - 10s - loss: 0.0948 - acc: 0.9685
 Epoch 6/100 - 10s - loss: 0.0901 - acc: 0.9708
 Epoch 7/100 - 10s - loss: 0.0872 - acc: 0.9716
 Epoch 8/100 - 10s - loss: 0.0863 - acc: 0.9721
 Epoch 9/100 - 10s - loss: 0.0834 - acc: 0.9740
 Epoch 10/100 - 10s - loss: 0.0828 - acc: 0.9732
 Epoch 11/100 - 10s - loss: 0.0835 - acc: 0.9728

... ..
 Epoch 89/100 - 10s - loss: 0.0540 - acc: 0.9828
 Epoch 90/100 - 10s - loss: 0.0537 - acc: 0.9832
 Epoch 91/100 - 10s - loss: 0.0575 - acc: 0.9817
 Epoch 92/100 - 10s - loss: 0.0543 - acc: 0.9824
 Epoch 93/100 - 10s - loss: 0.0542 - acc: 0.9830
 Epoch 94/100 - 10s - loss: 0.0544 - acc: 0.9825
 Epoch 95/100 - 10s - loss: 0.0527 - acc: 0.9838
 Epoch 96/100 - 10s - loss: 0.0521 - acc: 0.9839
 Epoch 97/100 - 10s - loss: 0.0546 - acc: 0.9831
 Epoch 98/100 - 10s - loss: 0.0525 - acc: 0.9830
 Epoch 99/100 - 10s - loss: 0.0523 - acc: 0.9835
 Epoch 100/100 - 10s - loss: 0.0537 - acc: 0.9827

***** AVALIANDO O MODELO (10) *****

Teste.... acc: 97.28%

97.41% (+/- 0.40%)

***** INICIANDO O MODELO COM UM NOVO DTASET *****

Carrega o modelo

>Modelo carregado C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof1.h5

Carrega o modelo

>Modelo carregado C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof2.h5

Carrega o modelo

>Modelo carregado C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof3.h5

Carrega o modelo

>Modelo carregado C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof4.h5

Carrega o modelo

>Modelo carregado C:\Users\suporte\Documents\Projeto_PYTHON\Save_model_le_prof\model_le_prof5.h5

Total modelos carregados...: 5

**** AVALIANDO A ACURÁCIA DE CADA MODELO SEPARADAMENTE *****

Acurácia do modelo: 0.983

Acurácia do modelo: 0.982

Acurácia do modelo: 0.983

Acurácia do modelo: 0.983

Acurácia do modelo: 0.985

Accuracia do teste: 0.986

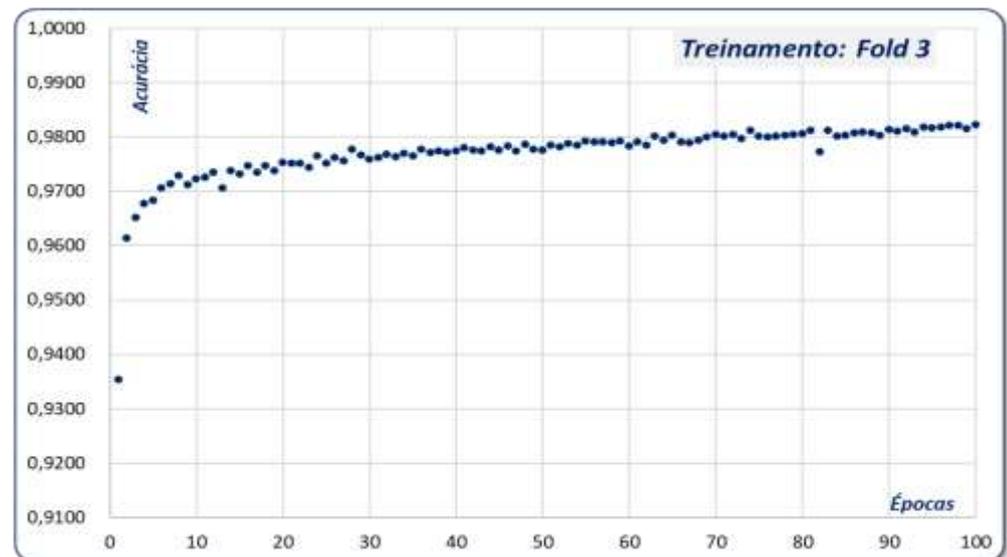
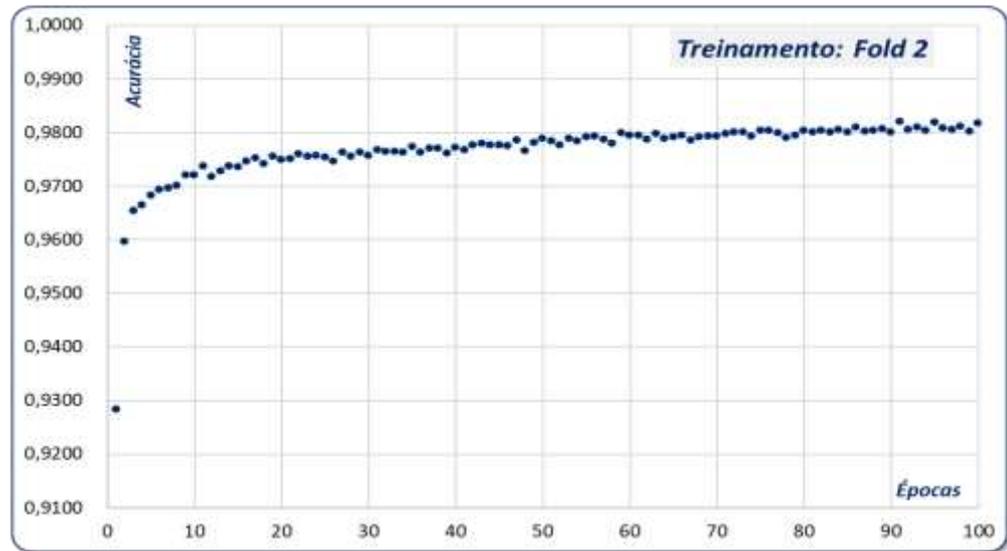
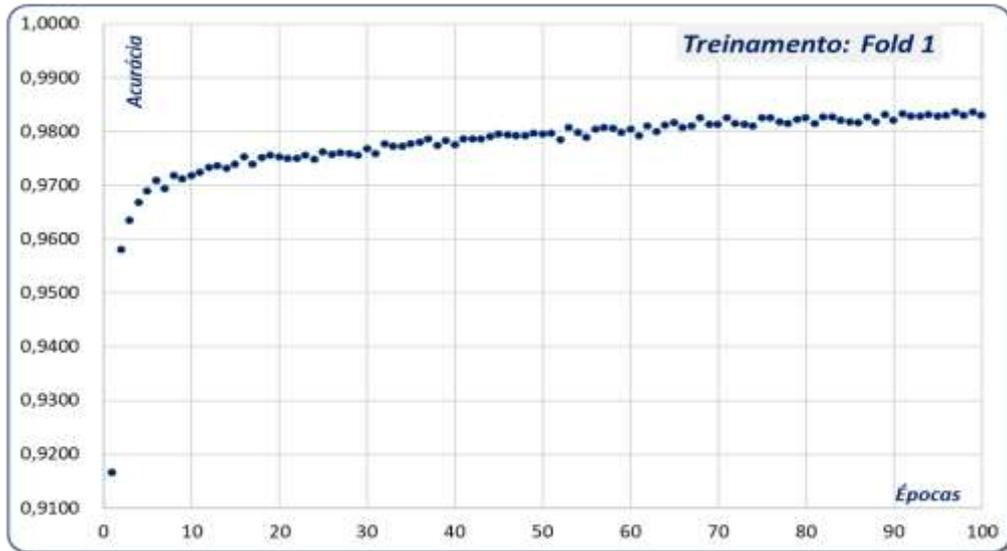
Modelo de regressão salvo

