

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação

Predição de dados em inquéritos populacionais utilizando
técnicas de aprendizado de máquina

Leonardo Zanini Ferreira

Pelotas, 2018

Leonardo Zanini Ferreira

Predição de dados em inquéritos populacionais utilizando técnicas de aprendizado de máquina

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação

Orientador: Prof. Dr. Ricardo Matsumura Araújo
Co-Orientadora: Prof^a. Dr^a. Ana Marilza Pernas

Pelotas, 2018

F383p Ferreira, Leonardo Zanini

Predição de dados em inquéritos populacionais utilizando técnicas de aprendizado de máquina. / Leonardo Zanini Ferreira ; Ricardo Matsumara Araújo, orientador. – Pelotas, 2018.

87 f. : il.

Dissertação (mestrado) – Universidade Federal de Pelotas ; Programa de Pós-Graduação em Computação, 2018.

1. Computação 2. Inteligência artificial. I. Araújo, Ricardo Matsumara II. Título.
CDD 004

AGRADECIMENTOS

Agradeço primeiramente a minha família pela educação, amor, incentivo e oportunidade de ter ingressado e cursado este mestrado. Também gostaria de agradecer aos meus orientadores, Ricardo Matsumura Araújo e Ana Marilza Pernas, pelo apoio, disponibilidade, e por terem aceito e me dado a oportunidade de estudar um tema do meu interesse.

Agradeço a minha equipe do Centro Internacional de Equidade em Saúde pela compreensão, apoio e também pela oportunidade de utilizar os dados. Agradeço ao Fernando Pires Hartwig por todo o suporte e disposição para auxiliar e discutir sobre os problemas enfrentados durante as análises, e a Cíntia Borges pelo apoio no design e plena disponibilidade.

Agradeço também a todos os meus amigos que ofereceram palavras de apoio e incentivo ao longo desta jornada.

RESUMO

FERREIRA, Leonardo Zanini. **Predição de dados em inquéritos populacionais utilizando técnicas de aprendizado de máquina.** 2018. 87 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2018.

O monitoramento do desenvolvimento dos indicadores de saúde e nutrição dos países depende de dados precisos e completos. Países de baixa e média renda dependem de inquéritos populacionais para tomada de decisões, já que seus sistemas de informação ainda são precários. Estes indicadores são desfechos de saúde que indicam a situação de uma intervenção, comportamento ou cuidado de saúde materno-infantil. Apesar de serem ótimas fontes de informação, os inquéritos populacionais também possuem informações incompletas que podem inviabilizar a construção destes indicadores de forma padronizada. O aprendizado de máquina surge como uma alternativa para completar estas lacunas existentes nos inquéritos com base em características sociodemográficas e preditores relacionados. Este trabalho avalia o desempenho de quatro técnicas de aprendizado de máquina (redes neurais, árvores de decisão, KNN e XGB) nesta tarefa, observando o comportamento de dois desfechos de saúde em quatro amostras diferentes. Os resultados apontam para a existência de uma grande variação no desempenho de acordo com o desfecho selecionado. A disponibilidade dos mesmos indicadores como preditores também é um desafio enfrentado. As técnicas de aprendizado de máquina estudadas apresentaram pouca variação de desempenho entre elas, o que sugere que a natureza dos dados é um fator determinante neste tipo de aplicação. Contudo, o primeiro desfecho apresentou resultados aceitáveis para uma aplicação prática, enquanto o segundo desfecho atingiu estimativas inferiores ao que especialistas consideram como um desempenho razoável.

Palavras-chave: aprendizado de máquina, desfechos de saúde, inquéritos populacionais de saúde.

ABSTRACT

FERREIRA, Leonardo Zanini. **Data prediction in populational surveys using machine learning techniques**. 2018. 87 p. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2018.

Monitoring the development of countries' health and nutrition indicators rely on accurate and complete data. Low-and middle-income countries depend on populational surveys for decision-making since their information systems are still unreliable. These indicators are health outcomes which point out the status of a maternal and child health intervention, behavior or care. Despite being great sources of information, the populational surveys also lack information which could prevent the creation of these standardized indicators. Machine learning arise as an alternative to fill the gaps found in the surveys based on sociodemographic characteristics and related predictors. This work evaluates the performance of four machine learning techniques (neural networks, decision trees, KNN and XGB) in this task by observing the behavior of two health outcomes in four different samples. The results indicate that a great variation can be found depending on the selected outcome. The availability of those indicators as predictors is also a challenge. The studied machine learning techniques presented a small performance variation between themselves which suggests that the nature of the data is determinant to this kind of application. Still, the first outcomes showed acceptable results for a practical application while the second outcome reached lower estimates compared to those which specialists consider as a reasonable performance.

Keywords: machine learning, health outcomes, populational health surveys.

LISTA DE FIGURAS

Figura 1 – Rede Neural Artificial com uma camada oculta de três neurônios	25
Figura 2 – Árvore de decisão baseada no banco de dados de plantas Iris	27
Figura 3 – Demonstração do KNN com dois valores para K.....	29
Figura 4 – Fluxograma do processo de mineração de dados.....	37
Figura 5 – Atributo país no formato original (acima) e no formato dummy (abaixo)	40
Figura 6 – Histograma da frequência dos inquéritos DHS ao longo do tempo....	46
Figura 7 – Pôster apresentado no Congresso Internacional de Epidemiologia ...	86

LISTA DE TABELAS

Tabela 1 – Atributos utilizados nas análises de AQP	46
Tabela 2 – Parâmetros disponíveis e valores atribuídos para Redes Neurais Artificiais	48
Tabela 3 – Parâmetros disponíveis e valores atribuídos para Árvores de Decisão	48
Tabela 4 – Parâmetros disponíveis e valores atribuídos para K-Vizinhos Mais Próximos	49
Tabela 5 – Parâmetros disponíveis e valores atribuídos para Otimização Extrema de Gradiente	49
Tabela 6 – Concordância e concordância esperada de cada amostra para AQP.	51
Tabela 7 – Atributos utilizados nas análises de AQP	51
Tabela 8 – Distribuição das classes do desfecho de AQP.....	52
Tabela 9 – Resultados para o desfecho AQP na amostra exploratória	53
Tabela 10 - Resultados para o desfecho AQP na amostra principal	54
Tabela 11 - Resultados para o desfecho de AQP na amostra específica	55
Tabela 12 - Resultados para o desfecho de AQP na amostra completa	56
Tabela 13 - Atributos utilizados nas análises de Déficit de Altura por Idade	58
Tabela 14 – Distribuição das classes do desfecho de Déficit de Altura por Idade	60
Tabela 15 – Parâmetros de SMOTE testados na amostra exploratória para Déficit de Altura por Idade.....	60
Tabela 16 - Resultados para o desfecho de Déficit de Altura por Idade na amostra exploratória	62
Tabela 17 - Resultados para o desfecho de Déficit de Altura por Idade na amostra principal	63

Tabela 18 - Resultados para o desfecho de Déficit de Altura por Idade na amostra específica	64
Tabela 19 - Resultados para o desfecho de Déficit de Altura por Idade na amostra completa	65
Tabela 20 – Resultados dos melhores modelos para o desfecho de AQP	66
Tabela 21 – Resultados dos melhores modelos para o desfecho de Déficit de Altura por Idade	67
Tabela 22 – Definição dos indicadores	83
Tabela 23 – Lista de inquéritos utilizados na construção das amostras	84

LISTA DE ABREVIações E SIGLAS

AQP	Atendimento Qualificado ao Parto
CRISP-DM	Cross Industry Standard Process for Data Mining
DHS	Demographic and Health Surveys
IMC	Índice de Massa Corporal
KDD	Knowledge Discovery in Databases
KNN	K-Nearest Neighbors
ODS	Objetivos de Desenvolvimento Sustentáveis
OMS	Organização Mundial da Saúde
PIB	Produto Interno Bruto
RNA	Redes Neurais Artificiais
SMOTE	Synthetic Minority Over-Sampling Technique
SVM	Support Vector Machine
USAID	United States Agency for International Development
UNICEF	United Nations Children's Fund
WEKA	Waikato Environment for Knowledge Analysis
XGB	Extreme Gradient Boosting

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO TEÓRICA.....	15
2.1	INQUÉRITOS POPULACIONAIS DE SAÚDE	15
2.2	DESFECHOS	16
2.3	MINERAÇÃO DE DADOS	18
2.3.1	SELEÇÃO E COMPREENSÃO DOS DADOS	19
2.3.2	PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS	21
2.3.3	APRENDIZADO DE MÁQUINA.....	22
2.3.4	AVALIAÇÃO DOS MODELOS	30
3	OBJETIVOS E METODOLOGIA	35
3.1	OBJETIVOS.....	35
3.2	METODOLOGIA	36
3.2.1	PREPARAÇÃO E PRÉ-PROCESSAMENTO DAS AMOSTRAS	38
3.2.2	EXECUÇÃO DOS EXPERIMENTOS.....	39
3.2.3	VALIDAÇÃO DOS MODELOS PREDITIVOS	41
3.2.4	FERRAMENTAS UTILIZADAS	42
4	RESULTADOS.....	45
4.1	ANÁLISES POR ATENDIMENTO QUALIFICADO AO PARTO	50
4.1.1	AMOSTRA EXPLORATÓRIA	52
4.1.2	AMOSTRA PRINCIPAL	53
4.1.3	AMOSTRA ESPECÍFICA	54

4.1.4	AMOSTRA COMPLETA	55
4.2	ANÁLISES POR DÉFICIT DE ALTURA POR IDADE	57
4.2.1	AMOSTRA EXPLORATÓRIA	61
4.2.2	AMOSTRA PRINCIPAL	62
4.2.3	AMOSTRA ESPECÍFICA	63
4.2.4	AMOSTRA COMPLETA	64
4.3	VALIDAÇÃO DOS MELHORES MODELOS	65
4.4	DISCUSSÃO DOS RESULTADOS	67
5	TRABALHOS RELACIONADOS	69
6	CONCLUSÕES	72
6.1	TRABALHOS FUTUROS	74
	REFERÊNCIAS	76
	APÊNDICE A - DEFINIÇÃO DOS INDICADORES	82
	APÊNDICE B - LISTA DE INQUÉRITOS DISPONÍVEIS	84
	APÊNDICE C – TRABALHO SUBMETIDO AO CONGRESSO INTERNACIONAL DE EPIDEMIOLOGIA	86

1 INTRODUÇÃO

A disponibilidade de dados padronizados e que permitam que a informação se transforme em conhecimento é essencial para a tomada de decisões em programas e políticas de saúde, tópico evidenciado recentemente nos Objetivos de Desenvolvimento Sustentáveis (ODS) (United Nations, 2015). Graças a programas como o *Demographic and Health Surveys* (DHS) e o *Multiple Indicator Cluster Survey*, a produção de inquéritos populacionais de saúde materno-infantil cresceu muito nos últimos anos, principalmente se comparar com décadas anteriores (Corsi *et al.*, 2012; Murray e Newby, 2012). Essa produção afeta diretamente as pesquisas científicas e políticas sociais, principalmente em países de baixa e média renda, onde os sistemas de informação são escassos e incompletos, tornando os inquéritos populacionais suas principais fontes de informação (Gakidou e Hogan, 2005). A metodologia dos programas conta com questionários padronizados, mas que frequentemente sofrem alterações por serem conduzidos em conjunto com os ministérios ou secretarias dos países. Essas alterações conseqüentemente limitam a coleta de algumas informações, implicando na ausência de estimativas para determinadas intervenções ou cuidados. Tais lacunas limitam a tomada de ações sobre determinado tema, assim como a comparação das estimativas de inquéritos em nível global, regional ou até em tendências no mesmo país (Victora *et al.*, 2012).