Duração: 0:00:27.979945

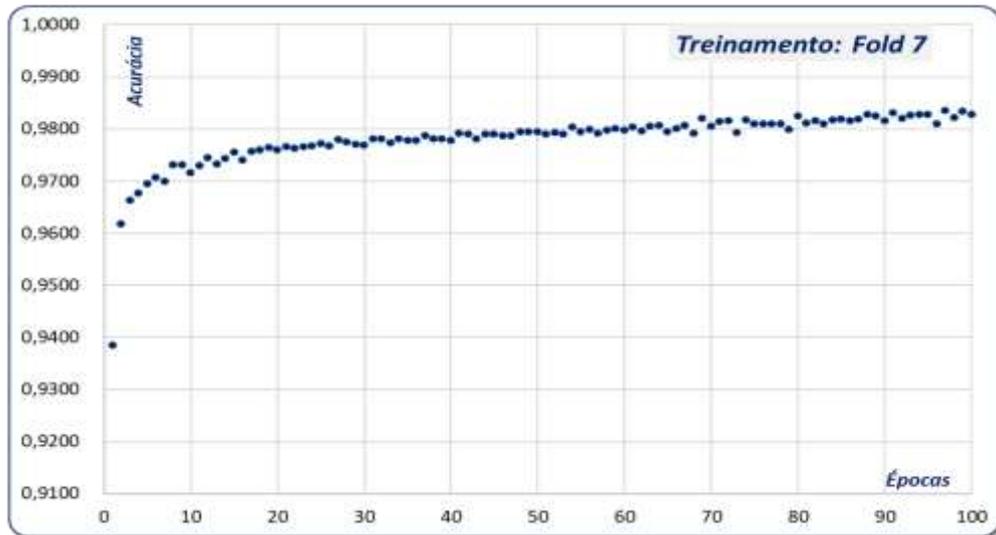
APÊNDICE VII – MATRIZ DE CORRELAÇÃO DAS VARIÁVEIS *FEATURES*

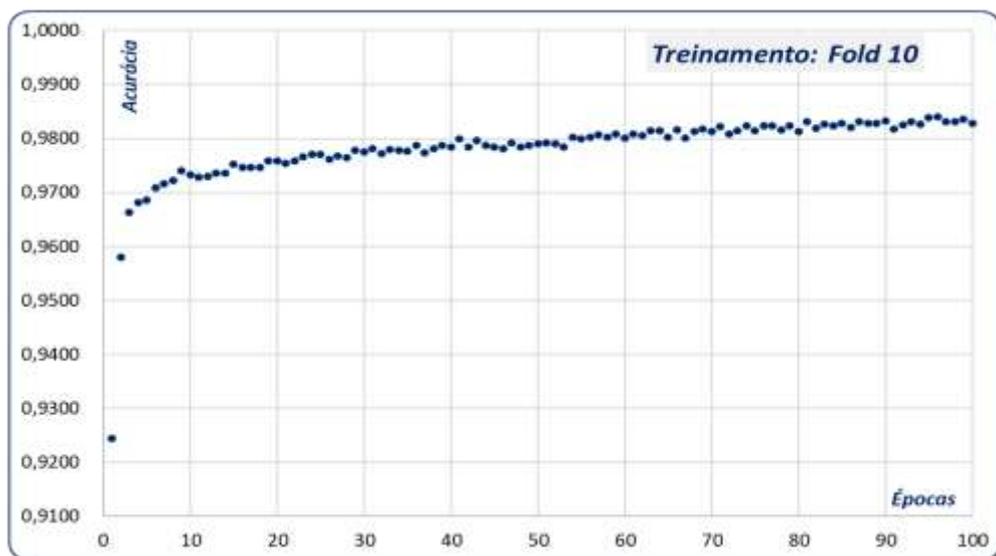
CORRELAÇÃO	HH_VIDA	SEXO	RAÇA	IDHM_MUN_RES	CD_REGIÃO	LOC_OCOR	IDADEMAE	ESCMAE	QT_FILVIVO	QT_FILMORT	GRAVIDEZ	GESTACAO	PARTO	PESO	ASSISTMED	APGAR1	APGARS	ESTCVIL	QTD_CONS
HORAS_VIDA		0,22816	0,07957	0,01906	0,02886	0,04810	0,04816	- 0,05788	- 0,05618	- 0,05829	- 0,15213	0,30542	0,02110	0,58139	- 0,05350	0,62382	0,53319	- 0,03142	0,27969
SEXO	0,22816		- 0,00478	0,00976	0,00184	0,00065	0,01402	- 0,02212	0,00534	0,01760	- 0,00244	- 0,00689	- 0,01554	0,00208	- 0,00765	0,02311	0,02698	0,00486	0,00395
RAÇA	0,07957	- 0,00478		0,00992	0,01729	- 0,05973	- 0,01883	- 0,08004	- 0,01230	0,00627	- 0,01997	- 0,02892	- 0,07130	0,09785	- 0,03867	0,05712	0,08403	- 0,02838	0,03208
IDHM_MUN_RES	0,01906	- 0,00976	0,00992		0,51512	- 0,05975	0,05766	0,07268	- 0,09108	0,00548	0,02117	- 0,05314	0,01707	- 0,04614	- 0,03026	- 0,00651	- 0,00937	- 0,08378	- 0,06172
CD_REGIÃO	0,02886	- 0,00184	0,01729	0,51512		- 0,05044	0,03050	0,02060	- 0,03601	- 0,01502	0,01803	- 0,05043	0,00293	- 0,03855	- 0,04095	- 0,00137	0,00057	0,01981	- 0,06949
LOC_OCOR	- 0,04810	0,00065	- 0,05973	0,05975	0,05044		- 0,02854	0,03227	0,04632	- 0,00726	- 0,01247	0,09547	0,09428	- 0,05891	0,31612	- 0,00470	- 0,04394	0,01866	- 0,04323
IDADEMAE	0,04816	0,01402	- 0,01883	0,05766	0,03050	- 0,02854		- 0,04844	0,39122	0,21506	0,03149	- 0,04583	- 0,04254	0,06579	- 0,03045	0,03914	0,04897	0,00732	0,06858
ESCMAE	- 0,05788	- 0,02212	- 0,08004	0,07268	0,02060	0,03227	- 0,04844		- 0,23359	- 0,08834	- 0,01862	0,31508	0,53704	- 0,22853	0,02251	- 0,03008	- 0,16618	- 0,05231	- 0,08824
QT_FILVIVO	- 0,05618	0,00534	- 0,01230	0,09108	0,03601	0,04632	0,39122	- 0,23359		0,21671	0,08485	- 0,09599	- 0,15908	0,00107	0,02704	- 0,03890	- 0,00939	0,11071	- 0,04796
QT_FILMORT	- 0,05829	0,01760	0,00627	0,00548	0,01502	- 0,00726	- 0,02854	0,03227	0,04632		0,03627	- 0,10526	- 0,07599	- 0,04288	- 0,00120	- 0,05952	- 0,03333	0,03119	0,00024
GRAVIDEZ	- 0,15213	- 0,00244	- 0,01997	0,02117	0,01803	- 0,01247	0,03149	- 0,01862	0,08485	0,03627		- 0,18385	- 0,03823	- 0,26187	- 0,00274	- 0,13448	- 0,12692	- 0,01266	- 0,09317
GESTACAO	0,30542	- 0,00689	- 0,02892	0,05314	0,05043	0,09547	- 0,04583	0,31508	- 0,09599	- 0,10526	- 0,18385		0,50834	0,29520	0,07185	0,29854	0,11657	0,02169	0,19155
PARTO	0,02110	- 0,01554	- 0,07130	0,01707	0,00293	0,09428	- 0,04254	0,53704	- 0,15908	0,07599	0,03823	0,50834		- 0,22385	0,07203	0,01797	- 0,16098	- 0,07260	- 0,05880
PESO	0,58139	0,00208	0,09785	0,04614	0,03855	- 0,05891	0,06579	- 0,22853	0,00107	- 0,04288	- 0,26187	0,29520	- 0,22385		- 0,05785	0,54601	0,49936	0,08540	0,46206
ASSISTMED	- 0,05350	- 0,00765	- 0,03867	0,03026	0,04095	0,31612	- 0,03045	0,02251	0,02704	0,00120	0,00274	0,07185	0,07203	- 0,05785		- 0,01648	- 0,03739	0,00850	- 0,01731
APGAR1	0,62382	0,02311	0,05712	0,00651	0,00137	0,00470	0,03914	- 0,03008	- 0,03890	0,05932	0,13448	0,29854	0,01797	0,54601	0,01648		0,73602	0,00037	0,27962
APGARS	0,53319	0,02698	0,08403	0,00937	0,00057	0,04394	0,04897	- 0,16618	0,00939	0,03333	0,12692	0,11657	- 0,16098	0,49936	0,03739	0,73602		0,00579	0,25098
ESTCVIL	- 0,03142	0,00486	- 0,02838	0,08378	0,01981	0,01866	0,00732	- 0,05231	0,11071	0,03119	0,01266	0,02169	- 0,07260	0,08540	0,00850	0,00037	0,00579		0,14388
QTD_CONS	0,27969	0,00395	0,03208	0,06172	0,06949	0,04323	0,06858	- 0,08824	0,04796	0,00024	- 0,09317	0,19155	- 0,05880	0,46206	- 0,01731	0,27962	0,25098	0,14388	

APÊNDICE VIII – DESEMPENHO DOS TREINAMENTOS ($K = 10$).