As lacunas, como referido às estimativas ausentes, são desfechos de saúde representados por medidas de monitoramento da situação de intervenções ou cuidados de saúde, definidas com base em referências internacionais, como pela Organização Mundial da Saúde (OMS) e pela *United Nation Children's Fund* (UNICEF), presentes nos bancos de dados dos inquéritos populacionais de cada país em determinado ano. Esses desfechos costumam ser binários, indicando a presença ou ausência de alguma cobertura ou intervenção, com base em perguntas dos questionários similares a “O parto foi realizado por profissional qualificado?” ou “A criança apresenta Déficit de Altura por Idade?”.

Em paralelo, as aplicações que utilizam inteligência artificial cresceram exponencialmente nos últimos anos, principalmente nas subáreas relacionadas à ciência dos dados e aprendizado de máquina (Holzinger, 2016). Tais aplicações incluem diversos desafios na área da saúde que consistem em lidar com dados provenientes de pacientes em tratamento, questionários aplicados na admissão ou durante a internação, custos hospitalares, utilização de planos de saúde, entre outros (Koh e Tan, 2005; Tomar e Agarwal, 2013; Yoo *et al.*, 2012). Dentre os problemas solucionados pelo aprendizado de máquina está a classificação de dados, contida nos algoritmos de aprendizado supervisionado, que a partir de um conjunto de informações existentes para uma determinada observação realiza a predição de um atributo de interesse, desde que o mesmo seja discreto (Tan, Steinbach e Kumar, 2006). O processo de estimar atributos de interesse contínuos é conhecida como regressão, também incluso nos algoritmos de aprendizado supervisionado, mas não abordados neste trabalho.

É possível imaginar um conjunto de dados com diversos pacientes que foram diagnosticados com determinada condição, e as características disponíveis, sejam elas biológicas, sociodemográficas ou resultantes de exames, sejam utilizadas para construir um modelo classificador de doenças. Este modelo seria então aplicado em novos pacientes cujas características estejam disponíveis, mas que o diagnóstico não tenha sido realizado.

Como pode ser percebido no exemplo do parágrafo anterior, um classificador baseado em aprendizado de máquina é utilizado para preencher lacunas em dados que possuam outras informações capazes de indicar o real resultado de determinado atributo. Este trabalho propõe uma alternativa para a predição de desfechos de saúde não coletados com o uso de técnicas de aprendizado de máquina, a partir de informações sociodemográficas e indicadores correlacionados presentes nos inquéritos populacionais de saúde da série DHS. Para o estudo foram avaliadas as predições em dois desfechos de saúde: **Atendimento Qualificado ao Parto (AQP)** e **Déficit de Altura por Idade**. Ambos desfechos foram avaliados em quatro cenários diferentes com a aplicação de quatro técnicas de aprendizado de máquina: 1) Redes Neurais Artificiais (RNA), 2) Árvores de Decisão, 3) K-Vizinhos mais próximos e 4) Otimização Extrema de Gradiente (XGB). Os resultados observados demonstram que as técnicas de aprendizado de máquina podem ser aplicadas para predizer este tipo de dado de acordo com o desfecho selecionado e a disponibilidade dos preditores.

Assim, este documento de dissertação está organizado em 6 capítulos, sendo que no Capítulo 2 será apresentado a base teórica utilizada para o desenvolvimento deste trabalho. A metodologia empregada e os objetivos propostos estão descritos no Capítulo 3 e os resultados obtidos em todas as análises são apresentados e discutidos no Capítulo 4. Os trabalhos relacionados encontram-se no capítulo 5, e por fim, o Capítulo 6 discorre sobre as conclusões e as recomendações de trabalhos futuros.

2 REVISÃO TEÓRICA

Neste capítulo serão introduzidos conceitos de epidemiologia e aprendizado de máquina, os quais embasam o desenvolvimento deste trabalho. A primeira seção descreve as fontes de dados que compuseram este trabalho. Em sequência, a segunda seção caracteriza os desfechos que serão avaliados por este trabalho. Já a última seção detalha as etapas do processo de mineração de dados, com ênfase nas técnicas de aprendizado de máquina que serão aplicadas com o intuito de prever os desfechos de interesse nos inquéritos de saúde.

2.1 INQUÉRITOS POPULACIONAIS DE SAÚDE

Os inquéritos populacionais cujos dados são abordados neste trabalho pertencem ao programa *Demographic and Health Surveys* (DHS), criado em 1985 e financiado pelo governo dos Estados Unidos através da USAID (“The DHS Program”, 2017). O programa é focado em países de baixa e média renda e atua em todo o processo de desenho amostral, coleta, análise e divulgação dos resultados em parceria com os ministérios e secretarias de saúde dos países. O método de coleta dos dados é feito através de entrevistas domiciliares com o auxílio de questionários padronizados, que tratam de temas relacionados à saúde e nutrição materno-infantil como amamentação, saúde reprodutiva, vacinação, entre outros. Além disso, dados demográficos e informações sobre cada domicílio

são coletadas, permitindo uma visão geral do contexto que está sendo avaliado (Corsi *et al.*, 2012). Os dados são públicos, mediante solicitação, e disponibilizados separadamente por unidade de amostra, ou seja, existe um banco de dados em que cada observação é uma criança, mulher, domicílio, entre outros. No âmbito deste trabalho, apenas os bancos de dados de crianças serão utilizados, visto que os desfechos são referentes as crianças e estão alocados nestes bancos.

Originalmente, os bancos de dados de crianças são limitados a todas as crianças menores de cinco anos cuja mãe foi entrevistada. Essa restrição é devido ao fato de que todas as perguntas que compõem os módulos das crianças são respondidas pela mãe, e por englobar informações como o período referente a gravidez de cada criança. Estruturalmente, cada observação (linha) é uma criança e cada atributo (coluna) é oriunda de uma ou mais perguntas do questionário. Em média, um banco de dados de crianças possui entre 1.000 e 1.500 variáveis, sendo que muitas delas não possuem relevância para a predição dos desfechos desejados por incluírem diversas perguntas em cascata, isto é, são dependentes de filtros anteriores que reduzem muito o número de crianças em determinados grupos, assim como perguntas específicas do contexto do país, que, ainda que algumas possam ser aproveitadas, não se encaixam na criação de modelos genéricos.

2.2 DESFECHOS

Com o propósito de acompanhar o progresso dos países, assim como garantir a comparabilidade no monitoramento da saúde materno-infantil, as organizações internacionais como UNICEF e a Organização Mundial da Saúde (OMS) possuem indicadores de saúde com definições padronizadas (World Health Organization, 2015). Entretanto, grande parte desses indicadores precisa ser estimada a partir das perguntas disponíveis nos inquéritos.

Os desfechos, ou indicadores de monitoramento em saúde, são variáveis binárias que indicam a presença ou ausência de determinada condição, intervenção ou comportamento em um indivíduo. Com isso, é possível produzir estimativas que reflitam o contexto nacional ou de determinados subgrupos, provendo conhecimento útil à tomada de decisões. O progresso e situação atual de saúde dos países depende de indicadores de monitoramento, como os desfechos, para guiar as políticas públicas e os investimentos para os mais necessitados, visando um acesso mais igualitário a todos os subgrupos da população (Victora *et al.*, 2012). A disponibilidade de indicadores de cobertura também fundamenta iniciativas globais de monitoramento, tais como o *Countdown to 2015*, iniciativa que acompanhou os Objetivos de Desenvolvimento do Milênio com base principalmente em dados oriundos de inquéritos nacionais de saúde (Requejo, Victora e Bryce, 2014).

Em conjunto com pesquisadores do Centro Internacional de Equidade em Saúde, os dois desfechos a seguir foram aprovados como candidatos aos processos de mineração de dados e aprendizado de máquina promovidos por esse trabalho. Dentre os critérios de seleção utilizados, levou-se em consideração a existência de inquéritos DHS nos quais pelo menos um dos desfechos selecionados não existe, tornando a obtenção de estimativas através de um modelo preditivo aplicável em dados reais. A definição do denominador e numerador de cada desfecho selecionado está presente na Tabela 21, inserida no Apêndice A.

O primeiro desfecho selecionado foi **Atendimento Qualificado ao Parto**, onde um profissional qualificado deve enquadrar-se na seguinte definição: profissionais de saúde autorizados tais como parteiras, médicos e enfermeiros que foram educados e treinados para dominar as habilidades necessárias para lidar com gravidezes normais (sem complicações), parto e o período pós-natal imediato, assim como a identificação, gestão e encaminhamento de complicações em mulheres e recém-nascidos (WHO, 2004). As informações

coletadas pelos inquéritos são classificadas a partir da pergunta “Quem auxiliou no seu parto?”, sendo a distinção entre profissionais qualificados e não-qualificados determinada por cada país.

O segundo desfecho trata-se do **Déficit de Altura por Idade**, reflexo de uma exposição prolongada a uma alimentação insuficiente, repetidas infecções e cuidado inadequado (Onis, de e Blössner, 2003). O déficit de altura é um problema crônico que afeta cerca de um quarto de todas as crianças do mundo e cujas consequências podem ser irreversíveis. Este indicador é estimado a partir do escore Z de altura para idade, ou seja, a distribuição das crianças em desvios padrão. Essa distribuição é comparada com as curvas de crescimento padronizadas da OMS que apresentam a expectativa de crescimento adequado para as crianças, de acordo com sua idade (WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP, 2006). Uma criança apresenta déficit de altura quando seu escore Z está abaixo de dois desvios padrão da curva de referência, indicando um crescimento inadequado. Crianças com desvios padrão inferiores a -6 ou superiores a +6 são classificadas como implausíveis.

2.3 MINERAÇÃO DE DADOS

Apesar da grande quantidade de informação que se encontra disponível em bases de dados em todos os segmentos e de diferentes fontes, transformar informação bruta em conhecimento útil de forma automática é um processo complexo, conhecido como mineração de dados. O foco da mineração de dados está em identificar padrões até então desconhecidos nos dados, onde uma de suas aplicações pode ser a predição de desfechos, visto que na medida em que dados futuros sejam recebidos, os mesmos possam ser classificados com base nos padrões identificados. Entretanto, a mineração de dados é apenas uma etapa da descoberta de conhecimento em bases de dados (do inglês *Knowledge Discovery in Databases*, ou KDD), introduzida por (Fayyad, Piatetsky-Shapiro e Smyth, 1996), que engloba todo o processo de extração de conhecimento dos dados. O

KDD aborda desde a obtenção dos dados até a interpretação dos resultados obtidos, consistindo em cinco etapas: seleção, pré-processamento, transformação, **mineração de dados** e interpretação/avaliação.