APÊNDICE VIII - DESEMPENHO DOS TREINAMENTOS ($K = 10$) Continua...

APÊNDICE VIII - DESEMPENHO DOS TREINAMENTOS ($K = 10$). Continua...

APÊNDICE VIII - DESEMPENHO DOS TREINAMENTOS ($K = 10$). Continua...

APÊNDICE IX – EXEMPLO DOS PESOS SINÁPTICOS OBTIDOS NO PROCESSO

		Pesos Sinápticos : DENSE									
		0	1	2	3	4	5	6	7	8	9
Features de Entrada	0	0,042863	-1,103604	-0,287059	-0,466819	0,253843	-0,728120	-0,409029	0,358168	0,575588	0,114529
	1	-0,494534	-0,375220	-0,202274	0,468450	0,065607	0,009266	-0,620935	-0,584590	-0,176908	-0,369471
	2	0,374428	0,021514	-0,057516	0,319282	0,444661	0,044307	-0,434998	0,076766	0,009256	-0,183062
	3	0,071184	0,090794	-0,024893	0,095269	0,446473	-0,136990	-0,564952	0,597222	0,293406	0,603630
	4	-0,049628	0,084757	-0,057929	0,303105	-0,591545	0,071579	0,431451	0,508352	0,210156	-0,481649
	5	1,291248	-1,309084	-0,619097	0,884449	0,693300	0,300862	0,804282	0,517254	-0,399057	-0,384681
	6	0,372319	0,474696	-0,111949	0,183790	-1,259657	-0,008944	-0,016075	0,087014	0,070013	0,595378
	7	-0,101723	0,053548	0,264728	-0,534860	0,042268	-0,152107	0,229700	0,004513	-0,264019	-0,124251
	8	-2,810811	0,509034	0,043581	0,523887	0,214046	0,008068	-0,315387	-1,635322	1,032517	-1,066496
	9	0,087882	-0,284756	0,038909	-0,998956	-0,659847	-0,003614	0,844906	-0,904091	-0,111982	0,861249
	10	-0,402414	-0,247400	0,107960	-0,896493	0,666614	-0,016897	0,503086	0,373706	-0,196080	0,663720
	11	-1,844517	-0,105479	0,160468	-0,797697	0,238447	-0,721254	-0,373887	0,237264	-0,507092	-0,965436
	12	0,842607	0,199795	0,055363	-3,936986	0,674696	-0,181873	1,856180	0,369099	-1,154453	-1,212708
	13	-0,846947	-0,024328	-0,214841	0,149384	-0,063159	0,236226	-0,175056	-0,517457	-1,381672	-0,363833
	14	0,000302	-0,008599	0,278293	0,291031	-0,062635	0,405906	0,243997	0,155720	-0,125656	-0,100771
	15	-0,130289	-0,446192	0,004972	-0,910106	1,668621	0,399716	-0,514996	-1,514350	0,938879	0,421179
	16	0,017811	-0,389264	-0,133524	0,130888	0,439455	0,158056	-0,166175	-0,101422	0,412641	0,459213
	17	0,360989	0,550730	-0,019001	-0,165300	-0,096556	-0,156880	0,037456	-0,099897	0,079392	0,602437
	18	-0,262187	0,473903	-0,102975	-1,369351	-1,249922	-0,001736	1,249787	-1,095625	0,207587	0,712174

		Pesos Sinápticos : DENSE									
		10	11	12	13	14	15	16	17	18	19
Features de Entrada	0	0,410687	0,942224	0,147538	-0,043144	0,174990	0,035713	-0,179643	0,544051	-0,846415	-0,720510
	1	-0,099912	0,017097	-0,418262	-0,479638	0,568475	-0,127318	-0,231073	-0,101448	0,118637	-0,046763
	2	0,059717	0,044116	0,257922	-0,347392	-0,761857	0,112672	0,178753	-0,148927	0,046436	0,253938
	3	0,072928	-0,106422	-1,158166	0,126116	0,716241	0,273229	-0,324512	0,294201	0,153878	-0,041216
	4	0,287314	0,059634	-0,746050	-0,230843	0,219532	-0,073108	-0,331391	-0,556148	-0,181295	0,310572
	5	1,035389	-0,077822	0,571115	-0,556973	0,162088	-1,592278	0,084307	-1,208655	-0,087482	0,145091
	6	-0,637300	0,176116	0,368608	-0,364689	-0,615529	-0,117078	-0,342310	-0,123182	0,240955	-0,326624
	7	-1,726656	0,141742	-0,536921	-0,143612	-0,423509	-0,340739	0,415817	-0,213166	-0,038521	-0,970308
	8	0,932110	-0,319004	-0,239530	-0,255649	-0,878585	-1,995036	0,091845	0,451298	-0,096799	-0,022147
	9	0,326629	0,305631	0,391211	0,764893	-1,227846	-1,047392	-1,124639	-0,829814	-0,742339	0,016951
	10	-1,022268	0,085635	0,164641	-0,647447	-1,200267	0,141196	-0,380558	0,032841	-0,247069	-0,169694
	11	-0,495051	0,405854	0,305434	-1,318498	-0,066698	-0,099998	0,058995	-0,415947	-0,956659	0,770445
	12	0,551791	-0,449515	2,035974	1,096808	1,665170	-0,149651	-1,076993	0,018297	0,534211	0,942125
	13	-0,062086	0,069967	0,469706	0,905344	-0,283066	0,565403	0,692797	0,693388	0,179500	0,269443
	14	-0,105132	-0,120161	0,124991	0,121071	-0,116763	-0,140226	0,270059	0,085073	0,539637	0,098185
	15	-1,048263	0,213128	-0,734933	-1,063411	-1,725841	1,987690	-0,848518	1,983932	0,446199	-0,573709
	16	-0,425825	-0,037135	-1,516810	-0,229685	-0,662544	1,661686	-0,465327	1,043579	0,471475	-0,295946
	17	-2,023115	-0,311226	0,146990	-0,128404	-1,104938	0,868493	-0,681055	0,251351	0,199055	0,025147
	18	-0,020556	0,251975	-0,997975	-0,045437	-0,781149	0,470664	-0,442663	0,011584	-0,151106	-0,357220

		Pesos Sinápticos : DENSE									
		20	21	22	23	24	25	26	27	28	29
Features de Entrada	0	-0,419468	0,200527	-0,569838	0,013584	0,937085	-0,143646	0,115651	0,775775	0,356542	-0,622068
	1	-0,108399	-0,078726	-0,033663	-1,084095	0,163092	0,475543	-0,123506	-0,241889	-0,041872	-0,188895
	2	0,026393	0,096983	0,398376	-2,306381	-0,084005	-0,649713	0,498756	0,311084	-0,216347	0,660542
	3	-0,251479	-1,162276	-0,644806	-0,060666	0,255437	-0,437946	0,598016	-0,068838	-0,482012	-0,359561
	4	-0,194679	0,688963	-0,062934	-0,430583	0,167908	-0,665646	0,042091	-0,010984	0,520223	-0,008322
	5	0,185283	0,513368	0,320383	1,577245	-0,254449	1,069227	-1,557307	-0,002388	-0,375079	-0,243445
	6	0,127975	-0,305754	0,368283	0,197633	0,153785	0,352865	-0,439323	0,162754	-0,243151	-0,131488
	7	0,121416	-1,015555	0,988481	-1,166941	-0,742245	-0,156965	-0,304022	-0,045781	-1,735722	0,005717
	8	-0,017855	-0,786171	-1,877858	0,977292	-0,039608	-0,185783	-0,279685	0,145525	0,376306	0,141460
	9	-0,017992	1,141674	0,228720	0,169810	-0,203286	1,011480	0,606196	0,105939	-0,401968	-0,038717
	10	0,201501	0,605346	-0,000133	-0,414982	0,365815	-0,416764	-0,472163	0,096744	0,285554	0,121371
	11	-0,256069	0,289544	-0,662992	-0,095545	-0,378638	0,046925	0,068434	-0,221755	0,971752	-0,922623
	12	-0,269722	0,123026	0,638528	-2,519319	-0,127443	2,411976	0,769259	0,017447	-1,285766	0,631365
	13	-0,259503	0,074053	0,170670	-0,071263	0,368805	-0,242421	0,250547	0,159436	0,039992	0,215661
	14	0,035879	-0,236350	0,241162	-0,019440	0,001199	-0,162568	-0,156199	0,033706	0,151981	0,138303
	15	0,175751	0,264231	1,003747	0,402931	-0,395035	-0,499897	1,755209	0,760925	-0,599358	0,219066
	16	0,067968	0,219285	0,457567	-0,025899	-0,232654	-0,548438	1,107633	0,646911	-0,782020	0,231035
	17	0,349823	0,438637	0,223694	0,621942	-0,299888	0,562719	-1,696238	0,639098	0,074758	-0,295558
	18	-0,066164	-0,940556	0,172656	1,031221	-0,357925	-0,632314	-0,875693	0,199138	-0,143535	-0,229262

APÊNDICE IX (Continuação...)

		Pesos Sinápticos : DENSE									
		30	31	32	33	34	35	36	37	38	39
Features de Entrada	0	-0,016612	0,128898	0,196772	0,026801	-0,087988	-0,728995	-0,144229	0,755172	-0,841382	0,281938
	1	0,288981	0,024532	0,137012	-0,209590	-0,040250	-0,218036	0,081188	-0,347441	0,065012	0,003801
	2	-0,640016	0,577505	-0,273138	-0,026257	-0,076652	-0,339922	-0,594147	0,197695	0,132024	0,105857
	3	0,544056	0,349439	-0,601148	0,138434	-0,109473	0,104666	0,182391	-0,203389	-0,059715	0,028992
	4	0,549053	-0,619889	0,231888	-0,174377	-0,036628	0,362269	-0,823765	-0,207759	0,011641	-0,152776
	5	-0,031291	-0,700205	-0,093498	0,041273	-0,042064	-0,325330	0,742127	-0,382659	-0,137103	0,087121
	6	0,233559	0,128990	-1,252146	-0,074254	-0,168177	-0,016184	-0,144350	-0,093722	0,432677	0,038656
	7	-1,052809	-0,097081	-0,331774	-0,160491	-0,192174	0,315035	-0,685297	0,151334	-0,407795	0,044973
	8	0,384539	0,454333	0,388208	0,062571	-0,167852	-0,579083	0,501662	0,030412	-0,583620	-0,142861
	9	1,695635	-0,363002	0,944497	-0,162256	0,194872	1,347426	0,274149	0,253255	0,217908	-0,082463
	10	0,522992	-0,253908	1,005133	-0,036602	0,063801	-1,086083	-1,599540	-0,511085	-0,210690	0,122472
	11	-0,186099	-0,372059	-0,526785	-0,216722	-0,210728	0,494237	0,520718	0,079420	0,835377	-0,204181
	12	0,865755	-0,766319	-0,696236	-0,107196	0,195543	-0,642565	1,772144	-0,440097	0,274203	-0,372684
	13	-0,055261	0,606346	0,731232	-0,210641	0,058962	0,317047	0,165736	-0,160364	-0,146879	-0,041524
	14	-0,029523	0,068686	0,138992	0,054027	0,030821	0,148342	-0,041664	-0,192133	-0,129399	-0,088350
	15	-1,366131	1,034883	-2,364136	-0,199447	0,135735	-1,101646	-0,722103	-0,693046	-0,430180	-0,014427
	16	-1,615713	1,362370	-0,920211	0,004460	-0,066587	-0,620039	-0,765415	-0,348688	-0,072258	0,202276
	17	-0,678341	0,395008	-0,731378	-0,151084	0,022222	0,331934	0,265654	-0,373037	-0,291402	0,183738
18	-0,182665	-0,663045	0,556800	0,013054	0,005209	-0,807295	0,337173	-0,448292	0,319838	-0,168506	