Dado que o KDD foi pensado como um modelo teórico, algumas extensões surgiram com o propósito de implementá-lo ou transformá-lo em uma metodologia para a aplicação das diretrizes de extração de conhecimento. Dentre eles, o CRISP-DM (*Cross Industry Standard Process for Data Mining*) destaca-se por ter sido uma das mais populares metodologias, contando com um processo completo e bem definido para a aplicação em problemas de mineração de dados (Azevedo e Santos, 2008). Consolidado no final da década de 90 por um consórcio formado por indústrias do segmento (Shearer, 2000), a falta de atualizações limitou a abrangência do CRISP-DM, principalmente se forem considerados os avanços tecnológicos alcançados em quase duas décadas.

O processo de mineração de dados englobado pelo KDD pode ser aplicado com diferentes fins. No contexto deste trabalho serão apresentadas as etapas dos processos desenvolvidos para fins de classificação de dados.

2.3.1 SELEÇÃO E COMPREENSÃO DOS DADOS

Um termo presente em todas as etapas dos processos introduzidos anteriormente e que define o foco, o desempenho e a viabilidade de se produzir informação útil é a compreensão dos dados. A obtenção de resultados relevantes é essencialmente definida pela melhor utilização das bases de dados disponíveis, visto que no mundo real frequentemente são obtidos resultados incompletos, inconsistentes e com algum tipo de erro (Han, Kamber e Pei, 2012). Dentre as principais premissas a serem observadas na construção de uma base de dados está a representatividade do domínio estudado. Isto é, combinações dos atributos disponíveis devem ser capazes de determinar a classificação adequada, caso o desfecho seja uma variável discreta, ou estimar o valor apropriado, caso o

desfecho seja uma variável contínua (Soman, Diwakar e Ajay, 2006). As técnicas de aprendizado de máquina a serem aplicadas são capazes de identificar estes padrões, cada uma utilizando sua abordagem, desde que eles existam na base de dados disponibilizada.

Outro ponto importante de ser observado ao preparar uma base de dados é se a mesma reflete o problema real. Espera-se que os padrões extraídos pelos modelos preditivos seja válidos em novos dados (Fayyad, Piatetsky-Shapiro e Smyth, 1996). Isso é conhecido como a capacidade do modelo de generalizar, pois não se espera que o mesmo funcione apenas para os dados que foram apresentados a ele. Com isso, a base de dados preparada precisa ter dados compatíveis com os dados cujo os modelos serão aplicados posteriormente. É irrelevante utilizar atributos que não existirão em bases de dados futuras, assim como fazer proveito de categorias de atributos que serão descontinuadas ou alteradas. Evidentemente, nem sempre é possível identificar todas as características dos novos dados, especialmente se eles ainda não estão disponíveis para consulta.

Enquanto as premissas anteriores exigem uma abordagem mais conceitual, definir os critérios de seleção tanto dos exemplos quanto dos atributos geralmente é percebido como uma etapa prática. Entretanto, decisões referentes a incluir ou excluir exemplos ou aproveitar determinados atributos devem ser tomadas com cuidado. É possível que atributos sem relevância sejam considerados por algoritmos ao acaso (Guyon e Elisseeff, 2003). Se o objetivo é diagnosticar uma doença, por exemplo, a data em que os dados foram coletados ou o usuário que os inseriu no sistema são irrelevantes e devem ser descartados, mesmo que os algoritmos indiquem que exista uma relação com os dados. No caso dos exemplos, os critérios de inclusão adotados não devem favorecer nem menosprezar nenhum subgrupo, ou os mesmos podem acabar sub-representados ou recebendo importância superior aos demais. O tratamento de dados incompletos, principalmente ao utilizar a abordagem de análise de casos

completos, também pode afetar a representatividade de grupos de dados e será abordada na próxima seção.

A importância da compreensão dos dados em todo o processo justifica o fato de ser uma das primeiras etapas caracterizada pelas estratégias de extração de conhecimento em bases de dados, visto que mudanças identificadas nessa etapa podem redefinir o andamento de todo o processo.

2.3.2 PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS

Após a definição da base de dados, a etapa de pré-processamento dos dados visa prover a melhor estrutura de dados para as técnicas de aprendizado de máquina. Essa é primeira das três etapas cíclicas do processo mineração de dados. O objetivo dessa etapa, de acordo com (Guyon e Elisseeff, 2003), pode ser dividido em três partes:

- 1) Melhorar o desempenho dos preditores selecionados
- 2) Simplificar e otimizar os atributos visando diminuir a velocidade de treinamento e o custo
- 3) Compreender o processo que gerou os dados e suas especificidades.

Efetivamente, esse processo é mais complexo do que determinar a melhor combinação de atributos para um desempenho ótimo. O processo em uma determinada amostra deve considerar circunstâncias como recursos computacionais disponíveis, tempo necessário para o término do projeto, além de todas as implicações descritas no capítulo anterior sobre o conteúdo dos dados (Guyon e Elisseeff, 2003). Durante este processo, inclusive, não basta apenas concentrar esforços em filtrar e reduzir o número de atributos para o mínimo possível. Frequentemente, transformações nos atributos originais como a utilização de técnicas de redução de dimensionalidade, normalização dos dados ou inclusão de informações externas apresentam bons resultados (Tan, Steinbach e Kumar, 2006).

Ao realizar essa etapa, o analista tem a opção de utilizar algoritmos especializados que fornecem informações relevantes no momento de filtrar os atributos. Esses algoritmos produzem uma espécie de ranking dos atributos disponíveis através de um método de avaliação, como por exemplo ganho de informação, e um método de busca (Han, Kamber e Pei, 2012). Mesmo com o uso de tais algoritmos, o conhecimento sobre os dados continua de extrema importância na interpretação dos atributos identificados, correlação entre os mesmos, assim como associações entre atributos já comprovadas na literatura. Aproveitar determinado conhecimento para eliminar atributos sem relevância ou com informações incompletas costuma impactar na minimização do ruído presente (Dash e Liu, 1997). O ruído é um termo atribuído as informações que atrapalham o aprendizado dos modelos preditivos, seja por falta de associação com o desfecho ou dados incorretos.

Além da preocupação necessária com a preparação dos atributos, existem decisões que devem ser tomadas mediante exemplos onde algum dos atributos selecionados está ausente. Uma abordagem comum é a análise de casos completos, onde qualquer exemplo que possuir alguma informação indisponível para algum dos atributos considerados para a análise são descartados. Entretanto, o impacto de desprezar estas informações precisa ser considerado, tendo em vista que, ao descartá-los, presume-se que sua inexistência seja devido ao acaso. Se esse pressuposto não estiver correto, o ato de excluí-los produz viés em favor de um ou mais subgrupos. A publicação de (Brown e Kros, 2003) detalha os possíveis tipos de informação ausente e seus impactos nos modelos preditivos.

2.3.3 APRENDIZADO DE MÁQUINA

Dado que o aprendizado de máquina pode ser considerado o núcleo da mineração de dados (Mannila, 1996), e a etapa responsável pela aplicação de algoritmos específicos com o propósito de extrair padrões relevantes nos dados

ter sido denominada mineração de dados (Fayyad, Piatetsky-Shapiro e Smyth, 1996), este trabalho irá considerar o termo mineração de dados como extração de conhecimento em bases de dados e a aplicação de técnicas de aprendizado de máquina como a etapa de mineração de dados do modelo KDD.

De acordo com (Kohavi e Provost, 1998), o aprendizado de máquina é considerado um campo de estudo onde os algoritmos têm a capacidade de aprender. Para entender melhor a representação do aprendizado de um algoritmo, podemos analisar a seguinte definição apresentada por Mitchell, amplamente conhecida: *"Diz-se que um programa de computador aprende pela experiência E, com respeito a algum tipo de tarefa T e performance P, se sua performance P nas tarefas em T, na forma medida por P, melhoram com a experiência E."* (MITCHELL, 1997, p. 2).

Dentro do contexto coberto pelos algoritmos de aprendizado de máquina, existem diversos problemas em que eles mostram-se eficazes, tais como classificação, regressão, associação, agrupamento, entre outros. O aprendizado de máquina também pode ser dividido em classes de algoritmos, de acordo com a sua abordagem e o tipo de dado disponível. Quando lidamos com problemas de classificação ou regressão, a classe dos algoritmos de aprendizado supervisionado normalmente é a abordagem utilizada. De uma forma prática, o aprendizado supervisionado depende que o banco de dados possua um atributo "resposta" que seja capaz de direcionar os algoritmos na etapa conhecida como treinamento, indicando acertos e erros para cada exemplo processado. Em uma perspectiva paralela, o atributo de "resposta" pode ser considerado como o professor do algoritmo, responsável por induzi-lo.

As quatro técnicas de aprendizado de máquina apresentadas nesta seção diferenciam-se por possuírem estratégias diferentes para abordar um problema de classificação e isso foi considerado como um critério de seleção. Apesar da grande disponibilidade de técnicas avançadas, com alto custo e complexidade, a

extensão do desempenho que pode ser alcançado por cada técnica depende do problema abordado e do tipo de dado disponível, tornando a classificação das técnicas de aprendizado de máquina com base em seu desempenho uma estratégia não recomendada.

2.3.3.1 REDES NEURAIAS ARTIFICIAIS

As Redes Neurais Artificiais (RNA) são técnicas de classificação ou regressão capazes de solucionar uma grande quantidade de problemas existentes, sendo considerada atualmente estado da arte em reconhecimento de objetos e voz (LeCun, Bengio e Hinton, 2015). Inspiradas no funcionamento dos neurônios do cérebro humano, as RNA são compostas de nós interligados através de conexões diretas (Tan, Steinbach e Kumar, 2006), onde cada caminho possui um peso que é ajustado para aumentar ou diminuir a importância de determinada conexão. Em uma hierarquia conhecida como Redes Neurais Artificiais Multi-camadas, estes nós são distribuídos em camadas que propagam os valores dos atributos de entrada ao longo da rede e encontram representações intermediárias para o problema. Um dos processos de treinamento da rede é conhecido como *backpropagation*, o qual retorna o erro encontrado na saída para os neurônios com o propósito de ajustar quais pesos detêm maior impacto no erro propagado. A Figura 1 ilustra uma RNA Multi-camada com cinco atributos na camada de entrada, uma camada oculta com três neurônios e a camada de saída com um resultado. Por tratar-se de uma técnica de aprendizado supervisionado, as RNA precisam ser treinadas com exemplos que possuam um atributo de resposta válido, pois o necessitam para mensurar o erro e otimizar os pesos de toda a rede. Entretanto, devido a grande quantidade de interconexões, as RNA geralmente exigem um grande número de exemplos para apresentar bons resultados, e por possuírem um grande número de pesos, o processo costuma ser lento.