		Pesos Sinápticos : DENSE									
		40	41	42	43	44	45	46	47	48	49
Features de Entrada	0	-0,070931	0,153580	0,404954	-0,172296	0,579143	-0,959644	-0,124926	-0,376919	0,004068	-0,170602
	1	-0,863906	0,060748	-0,080562	0,209039	0,181132	0,147170	0,287228	-0,025075	0,253064	0,270304
	2	0,014776	0,410232	0,073121	-0,121978	-0,438490	0,451741	0,412383	-0,351070	-1,265148	0,104754
	3	-0,366623	0,374085	0,017916	0,297296	0,454741	0,275975	-1,010617	-0,155332	-0,226286	-0,446956
	4	-0,512949	-0,202119	0,005467	-0,122808	0,029260	-0,064510	0,059565	0,832433	0,220503	0,222892
	5	1,157632	0,570220	-0,367402	-2,687004	0,293652	0,356447	-1,353834	-0,479150	-3,524058	0,059808
	6	1,082710	0,423612	0,073747	0,277272	0,354302	-0,114278	0,747761	0,102721	-0,543477	0,576629
	7	0,588097	0,496030	0,421959	1,356735	-0,497163	-0,107870	0,384092	0,808819	-0,111776	-0,013310
	8	0,228512	0,263555	-1,351283	0,852429	0,308243	-0,335342	0,459613	-0,289174	0,331385	-0,292063
	9	0,476067	0,251987	0,398859	-0,287010	-0,250718	0,485144	-0,229462	0,503348	0,359365	0,116880
	10	0,583447	-1,206589	-0,132950	-0,463007	0,207112	-0,483975	-0,353123	0,422371	0,542417	0,158977
	11	0,607518	0,042546	-0,420619	-1,015623	-0,422469	0,717475	-0,049679	-0,814383	0,153361	-0,963182
	12	1,038696	1,179853	1,131984	-1,028583	-1,632321	-2,362716	-0,373028	-0,425556	1,217753	-0,780387
	13	-0,692035	-0,240316	-1,573678	-0,061961	-0,529571	-0,021208	0,072506	-0,639325	0,271942	-0,670187
	14	0,279347	0,046617	0,024528	0,121498	0,074499	-0,026043	-0,008506	0,121045	0,109982	0,150626
	15	-0,510574	1,636212	-2,683697	-0,054929	2,011038	0,097099	1,194511	0,007558	1,315743	-0,836614
	16	-1,637459	0,430058	-0,657325	-0,601894	0,529924	0,805007	0,693217	-0,306274	1,293802	0,035646
	17	-0,501830	-0,614192	0,340745	-1,127904	-0,078389	0,289237	0,687872	-0,000616	1,028327	-0,056829
18	0,126534	-1,508742	0,172777	0,156535	0,613551	-0,046127	-1,182626	0,041771	0,491247	0,133169	

		Pesos Sinápticos : DENSE									
		50	51	52	53	54	55	56	57	58	59
Features de Entrada	0	-0,508752	0,293552	-1,096655	0,269958	-0,015633	0,289528	-0,400781	-0,644782	0,456398	0,172953
	1	0,497284	-0,892416	0,002493	0,125504	-1,243098	-0,235271	0,021431	-0,298535	-0,009312	-0,152434
	2	0,303691	0,045141	-0,010794	-0,298920	-0,132937	-0,277905	-0,467778	-0,471245	-0,017646	-0,091539
	3	-0,615278	0,316452	-0,127048	-0,221614	0,777975	-0,194920	0,293315	0,559358	-0,085244	-0,132158
	4	0,605001	-0,734631	0,142823	-0,214787	-0,386513	-0,222456	-0,651927	-0,328217	-0,293559	-0,004821
	5	-0,705985	-0,574047	0,261082	-0,247532	1,077886	0,782252	-0,649249	-1,805824	-0,030220	0,183529
	6	0,364668	-1,150417	0,125412	-0,298392	0,088483	0,672173	0,301819	0,320704	-0,019777	0,095922
	7	0,293965	-0,735422	0,236487	0,579401	0,632213	0,111122	0,500316	0,363905	-0,051935	-0,208122
	8	0,140941	1,209117	0,064749	0,226604	-0,140233	-0,202409	-0,060928	0,233350	0,044819	-0,265031
	9	-0,319430	0,529847	0,115876	-1,816598	-1,038445	0,806092	-0,203938	-0,577660	0,017097	-0,281603
	10	-1,376504	-0,108357	-0,052223	0,550720	-0,667483	1,030074	0,558240	-0,074134	-0,001180	0,267113
	11	-1,144294	-0,213373	0,594253	1,083211	-1,211757	0,188554	0,871975	0,855545	-0,035180	-0,213260
	12	-1,494245	0,331280	1,213428	1,167362	-1,285240	1,195779	0,744927	-1,421597	0,099440	-0,111235
	13	-0,264163	-0,253914	0,071384	-0,613738	-0,016899	0,120505	-0,128779	-0,209961	-0,001956	0,109652
	14	0,000476	0,178948	0,244898	0,163640	0,140826	-0,136081	-0,104468	-0,018322	-0,179520	-0,001116
	15	-1,150840	-1,297912	-0,484985	-0,899100	-0,828549	-0,307569	-1,006832	-0,825693	0,095780	0,291893
	16	-0,259207	-0,590774	-0,492998	-0,597968	0,030635	-0,423785	-0,512452	-0,199320	0,100219	0,220853
	17	-0,684513	-3,105706	0,143570	-0,731543	-0,053638	-0,293547	0,616067	-0,232415	0,109746	0,157371
18	0,301697	0,790559	0,172485	-1,105458	-0,467498	0,688448	-1,353669	-0,230006	0,071971	-0,011707	

APÊNDICE IX (Continuação...)