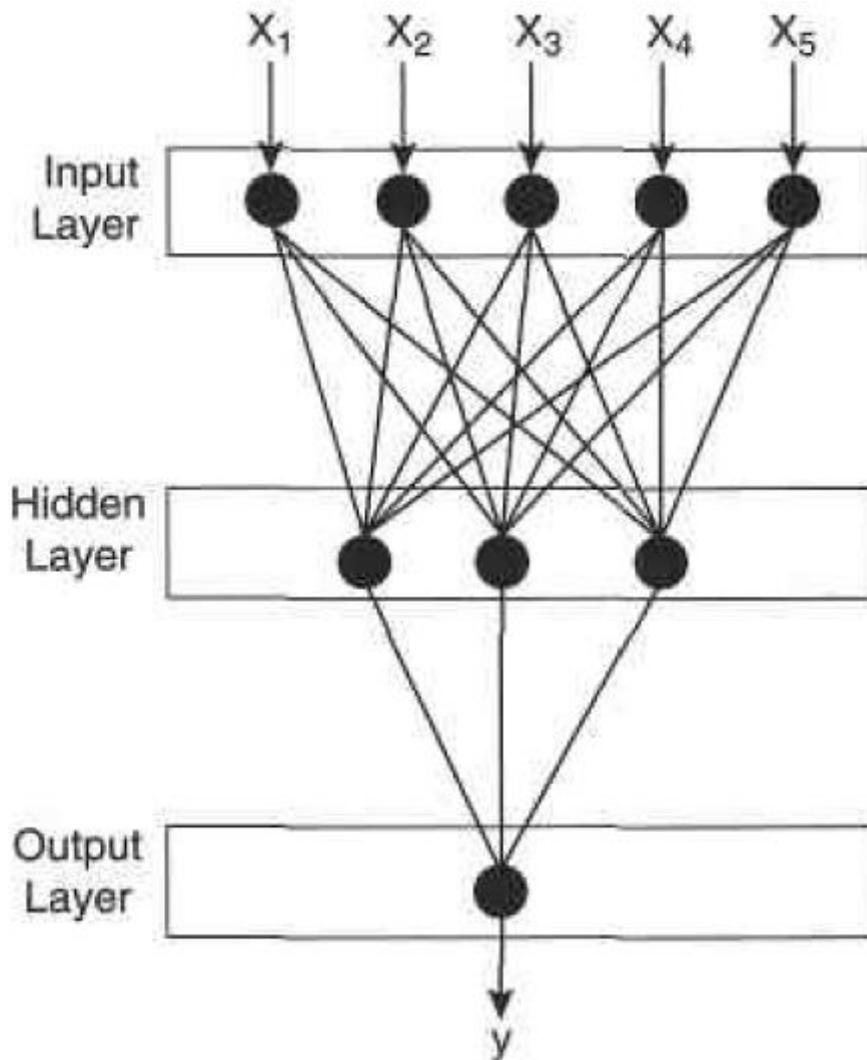


Figura 1 – Rede Neural Artificial com uma camada oculta de três neurônios

Fonte: (Tan, Steinbach e Kumar, 2006)

O desempenho das RNA costuma ser superior em conjuntos de dados que possuam muitas dimensões e atributos contínuos (Kotsiantis, 2007). Na área da saúde, as RNA apresentam bons resultados em pacientes hospitalizados com doenças agudas (Song *et al.*, 2004), no diagnóstico de doenças crônicas (Tomar e Agarwal, 2013) e diagnóstico de doenças através da análise de imagens clínicas (Kaur e Wasan, 2006).

2.3.3.2 ÁRVORES DE DECISÃO

As Árvores de Decisão (AD) são estruturas similares a um fluxograma (Quinlan, 1986). Segundo (Tan, Steinbach e Kumar, 2006), são compostas por:

1. Nó principal: localizado no topo da árvore representado pelo atributo mais relevante, o qual inicia o processo distribuindo o fluxo.
2. Nós internos: Análogos aos galhos, cada nó consiste em um teste realizado sobre um atributo, o qual possui um fluxo de entrada e dois ou mais fluxos de saída.
3. Folhas: Representam uma classe do desfecho, onde cada folha indica o fim de um fluxo apenas, não possuindo fluxos de saída.

A construção das árvores é realizada de maneira hierárquica, visto que os atributos mais relevantes, ou que possuem maior ganho de informação estão localizados nos nós mais altos da árvore. Além disso, não existem fluxos horizontais e um fluxo não pode retornar para um nó que esteja no mesmo nível ou em um nível anterior. Partindo do nó principal, cada exemplo é direcionado para um fluxo na árvore até atingir uma folha, responsável pela sua classificação. A Figura 2 apresenta um exemplo de uma árvore de decisão baseada no conhecido banco de dados de plantas Iris. Nesta figura, é possível observar que cada forma azul representa um nó com um atributo, dividido em dois testes. Se o valor do atributo for inferior a 2,4, ele atinge uma folha e é classificado como Setosa. Caso contrário, o fluxo direciona para um novo atributo, passa por um novo teste e segue o respectivo fluxo.

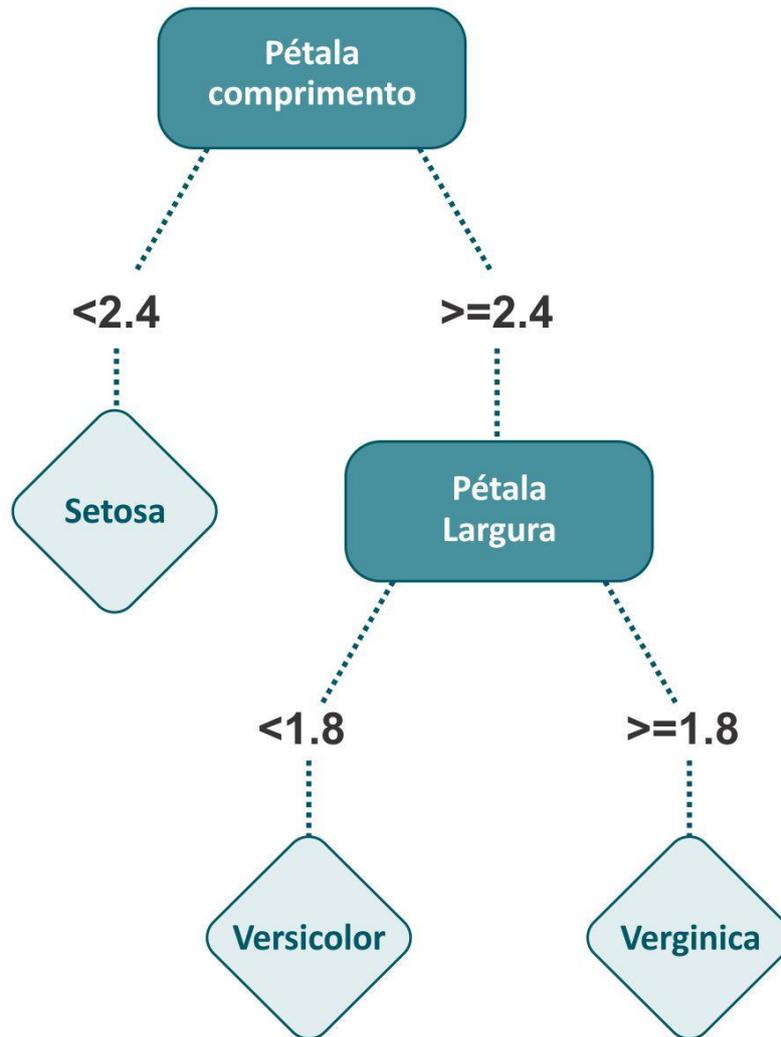


Figura 2 – Árvore de decisão baseada no banco de dados de plantas Iris

Um processo frequente na construção de árvores é a utilização de uma técnica de poda, responsável por eliminar os nós inferiores da árvore com o intuito de evitar o sobreajuste (Witten, Frank e Hall, 2005). Como cada nó é representado por um teste, as árvores de decisão não são capazes de lidar com atributos contínuos. Para utilizá-los, é necessário realizar um processo de discretização, separando-os em faixas (Witten, Frank e Hall, 2005).

Por ser uma técnica extremamente rápida e simples, além de permitir que o usuário visualize quais são os preditores que afetam a classificação de cada classe, as árvores de decisão são amplamente utilizadas em problemas de todas as áreas de conhecimento. Seu desempenho frequentemente é superior quando

atributos discretos compõem os conjuntos de dados (Kotsiantis, 2007). As aplicações que utilizam árvores de decisão na área da saúde são inúmeras, com estudos na literatura indicando sobrevivência de pacientes com câncer de mama (Yoo *et al.*, 2012), padrões de fumo, doenças de pele (Tomar e Agarwal, 2013), diagnóstico de diabetes (Koh e Tan, 2005), entre outros.

2.3.3.3 K-VIZINHOS MAIS PRÓXIMOS

A técnica de K-vizinhos mais próximos, do inglês *K-Nearest Neighbors* (KNN), é uma das técnicas mais simples de aprendizado de máquina. A estratégia utilizada tem base em instâncias, onde os exemplos de treinamento são armazenados e o processo de classificação ocorre diretamente no conjunto de interesse (Mitchell, 1997). A lógica de aprendizado do KNN é baseada na ideia de que ao comparar uma determinada observação com as existentes, a classe de tal observação deve corresponder a classes das observações similares a ela (Han, Kamber e Pei, 2012). A distância entre os k exemplos vizinhos é calculada e a classe atribuída é referente aos vizinhos mais próximos. Existem diversas métricas de distância que podem ser utilizadas no processo de comparação, como Euclidiana e Manhattan.

O principal parâmetro é o valor de k , que de todos os exemplos disponíveis, indica qual o número de exemplos mais próximos será utilizado para classificar o novo exemplo. Em geral, quanto maior o número de exemplos disponíveis no treinamento, maior será o número de k (Han, Kamber e Pei, 2012). A Figura 3 apresenta uma aplicação do KNN com dois valores diferentes de K . De acordo com a figura, a aplicação do KNN com um valor de k igual a 3 resultaria em uma classificação do ponto central em verde pois existem dois pontos verdes e um ponto azul dentre os três vizinhos mais próximos. Aumentando o valor de k para 7, é possível observar que a classificação muda para azul, dado que o número de pontos azuis superou o número de verdes no círculo externo.