		Pesos Sinápticos : DENSE									
		60	61	62	63	64	65	66	67	68	69
Features de Entrada	0	-0,099128	-0,630947	-0,557636	-0,351325	-0,102586	-0,011873	-0,292464	-0,962174	0,415585	-0,170980
	1	-0,401781	-0,256132	-0,616357	0,179405	-0,053155	-0,035774	-0,642500	-0,019314	-1,317693	0,171759
	2	-0,478698	0,514205	0,290120	0,559153	0,486000	0,021882	0,163423	0,305292	0,126980	-0,785590
	3	0,525946	-0,500353	0,183853	-0,048243	0,069396	-0,112237	0,438362	0,239109	-0,082947	0,506448
	4	-0,335076	0,376874	-0,115946	-0,361513	0,424611	-0,169321	-0,171473	-0,065390	-0,108061	0,492911
	5	1,010800	0,085062	0,808373	0,165307	-2,032401	0,130360	0,739776	-0,367883	-0,444246	0,448569
	6	0,344481	0,022772	-0,293653	0,853815	-0,785248	-0,105550	-0,408213	0,079587	0,096216	-0,701408
	7	0,266393	0,342039	-0,345441	-1,175299	-0,397622	-0,160355	0,920378	-0,135729	0,544472	-1,236058
	8	0,766169	0,437164	0,459622	-0,707084	-0,153642	-0,235230	0,352457	-0,362626	-0,780208	0,253514
	9	0,853309	-0,754556	0,923398	0,477578	1,547918	0,165176	-0,145950	-0,173084	0,777434	0,482744
	10	0,058089	-0,144272	-0,456089	0,628466	0,485963	-0,195798	0,142472	0,619451	-0,386223	0,087307
	11	-0,486992	-0,323050	0,646841	0,710098	-0,074506	0,010252	-0,465957	0,725375	0,255560	0,665944
	12	-0,947234	-1,072582	-0,038148	-0,106367	-0,239647	0,051505	1,430659	-0,835640	0,078939	0,318324
	13	0,143674	0,251519	0,436828	-0,340053	0,341856	-0,058831	-0,367448	-0,107979	-0,491077	-0,364169
	14	-0,171575	0,233291	0,334394	0,051807	-0,043598	0,097211	0,288171	0,365869	-0,045047	0,083367
	15	2,012425	-0,343281	-1,700498	-0,327100	1,313926	-0,137523	0,460281	0,199597	-0,904699	-2,080634
	16	0,678061	-0,415405	-1,272998	-0,184642	0,846676	0,115824	0,045771	0,166769	-0,843814	-0,460932
	17	0,778616	-0,446558	-1,563080	-0,830523	-0,146246	0,058425	-0,633820	-0,040921	-3,651137	0,077508
18	0,685841	-0,850494	0,172715	-0,677337	0,792023	-0,143153	0,158117	0,053396	0,174155	0,691352	

		Pesos Sinápticos : DENSE									
		70	71	72	73	74	75	76	77	78	79
Features de Entrada	0	-0,043420	-0,080741	-0,136721	-0,071431	0,316958	-0,188783	-0,839930	-0,352972	0,076073	-0,059109
	1	0,003254	0,468480	-0,772753	-0,097418	-0,332870	-0,170912	0,001041	0,048841	0,345231	0,204061
	2	-0,112662	-0,401762	0,544213	-0,165393	0,088720	-0,146706	-0,161607	-1,017391	-0,135236	-0,371486
	3	0,084166	-0,073749	-0,415409	-0,394060	-0,912935	-0,001424	0,006498	-0,863348	0,519375	-0,509981
	4	-0,007323	0,272283	-0,616058	0,037864	0,198255	-0,149043	0,006959	-0,004090	-0,798274	0,248643
	5	-0,190740	0,427952	-1,219177	1,147810	-0,885822	0,136408	-0,178632	-1,793647	-1,242767	0,972604
	6	-0,270618	-1,071955	-0,068957	0,279629	-0,524600	0,048307	-0,188194	-0,926575	0,053158	-0,971891
	7	-0,254760	0,332407	0,974639	0,262050	-0,453108	-0,006027	0,021135	0,698658	0,449270	0,927994
	8	-0,171697	0,908460	0,931086	0,780224	-1,323105	-0,035977	0,464128	0,275891	0,794451	0,127325
	9	0,102195	-0,801099	-0,004040	0,974756	0,806617	0,066901	0,376410	0,940766	-0,069389	1,095400
	10	0,006267	0,666009	0,112221	0,028449	-0,209908	0,160734	0,039084	1,240307	-0,120563	1,309111
	11	-0,011880	-0,270251	-0,177386	-0,445189	-0,379664	-0,013418	0,211389	0,599015	-0,494647	0,445760
	12	-0,269257	0,1016470	0,022262	0,945765	1,354973	0,051519	-0,811305	-1,613683	-0,365838	0,680830
	13	0,167339	-0,766922	-0,220078	-0,624584	0,658630	-0,156483	-0,238487	0,228947	-0,794813	-0,806843
	14	0,083910	0,192347	0,247270	-0,101835	-0,019458	0,000140	0,401769	-0,001331	-0,014853	0,154224
	15	-0,030039	0,283353	-0,880444	-0,036672	1,055717	-0,171147	1,132979	0,363896	1,265693	-1,309279
	16	0,004549	-0,195232	-0,812605	0,020512	0,512533	0,010419	0,304313	0,755176	0,094470	-0,281163
	17	0,166991	0,204972	0,671798	0,705383	0,801555	0,002039	0,221494	0,219446	-1,318329	-0,085662
18	-0,174486	-0,367834	-0,031053	-0,402185	-0,631911	-0,231766	0,052963	0,294255	-0,110410	0,270048	

		Pesos Sinápticos : DENSE									
		80	81	82	83	84	85	86	87	88	89
Features de Entrada	0	-1,005254	0,696623	0,955302	0,827317	-0,143386	-0,844557	0,749322	0,384300	-1,017035	-0,758288
	1	0,002493	0,122448	-0,002474	0,334198	0,031389	0,151921	-0,101094	0,097563	-0,075121	0,203422
	2	0,057656	0,043021	0,332153	0,112741	-0,207847	-0,244535	0,076533	-0,505338	0,081572	-0,221784
	3	-0,024005	-0,206992	-0,067338	-0,122399	0,076077	-0,226146	-0,099551	-1,524803	-0,003415	-0,241659
	4	0,035636	-0,071643	-0,089207	0,327241	-0,276950	0,657073	-0,028727	0,299591	0,033595	0,057315
	5	-0,190895	-1,745298	0,000308	0,089762	-0,012022	-0,323236	0,373040	-1,520119	-0,236324	-0,531637
	6	0,118478	0,189618	0,450263	0,216946	0,013679	-0,026733	0,336525	0,704098	0,057885	0,190972
	7	-0,157221	-0,129249	-0,024105	-0,109221	-0,144680	0,150790	-0,186107	-0,331892	-0,179221	0,588852
	8	-0,023874	-0,181608	-0,206472	-0,490348	0,135722	0,116037	-0,016234	0,425820	-0,002034	-0,588539
	9	0,090980	0,238590	-0,113963	-1,341998	-0,056415	-0,555527	-0,016045	0,998664	0,098958	0,008800
	10	0,159483	-0,060992	0,285355	0,051368	-0,097163	-0,435038	0,085094	-0,434689	0,338696	0,543296
	11	0,146089	-0,212572	0,035079	0,113520	0,117405	0,418972	-0,243210	-0,120153	0,370807	1,266958
	12	-0,312394	0,558974	-0,064506	-1,159002	0,036229	1,026291	0,013914	0,581053	-0,180621	-1,570849
	13	0,056046	0,201989	-0,115361	-0,311595	0,114848	0,454824	0,268385	-0,013563	0,023835	0,220852
	14	0,212461	-0,171301	-0,187774	0,324571	-0,221566	0,100675	0,042100	-0,058347	0,314783	0,011944
	15	0,605057	0,407575	-0,159363	-2,150085	-0,228493	-2,139764	0,652179	0,023805	0,230253	-0,742920
	16	0,410865	0,351550	-0,145668	-1,219547	0,121393	-0,870299	0,460026	0,079018	0,052946	-0,583696
	17	-0,035593	-0,086348	-0,109044	-0,126152	-0,003843	-0,958978	0,317049	0,331151	0,021826	-0,752477
18	0,221312	0,160071	-0,201546	-0,043569	-0,283562	0,193613	0,342580	-0,165908	0,033934	-0,134958	

APÊNDICE IX (Continuação...)

		Pesos Sinápticos : DENSE									
		90	91	92	93	94	95	96	97	98	99
Features de Entrada	0	-0,077965	-0,439084	-0,576250	0,409522	0,491775	0,009763	0,522283	-0,221208	-0,645332	-0,489897
	1	0,637115	0,296121	-0,031274	0,891460	-0,012243	-0,294987	0,004552	-0,103784	0,136292	0,041815
	2	-0,299263	-0,033273	0,428454	-0,127729	-0,160752	0,632343	-0,029812	0,099151	0,331098	-0,155878
	3	-0,249851	-0,288016	0,175688	-0,069818	-0,204777	-0,467661	-0,023102	0,026143	-0,163943	-0,156823
	4	-0,228581	-0,009372	-0,214744	-1,239168	0,048918	-0,055535	0,021630	0,045074	-0,156051	-0,260154
	5	-1,309693	-1,582090	-0,423067	0,305278	-0,168926	-1,615902	-0,012329	-0,047997	0,635027	0,282020
	6	-0,509554	-0,020917	-0,063126	-0,584397	-0,170382	0,731358	0,032133	-0,186948	0,439353	-0,622186
	7	0,700859	0,367861	-0,150108	-0,024769	-0,168325	-0,176164	0,036926	0,026957	0,144937	0,380379
	8	0,597462	-0,908054	0,004783	-0,858826	0,080897	1,370002	-0,004943	0,065997	-0,111941	0,690564
	9	1,222249	-1,834715	-0,065361	-1,011341	-0,365578	-0,223455	-0,024373	-0,014967	0,246573	0,287777
	10	1,031991	-0,146845	0,177929	-1,093347	-0,010835	-0,381531	0,043739	-0,078620	0,042786	-0,205840
	11	-0,909399	-0,116320	-0,214697	-0,239578	-0,065553	-0,259121	0,039276	-0,077524	-0,950224	1,008107
	12	-2,034990	0,635986	0,544190	0,087625	0,360630	1,102135	-0,143290	-0,320071	0,025186	-0,567362
	13	-0,289317	0,293668	0,335583	0,071457	0,008403	0,358658	-0,002954	0,047640	-0,170958	-0,539860
	14	0,269185	-0,083597	0,097502	0,291429	0,090245	0,114313	-0,269295	-0,109279	0,143367	0,111233
	15	0,934430	0,393990	-0,498078	-1,095280	0,158397	0,133788	0,243239	-0,167760	0,050298	-0,280538
	16	0,534513	0,276476	-0,336340	-0,127028	0,345269	0,563985	0,684961	0,063742	-0,123472	0,216310
	17	0,579193	1,317244	-1,349052	0,108195	0,092829	0,487376	-4,517613	0,004993	-0,097160	0,484689
	18	0,032836	0,346591	-0,023984	-0,919107	-0,591243	-0,504867	0,019742	0,044814	-0,643059	-0,183480