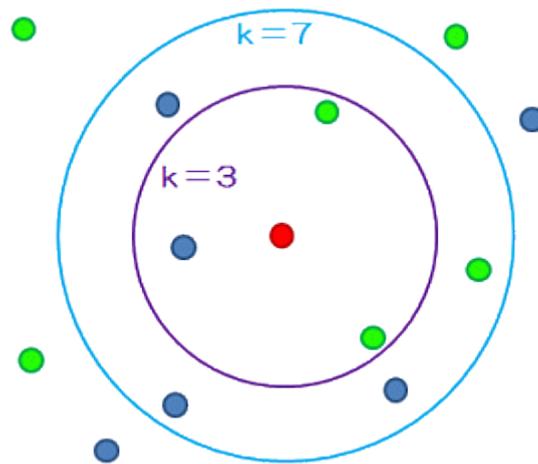


Figura 3 – Demonstração do KNN com dois valores para K

Fonte: (PEDDI, 2016)

O KNN pode ser encontrado em aplicações na área da saúde que envolvem diagnósticos de doenças cardíacas, hipertensão e doenças crônicas (Tomar e Agarwal, 2013).

2.3.3.4 OTIMIZAÇÃO EXTREMA DE GRADIENTE

Pertencente ao grupo de algoritmos que trabalha com técnicas de *boosting*, a otimização extrema de gradiente, traduzida de *Extreme Gradient Boosting* (XGB), é uma técnica de aprendizado de máquina desenvolvida por (Chen e Guestrin, 2016) como uma extensão da otimização de gradiente proposta por (Friedman, 2001).

A estratégia de *boosting* é uma melhoria do processo de *bagging*, que consiste em treinar diversos submodelos com subamostras aleatórias no processo de treinamento e combiná-los para obter um desempenho menos especializado e conseqüentemente com menos sobreajuste. Essa abordagem é utilizada na conhecida técnica de florestas aleatórias introduzida por (Breiman, 2001). No caso do *boosting*, cada nova subamostra prioriza exemplos classificados incorretamente no submodelo anterior. No caso do algoritmo XGB, as árvores de decisão também são utilizadas para construir os submodelos.

Inclusive, a eficiência do XGB resultou em sua premiação no *John M. Chambers Statistical Software Award* em 2016.

Por ser uma técnica relativamente recente, ainda não existem muitos trabalhos na literatura que apresentem problemas que possam ser solucionados por esta abordagem, mas um bom indicativo é a utilização desta técnica por todos os 10 melhores times do KDD CUP 2015 (Chen e Guestrin, 2016).

2.3.4 AVALIAÇÃO DOS MODELOS

Durante o processo de treinamento dos algoritmos, os modelos necessitam mensurar seu desempenho para identificar e quantificar o que está sendo aprendido. Nessa etapa, é possível identificar se o modelo está de fato aprendendo padrões referentes ao conceito do desfecho que deseja ser predito, ou se está absorvendo especificidades do banco de dados disponível para treinamento (Weiss e Zhang, 2003). A validação dos modelos afeta a comparação entre os parâmetros de uma própria técnica assim como comparação do desempenho entre modelos de técnicas diferentes. Essa avaliação ocorre ao expor o modelo preditivo a uma base de dados que ele ainda não teve contato, frequentemente através de uma das seguintes estratégias: particionamento do banco de dados ou validação cruzada.

A primeira estratégia consiste em dividir o banco de dados em duas ou três partições, onde a maior será disponibilizada para a etapa de treinamento e as demais para validação. De acordo com (Witten, Frank e Hall, 2005), a divisão pode ocorrer de duas formas:

- 1) Em **duas** partes: treinamento e validação. O treinamento é utilizado no aprendizado do modelo e a validação atua para validar o desempenho de cada combinação de técnicas e parâmetros. Uma proporção comum é separar em torno de 2/3 das observações para treinamento e a terça parte remanescente para avaliar o desempenho (Witten, Frank e Hall, 2005)

- 2) Em **três** partes: treinamento, validação e teste. O treinamento é utilizado no aprendizado do modelo, a validação atua na seleção dos parâmetros, identificando os melhores modelos, e o teste é utilizado para medir o desempenho dos melhores modelos.

Dentre as principais vantagens de particionar o banco de dados está o ganho de velocidade de processamento na criação dos modelos, visto que nenhuma etapa adicional é necessária (Weiss e Zhang, 2003). Em compensação, toda a informação presente nas observações pertencentes a partição de testes não será aproveitada no aprendizado. Essa abordagem é preferida quando o número total de observações disponíveis é alto, onde a remoção de uma parte do banco de dados não implica em perda de representatividade do conceito que se espera que os modelos extraiam.

Já a validação cruzada é uma estratégia que busca dividir o conjunto de treinamento em n partes, utilizando $n-1$ partes para o processo de treinamento e a parte restante a validação (Kohavi, 1995). Com isso, n modelos são gerados com cada parte atuando uma vez como validação e as demais vezes como treinamento, sendo a média de desempenho dos n modelos o determinante do desempenho real. O valor de n é arbitrário, sendo a divisão em 10 partes um parâmetro frequente. Uma variação dessa abordagem, conhecida como *leave-one-out*, ou deixando um de fora, fixa o n no número total de observações, realizando todas as permutações possíveis para determinado banco de dados (Kohavi, 1995). Essas abordagens se destacam quando o conjunto de dados é de pequeno a médio porte, onde a perda de observações no caso de separação pode implicar consideravelmente no desempenho do modelo. Em compensação, a grande desvantagem desse método é que o procedimento é computacionalmente pesado, devido ao tempo necessário para realizar todas as permutações (Witten, Frank e Hall, 2005). O valor atribuído a n também indica a quantidade de modelos que precisam ser treinados, ou seja, quanto maior o n , mais lento o processo. Dado que a quantidade de observações total no banco de

dados também determina a velocidade do processamento das técnicas de aprendizado de máquina, aplicar essas abordagens em bancos de dados muito grandes pode ser inviável.

Independentemente da estratégia de validação aplicada, a comparação dos modelos preditivos ocorre por meio de diferentes métricas, onde cada aplicação deve determinar qual métrica é mais importante de ser maximizada. Abaixo são detalhadas as principais métricas de avaliação dos modelos:

- **Acurácia:** Também conhecida como concordância, é uma métrica que considera o número total de acertos do modelo sobre o número total de observações disponíveis. A acurácia apresenta uma boa medida de desempenho geral do modelo, sem levar em consideração as especificidades como a direção dos erros obtidos ou a distribuição das categorias do desfecho. Na equação abaixo: TP = Verdadeiros positivos; TN: Verdadeiros negativos; P = Total de positivos; N = Total de negativos.

$$A = \frac{(TP + TN)}{(P + N)}$$

- **Sensibilidade e Especificidade:** Essas duas métricas são importantes principalmente em aplicações onde as consequências de um falso positivo ou negativo podem ser graves. A sensibilidade mede o número verdadeiros positivos sobre total de positivos reais, enquanto a especificidade mede o número de verdadeiros negativos sobre o total de negativos reais (POWERS, 2011). Por exemplo, um modelo preditivo que realiza uma triagem para identificar a compatibilidade de um paciente com determinada droga ou tratamento não deve permitir que pacientes incompatíveis a recebam. É preferível ter um modelo que identifique menos pacientes aptos, com uma acurácia menor, do que um modelo que aceite pacientes incompatíveis.

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad \text{Especificidade} = \frac{VN}{VN+FP}$$

- **Estatística Kappa:** Também conhecido como coeficiente kappa de Cohen, essa métrica considera a possibilidade de a concordância ter ocorrido pelo acaso. O kappa é construído com base na acurácia e na concordância esperada se a classificação ocorresse ao acaso (POWERS, 2011). Na equação abaixo: p_e = Concordância esperada ao acaso; p_o : concordância observada (acurácia);

$$K \equiv \frac{p_o - p_e}{1 - p_e}$$

Com base nas descrições das métricas, percebe-se que o impacto na aplicação define qual deve ser priorizada. Entretanto, algumas métricas individualmente podem resultar na seleção de um modelo preditivo ruim. Se um modelo assume que a maior parte das observações pertence à classe positiva, o número de falsos negativos tende a ser baixo. Encontrar um equilíbrio entre duas ou mais métricas ajuda a descartar modelos extremos.

Ao lidar com soluções de aprendizado de máquina, uma situação que deve ser minimizada é o quanto o modelo preditivo se especializa nos dados. Visto que o conceito ou a abstração do problema deve ser ensinado para os modelos, e isto é feito através de um conjunto específico de dados (conjunto de treinamento), o que se busca é a uma generalização do problema que possa ser aplicado a novas bases de dados (Weiss e Zhang, 2003). Quando determinado modelo introduz nas predições as características específicas do conjunto estudado, este comportamento é conhecido como sobreajuste. Validar o desempenho dos modelos em conjuntos de dados não utilizados na etapa de treinamento são essenciais para identificar esse comportamento.

O próximo capítulo apresenta a aplicação do que foi introduzido nesta revisão teórica com o propósito de predizer dados ausentes. Os inquéritos populacionais de saúde posicionam-se como as fontes de dados na qual os

processos de mineração de dados atuam. Na aplicação destes processos, os desfechos de saúde apresentados são os alvos da predição.

3 OBJETIVOS E METODOLOGIA

Neste capítulo são detalhados os objetivos que este trabalho visa atingir, e a metodologia proposta para solucionar os problemas de ausência de informação para indicadores de saúde. A primeira seção contém o objetivo geral e os objetivos específicos deste trabalho. Na segunda seção é introduzida de forma geral a metodologia e o processo resumido. Este processo é detalhado em quatro subseções, iniciando com a preparação e pré-processamento dos dados, sucedido pelas análises conduzidas com cada técnica de aprendizado de máquina, e em seguida a avaliação do desempenho dos modelos. A última subseção apresenta as ferramentas utilizadas no desenvolvimento deste trabalho.

3.1 OBJETIVOS

Este trabalho visa primordialmente avaliar a possibilidade de prever desfechos relacionados a intervenções ou cuidados de saúde materno-infantis, com base em um conjunto de dados de inquéritos DHS, através de técnicas de aprendizado de máquina. A capacidade dessa predição, após observada, define a aplicabilidade do estudo em problemas reais, visto que a magnitude do erro resultante implica na aceitação ou rejeição do modelo por especialistas em um estudo de caso apresentado, e ultimamente por cada usuário baseado nas métricas disponibilizadas.

Especificamente, durante a aplicação da proposta apresentada acima, o trabalho propõe-se a:

1. Ao aplicar técnicas de aprendizado de máquina na predição de desfechos de saúde, identificar se o desempenho varia e quais técnicas produzem melhores resultados.
2. Comparar os resultados obtidos entre os diferentes desfechos e identificar, se possível, razões para variações no desempenho.
3. Para o desfecho que possui um preditor altamente correlacionado, atingir o maior ganho de desempenho possível sob este preditor ao agregar novos atributos, e alcançar um balanceamento no número de falsos positivos e falsos negativos, visando minimizar a possibilidade do modelo de superestimar ou subestimar as estimativas, em média.

3.2 METODOLOGIA

Em síntese, o desenvolvimento deste trabalho pode ser descrito como a aplicação das etapas de mineração de dados em diferentes cenários cujo dados são provenientes de uma série de inquéritos nacionais de saúde conhecidos como DHS. Esta aplicação tem como propósito prever a ocorrência de eventos caracterizados como desfechos de saúde. As etapas conduzidas neste trabalho baseiam-se nos processos apresentados pelo modelo do KDD, levando em consideração as duas primeiras fases do CRISP-DM: entendimento de domínio e entendimento dos dados, que, apesar de conceitualmente englobadas na primeira fase do KDD, merecem destaque e atenção especial, como priorizado no processo do CRISP-DM. É importante salientar que apesar das fases possuírem uma sequência, elas são recorrentes, dado que em qualquer momento pode ser necessário retornar a uma etapa anterior.

Os dois desfechos avaliados, Atendimento Qualificado ao Parto e Déficit de Altura por Idade, foram avaliados em 4 cenários diferentes: amostra de caráter

exploratório, amostra correspondente a situação atual, amostra específica de um país e amostra completa de países ao longo do tempo. Inicialmente, o processo de mineração de dados foi aplicado ao primeiro cenário com o propósito de obter uma melhor compreensão do tipo de dado disponível, assim como avaliar os atributos que poderiam ser úteis e, observar o desempenho preliminar das técnicas de aprendizado de máquina em determinada aplicação. Em sequência, o mesmo processo incrementado pela experiência obtida na primeira análise foi conduzido para os demais cenários e novamente para a amostra exploratória. A Figura 4 ilustra as etapas do processo de mineração de dados utilizados. O detalhamento de cada uma destas etapas e as observações extraídas da abordagem exploratória serão detalhadas nas subseções deste capítulo.

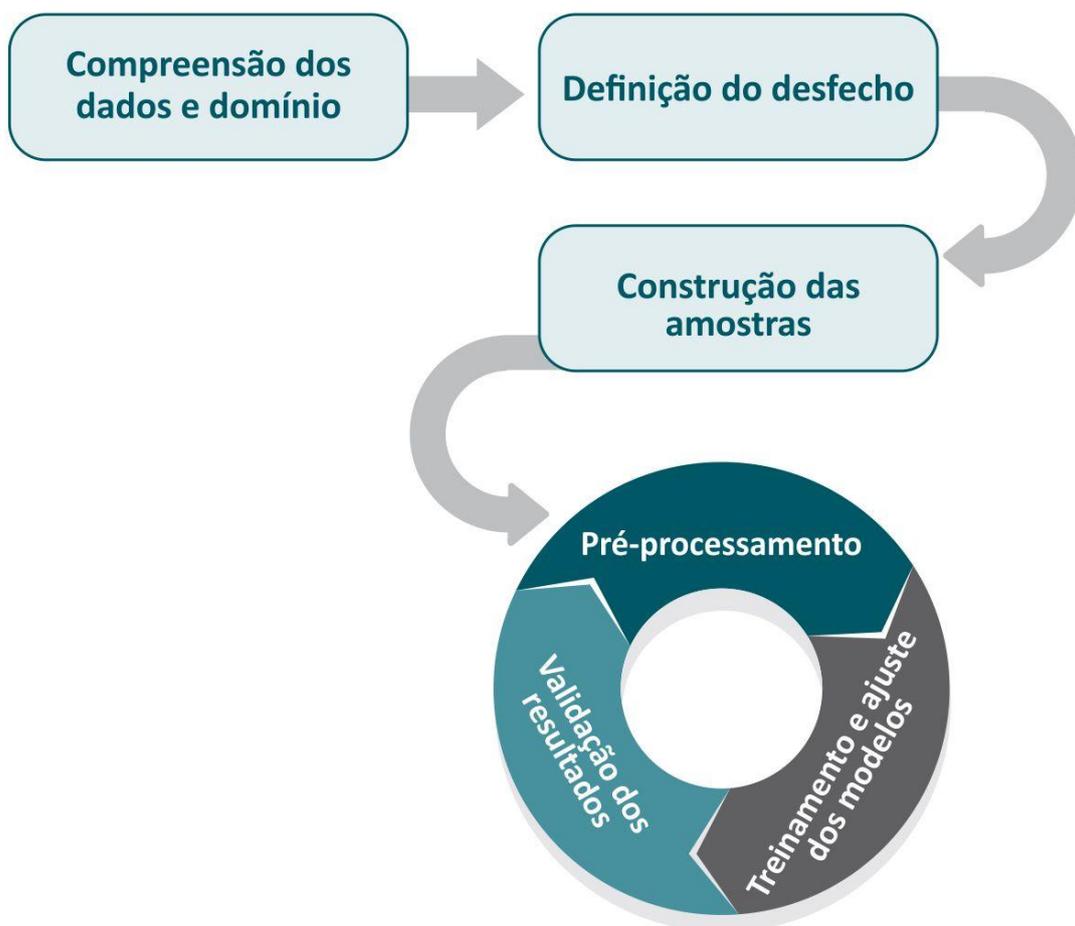


Figura 4 – Fluxograma do processo de mineração de dados

3.2.1 PREPARAÇÃO E PRÉ-PROCESSAMENTO DAS AMOSTRAS

Todas as amostras foram construídas a partir de bancos de dados de crianças de inquéritos DHS. Dentre os inquéritos disponíveis para a construção das amostras estavam todos os inquéritos padrão DHS (*standard DHS*), conduzidos a partir da fase II do programa, com dados publicamente disponíveis no website no período de 5 a 15 de fevereiro de 2017. As amostras foram preparadas com base em todos os inquéritos disponíveis, sendo descartados durante a análise de cada desfecho os inquéritos em que sua informação não estava presente. A seguir são descritas as especificidades de cada amostra:

1. Amostra **exploratória**: Composta de 25 inquéritos selecionados aleatoriamente conduzidos entre 2010 e 2015. Este intervalo de 6 anos foi utilizado para caracterizar o panorama atual dos países, cuja restrição é idêntica a aplicada para a amostra **principal**.
2. Amostra **principal**: Conta com todos os inquéritos disponíveis no intervalo de até 6 anos observados desde o inquérito mais recente (2010-2015)
3. Amostra **específica**: Estudo de caso de um país que possua pelo menos 5 inquéritos disponíveis ao longo do tempo.
4. Amostra **completa**: Incluir os inquéritos de todos os países para todos os anos disponíveis.

A construção de cada amostra foi realizada através de um processo de união de todos os bancos de dados de crianças dos inquéritos selecionados. Em sequência, foi realizada a divisão de cada amostra em 3 partes: 80% para treinamento, 10% para validação e 10% para teste. Essa divisão foi conduzida de forma estratificada, garantindo a proporção original dos indivíduos em cada categoria do desfecho.

Cada inquérito foi submetido individualmente a um processo padronizado para estimar os indicadores internacionais de monitoramento seguindo as definições e orientações da UNICEF e da Organização Mundial da Saúde, assim

como demais indicadores estimados rotineiramente pelo Centro Internacional de Equidade em Saúde. A criação dos dois desfechos que o trabalho visa predizer está englobada neste processo. O Apêndice A sintetiza a definição aplicada para a criação de cada um dos indicadores aproveitados neste trabalho.

Inicialmente, todos os atributos originais foram mantidos, com exceção das perguntas não padronizadas que possuem especificidades de cada país e não são comparáveis com o restante da amostra. A primeira triagem removeu todos os atributos cuja proporção da categoria sem resposta (*missing*) era superior a 75%.

Complementarmente, um atributo contendo dados agregados para cada combinação de país/ano foi incluído nas amostras. Esse atributo foi desenvolvido por (Fink *et al.*, 2017) a partir do Produto Interno Bruto (PIB) dos países no ano em questão e outras informações, representando sua renda absoluta. É uma boa medida de comparação econômica entre contextos de diferentes países em diferentes momentos. Esse valor, cuja unidade é estabelecida em dólares, é estimado para cada percentil de riqueza e vinculado aos inquéritos estudados através dos percentis criados com base no escore do índice de riqueza.

3.2.2 EXECUÇÃO DOS EXPERIMENTOS

Com as amostras definidas, a análise de cada desfecho iniciou pela seleção dos atributos que compuseram as amostras reduzidas submetidas aos algoritmos de aprendizado de máquina. Este processo foi realizado com duas abordagens:

1. Abordagem conceitual, onde o estudo de possíveis variáveis relacionadas a cada desfecho, consultadas na literatura e com especialistas, foram consideradas.
2. Abordagem técnica, contando com o auxílio de técnicas especializadas que baseiam-se nas relações existentes entre os dados avaliados. Foram

utilizadas duas combinações de algoritmos disponibilizados pelo software *Waikato Environment for Knowledge Analysis (WEKA)*.

As análises consistiram na aplicação de 4 técnicas de aprendizado de máquina diferentes em cada amostra: árvores de decisão, redes neurais artificiais, otimização extrema de gradiente e k-vizinhos mais próximos.

Diversas técnicas de classificação não permitem ou não funcionam bem ao se depararem com atributos categóricos. Para contornar essa adversidade, uma estratégia chamada *Dummification*, converteu os atributos categóricos em n variáveis *dummy* binárias, onde n é igual ao número de categorias existentes. A Figura 5 ilustra uma variável categórica no formato original e após o processo.

Original					
country	<chr>				
1	Albania				
2	Albania				
3	Albania				
4	Albania				
5	Albania				
6	Albania				
Dummy					
countryAlbania	countryArmenia	countryAzerbaijan	countryBrazil	countryBurkina_Faso	countryBurundi
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0

Figura 5 – Atributo país no formato original (acima) e no formato dummy (abaixo)

As amostras, reduzidas apenas aos atributos selecionados, foram utilizadas na criação de diversos modelos preditivos onde cada combinação de parâmetros, para cada técnica de aprendizado de máquina, gera um modelo diferente. Adicionalmente, todos os modelos gerados foram recriados aplicando duas abordagens de pré-processamento conhecidas como centralização e escala. O recurso de centralizar significa subtrair a média de todos os valores, com o propósito de transformar a média em zero. Já a escala de valores é resultado da

divisão de cada observação pelo desvio padrão da amostra, provendo valores aos modelos com desvio padrão igual a um.

Nas análises do desfecho de Déficit de Altura por Idade, uma técnica conhecida como SMOTE foi aplicada para corrigir o desbalanceamento entre as classes do desfecho. Esta técnica expande as classes sub-representadas do desfecho ao criar exemplos sintéticos ao invés de simplesmente expandir a amostra com substituição (Chawla *et al.*, 2002).

3.2.3 VALIDAÇÃO DOS MODELOS PREDITIVOS

A primeira etapa de validação ocorre ao selecionar um modelo preditivo para cada técnica de aprendizado de máquina mediante os diversos modelos gerados, um com cada combinação de parâmetros. Para este propósito, uma abordagem conhecida como validação cruzada *k-fold* foi aplicada com $k=10$ para selecionar o modelo com a melhor combinação de parâmetros, mantendo apenas um modelo preditivo por técnica. O melhor modelo é determinado pelo que obtiver a acurácia mais alta nesta etapa. O desempenho dos modelos selecionados é então avaliado em dados que ainda não foram observados pelos algoritmos. Isso permite analisar se os modelos gerados estão aprendendo de acordo com os padrões existentes ou se estão limitados aos dados observados no treinamento.

A validação externa dos melhores modelos foi conduzida com o auxílio de pesquisadores da área da saúde, cujo conhecimento dos dados avaliados, assim como do impacto dos resultados obtidos através das respectivas previsões, permite estimar uma aproximação do que seria um desempenho mínimo aceitável na aplicação em problemas reais. Essa validação será discutida no capítulo 6.

3.2.4 FERRAMENTAS UTILIZADAS

Os bancos de dados originais, obtidos junto ao website da DHS, encontram-se em uma extensão atribuída ao pacote estatístico Stata. O mesmo foi responsável por auxiliar em toda a primeira etapa de pré-processamento, na padronização de alguns atributos, assim como a estimação dos indicadores de saúde sejam eles desfechos ou possíveis preditores. O particionamento das amostras também foi realizado utilizando o Stata a partir do comando *sample*, estratificando apropriadamente de acordo com a proporção dos desfechos (StataCorp, 2013).

Os algoritmos de seleção de atributos utilizados na etapa de pré-processamento dos conjuntos de dados foram implementações presentes no software WEKA. Cada algoritmo é a combinação de um método de busca e um método de avaliação, e no contexto deste trabalho foram utilizadas duas combinações diferentes: o método de busca *ranker* em conjunto com o método de avaliação *InfoGainAttributeEval*, que classifica em ordem de relevância os atributos selecionados utilizando uma medida de entropia, similar a estratégia adotada pelas árvores de decisão, e o método de busca *BestFirst* com o método de avaliação, que expande a busca pelo caminho mais promissor de acordo com o custo de cada caminho identificado pela função de avaliação.

Optou-se por desenvolver uma plataforma para a condução das análises de mineração de dados na linguagem de programação R. As principais vantagens observadas foram a disponibilidade de um grande número de bibliotecas com diversos algoritmos implementados, possibilidade de paralelização dos processos de análise e estabilidade. O R destaca-se por ser uma linguagem de programação de código livre mantida por um grupo de profissionais desde 1997 (R Core Team, 2017).

O desenvolvimento da plataforma contou com o auxílio de quatro bibliotecas desenvolvidas por usuários, além dos pacotes requeridos (dependências) e das técnicas de aprendizado de máquina:

1. *dplyr*: biblioteca utilizada para auxiliar no gerenciamento de bases de dados (Wickham *et al.*, 2017).
2. *haven*: biblioteca desenvolvida para permitir a integração de bancos de dados do Stata (Wickham e Miller, 2017).
3. *doSNOW*: biblioteca que permite o uso de diretivas de paralelização de código para aumentar o desempenho (Analytics e Weston, 2015).
4. *caret*: biblioteca que conta com uma série de funções para pré-processamento, treinamento e validação de algoritmos de aprendizado de máquina, além de possuir um sistema de encapsulamento que permite o uso de bibliotecas de técnicas de aprendizado de máquina já consolidadas (Kuhn, 2016).
5. *DMwR*: biblioteca que implementa a técnica de correção de balanceamento de classe SMOTE (Torgo, 2010).

As quatro técnicas de aprendizado de máquina adotadas para o desenvolvimento deste trabalho contém suas respectivas implementações como bibliotecas para o R. Para cada implementação, uma série de parâmetros estava disponível para ajustar os modelos preditivos:

- Redes neurais artificiais: Foi utilizado a biblioteca *nnet* (Venables e Ripley, 2002) que possui dois parâmetros ajustáveis: *decay*, que representa a magnitude da penalidade aplicada aos pesos e *size*, que contabiliza o número de neurônios disponíveis na camada oculta.
- Árvores de decisão (J48): A implementação do J48 no R (Hornik, Buchta e Zeileis, 2009), compatibilizada da versão em Java do WEKA baseada no algoritmo C4.5, também dispõe de dois parâmetros ajustáveis: *Confidence Threshold*, que determina a magnitude da poda aplicada as árvores de

decisão, e *Minimum Instances Per Leaf*, diretamente traduzido como número mínimo de instâncias por folha, limita as pontas das árvores de decisão para remover especificidades, minimizando o sobreajuste.

- K-vizinhos mais próximos: A implementação do KNN é mantida no código base do R por ser uma técnica consolidada e com uma abordagem bastante simples. O único parâmetro disponível é o *K*, indicando o número de vizinhos a serem comparados com cada exemplo.
- Otimização extrema de gradiente: Desenvolvida para ser uma técnica escalável, portátil e distribuída, a biblioteca desenvolvida por (Chen *et al.*, 2017) conta com 7 parâmetros ajustáveis:
 - *nrounds*: número de iterações máximo para a estratégia de boosting.
 - *max_depth*: profundidade máximo das árvores construídas.
 - *eta*: conhecido como taxa de aprendizado, pondera as árvores com o objetivo de diminuir o sobreajuste.
 - *gamma*: define a redução mínima necessária na função de perda para que uma divisão na árvore seja permitida.
 - *colsample_bytree*: determina a porcentagem dos atributos que vão ser selecionados para construir cada árvore.
 - *min_child_weight*: valor mínimo resultante da soma dos pesos de cada exemplo em uma árvore. Quanto maior, mais conservador, portanto, menos sobreajuste.
 - *subsample*: determina a porcentagem da amostra que vai ser aleatoriamente selecionada para construir cada árvore

4 RESULTADOS

Neste capítulo são apresentados os resultados obtidos através da aplicação da metodologia proposta, os quais originaram uma análise para cada desfecho apresentado em cada uma das quatro amostras construídas. Sua construção, de acordo com as características de cada amostra apresentada na metodologia, foi originada a partir de 201 inquéritos DHS disponíveis conduzidos de 1993 a 2015, representando 72 países. A Tabela 1 sintetiza a composição de cada amostra e apresenta entre parênteses o número total de inquéritos disponíveis para cada desfecho. O Apêndice B apresenta a lista de todos os inquéritos disponíveis utilizando nas análises. O tamanho amostral de cada inquérito varia desde 500 até 40.000 crianças, com uma média de 10.000 e mediana em 8.000 crianças. A produção dos inquéritos é bem distribuída ao longo dos anos, com uma média de quase 9 inquéritos por ano, variando de 4 inquéritos conduzidos em 2002 até um pico de 15 inquéritos em 2005. A Figura 6 apresenta a distribuição dos inquéritos ao longo dos anos.

Tabela 1 – Composição das amostras

Amostra	Composição
Exploratória	25 inquéritos selecionados aleatoriamente entre 2010 e 2015
Principal	Todos os inquéritos disponíveis entre 2010 e 2015 (47,60)
Específica	Todos os inquéritos de Bangladesh (1993, 1996, 1999, 2004, 2007, 2011)
Completa	Todos os inquéritos disponíveis (177,156)

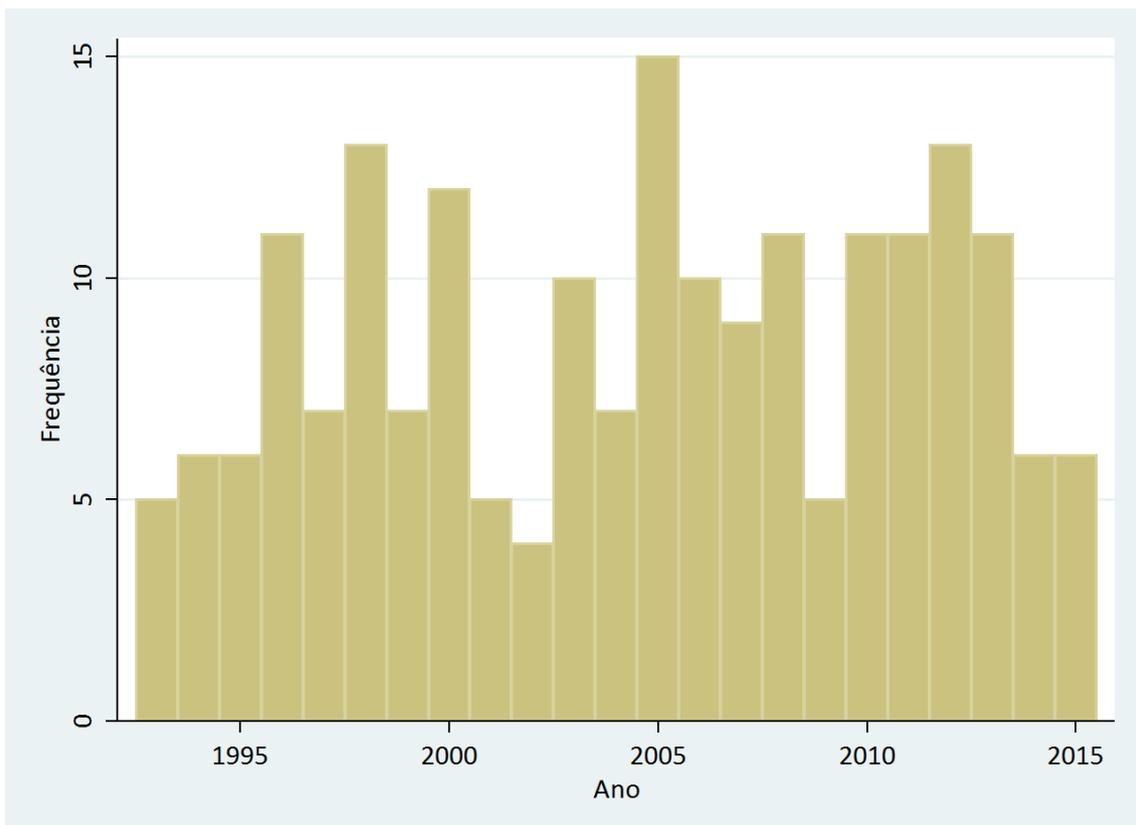


Figura 6 – Histograma da frequência dos inquéritos DHS ao longo do tempo

Em ambas análises foram utilizados atributos originais presentes nos inquéritos DHS, assim como indicadores derivados de atributos construídos a partir de definições padronizadas. Além dos indicadores padronizados

mencionados, alguns atributos foram criados ou categorizados de forma padronizada para garantir a comparabilidade entre os inquéritos. Dos atributos selecionados pelas análises estão: níveis de educação materna e o índice de bens. A educação materna foi classificada em três categorias: nenhuma, qualquer tipo de educação primária e qualquer tipo de educação secundária ou superior, ajustando para alguns países categorias de origem religiosa. Já o índice de bens foi criado a partir da aplicação da técnica de Análise de Componentes Principais nos bens de cada domicílio, assim como materiais de construção e sistemas de abastecimento de água e saneamento. O índice de bens é uma medida agregada que serve como proxy para a renda de um domicílio levando em consideração suas posses, assim como estruturas do domicílio e condições de moradia (Filmer e Pritchett, 2001). Esse índice é um pré-processamento realizado pela equipe da DHS, e apesar de ser um atributo original do banco de dados, é baseado em uma transformação de diversas perguntas combinadas.

Adicionalmente, após as análises serem conduzidas apenas com atributos originais, todas as amostras utilizaram o atributo agregado de renda absoluta. A partir do score do índice de bens, uma divisão em percentis é realizada por inquérito e o índice é adicionado ao banco de dados. Inquéritos cuja informação para o índice de bens não está disponível, assim como o inquérito de Myanmar que não possuía nenhuma informação sobre renda absoluta, foram excluídos das análises. O inquérito do Haiti no ano de 1994 não possuía valor de renda absoluta estimado e o valor predito através de uma regressão linear foi utilizado com base nos demais anos disponíveis.

As amostras foram submetidas a quatro técnicas de aprendizado de máquina onde cada técnica contou com uma variedade de combinações de parâmetros, observando a disponibilidade na implementação de cada algoritmo. As Tabelas 2-5 apresentam os parâmetros disponíveis para cada técnica, assim como os valores estabelecidos para ajustar cada modelo.

Tabela 2 – Parâmetros disponíveis e valores atribuídos para Redes Neurais Artificiais

Redes Neurais Artificiais	
Parâmetro	Valor
size	3
size	5
size	7
size	10
decay	0,1
decay	0,2
decay	0,3
decay	0,4
decay	0,5

Tabela 3 – Parâmetros disponíveis e valores atribuídos para Árvores de Decisão

Árvores de Decisão	
Parâmetro	Valor
C	0,2
C	0,3
C	0,4
M	2
M	3
M	5

Tabela 4 – Parâmetros disponíveis e valores atribuídos para K-Vizinhos Mais Próximos

K-Vizinhos Mais Próximos	
Parâmetro	Valor
K	3
K	5
K	7
K	11
K	15
K	19

Tabela 5 – Parâmetros disponíveis e valores atribuídos para Otimização Extrema de Gradiente

Otimização Extrema de Gradiente	
Parâmetro	Valor
nrounds	100
nrounds	500
eta	0,01
eta	0,001
max_depth	2
max_depth	4
max_depth	6
max_depth	8
gamma	0,5
gamma	1
subsample	1
min_child_weight	0,5

min_child_weight	1
colsample_bytree	0,8

4.1 ANÁLISES POR ATENDIMENTO QUALIFICADO AO PARTO

Os experimentos realizados nas análises do desfecho de Atendimento Qualificado ao Parto identificaram quatro atributos preditores além dos atributos utilizados para identificar os inquéritos individualmente em cada amostra. O principal atributo, com impacto já reconhecido na literatura e com alto desempenho nos algoritmos de seleção de atributos, foi a realização do parto em um ambiente institucional. Classificado como um atributo binário a partir da pergunta “Local onde o parto ocorreu”, ele apresenta diversas categorias, variando de país para país, desde hospitais, clínicas e postos de saúde até casa de amigos ou familiares. Esse preditor é altamente correlacionado com o desfecho, como observado nas taxas de concordância apresentada para cada amostra na Tabela 6. Os resultados apresentados nas análises deste desfecho incluem uma estimativa denominada como **base**, que indica o poder de predição de determinado algoritmo utilizando apenas o preditor principal. Ao comparar as demais estimativas com a estimativa de base, é possível observar o quanto os demais preditores contribuíram para a melhoria na estimativa. As concordâncias apresentadas na Tabela 6 representam a estimativa obtida em cada amostra antes do processo de particionamento. As estimativas apresentadas como base em cada uma das amostras analisadas são referentes ao conjunto de dados de validação, visto que as estimativas geradas pelos modelos são produzidas com base neste conjunto e não seria compatível compará-las com as estimativas obtidas através de uma amostra com composição diferente.

Tabela 6 – Concordância e concordância esperada de cada amostra para AQP

Amostra	Concordância	Concordância Esperada
Exploratória	95,24%	57,92%
Principal	94,27%	52,84%
Específica	95,67%	71,83%
Completa	92,52%	50,64%

Os três atributos originais selecionados que contém menor impactos são: Área de residência (zona urbana ou rural), educação materna em níveis (nenhuma, primária ou secundária/superior) e escore contínuo do índice de bens. Além destes atributos, o país foi incluído nas amostras exploratória, principal e completa, e o ano do inquérito foi incluído na amostra específica e completa. Todos os atributos utilizados nas análises deste desfecho são apresentados na Tabela 7.

Tabela 7 – Atributos utilizados nas análises de AQP

Atendimento Qualificado ao Parto	
Atributos	Valores
Parto institucional	Sim, Não
Zona de Residência	Urbano, Rural
Educação materna em níveis	Nenhuma, Primária, Secundária
Escore do índice de bens	[-25,0:50,0]
País	Albania, Bangladesh, Egito, Etiópia, Guiana, Honduras, Malawi, Senegal, Tanzania, etc...
Ano	[1993:2015]

Durante o detalhamento das análises de cada amostra é referenciada a distribuição das classes do desfecho em cada uma delas, presente na Tabela 8.

Tabela 8 – Distribuição das classes do desfecho de AQP

Atendimento Qualificado ao Parto		
Amostra	Proporção 'não'	Proporção 'sim'
Exploratória	28,56%	71,44%
Principal	35,61%	64,39%
Específica	81,12%	18,88%
Completa	41,52%	58,48%

4.1.1 AMOSTRA EXPLORATÓRIA

Dos 25 inquéritos aleatoriamente selecionados, um inquérito foi removido devido à ausência de informação para o desfecho, resultando em uma amostra com 205.042 crianças. A proporção de crianças que tiveram um Atendimento Qualificado no parto foi de 71,9% e o nível de concordância do principal preditor com o desfecho foi avaliado em 95,17%.

Os resultados preliminares dessa amostra, obtidos com a técnica de árvore de decisão implementada pelo algoritmo J48, foram submetidos ao Congresso Internacional de Epidemiologia¹. O modelo atingiu 96,1% de acurácia no conjunto de validação, valor Kappa de 90,5 e uma distribuição similar de falsos positivos e falsos negativos. Esse desempenho acrescentou 0,9 pontos percentuais sobre a concordância apresentada pelo preditor principal na Tabela 6. O apêndice C ilustra a versão final submetida.

¹ O Congresso Internacional de Epidemiologia ocorreu em agosto de 2017 no Japão, e o trabalho submetido foi aceito na modalidade pôster.

Com o término das análises para essa amostra, observou-se que a acurácia obtida ainda no processo exploratório não foi superada. As técnicas de árvore de decisão e XGB empataram com a melhor acurácia em 96,09%, um ganho de 0,92 pontos percentuais sob a acurácia do preditor principal. Apesar do desempenho entre as técnicas ter sido muito similar com a utilização do atributo de renda absoluta, o mesmo apresentou um leve acréscimo de desempenho nas árvores de decisão, incluindo a obtenção da melhor sensibilidade com 97,09%. O melhor Kappa foi observado também na técnica XGB, sem a incorporação do atributo de renda absoluta. Curiosamente, todos os modelos tiveram uma piora no desempenho da métrica de especificidade ao comparar com o preditor principal, que sozinho obteve 96,82%. A Tabela 9 apresenta todas as estatísticas para cada técnica.

Tabela 9 – Resultados para o desfecho AQP na amostra exploratória

Técnica	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta
	Acurácia		Kappa		Sensibilidade		Especificidade	
Base	95,17%	-	88,53	-	94,51%	-	96,82%	-
RNA	96,05%	96,03%	90,39	90,39	96,84%	96,44%	94,09%	95,01%
AD	96,08%	96,09%	90,44	90,45	96,88%	97,09%	94,07%	93,60%
KNN	95,96%	95,92%	90,14	90,05	96,87%	96,85%	93,66%	93,61%
XGB	96,09%	96,09%	90,48	90,46	96,88%	97,03%	94,13%	93,75%

4.1.2 AMOSTRA PRINCIPAL

A segunda amostra identificou 47 inquéritos conduzidos entre 2010 e 2014, dos quais dois foram excluídos por falta de informação para o desfecho. O banco de dados construído consistiu de 645.351 crianças das quais 415.606 (64,4%) tiveram o parto com atendimento qualificado.

Apesar de possuir uma gama maior de inquéritos em comparação com a amostra exploratória, o desempenho dessa amostra foi inferior tanto ao comparar as métricas individualmente, quanto ao comparar o ganho obtido sob o preditor principal. Todavia, uma diferença entre as amostras foi que o impacto do atributo de renda absoluta foi superior na amostra principal, visto que todas as métricas com o atributo foram superiores quando estimadas sem o mesmo. O desempenho das métricas também apresentou um padrão muito similar a amostra exploratória. A técnica XGB foi a vencedora na métrica de acurácia, com 95,39%, e também na estatística Kappa, com 89,99. Novamente a melhor sensibilidade ficou com a árvore de decisão, em 96,1% e a especificidade de todas as técnicas foi inferior a especificidade do preditor principal, com 95,98%. Todas as métricas podem ser observadas na Tabela 10.

Tabela 10 - Resultados para o desfecho AQP na amostra principal

Técnica	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta
	Acurácia		Kappa		Sensibilidade		Especificidade	
Base	94,73%	-	88,7	-	94,04%	-	95,98%	-
RNA	95,23%	95,37%	89,65	89,94	95,77%	95,97%	94,26%	94,29%
AD	95,20%	95,36%	89,59	89,9	95,81%	96,10%	94,11%	94,02%
KNN	94,72%	94,98%	88,65	89,08	94,21%	95,84%	95,64%	93,43%
XGB	95,21%	95,39%	89,64	89,99	95,53%	95,87%	94,64%	94,52%

4.1.3 AMOSTRA ESPECÍFICA

Diferentes das demais amostras, a perspectiva de obter um modelo generalizado foi substituída por uma perspectiva de especialização. Nesta abordagem foram utilizados 5 inquéritos de um mesmo país, um estudo de caso de Bangladesh, situados em diferentes pontos no tempo: 1996, 1999, 2004, 2007

e 2011. O inquérito de 1993 foi removido da amostra por não conter informação disponível para AQP. Ao total, 38.706 crianças foram contabilizadas no banco de dados, das quais apenas 18,9% teve seu parto conduzido através de um atendimento qualificado.

O estudo de caso obteve um desempenho frustrante, independente de se tratar do contexto de um único país, como pode ser observado na Tabela 11. Sem o atributo de renda absoluta, apenas a técnica de redes neurais artificiais obteve um desempenho minimamente superior ao preditor principal, com 96,23% de acurácia contra 96,17%. As técnicas de árvores de decisão e XGB, com o auxílio do atributo externo, alcançaram um resultado idêntico, porém com um ganho de apenas 0,09 pontos percentuais sobre a acurácia do preditor principal, o que implica em aprendizado quase nulo sobre o desempenho já existente. A especificidade de todas técnicas foi exatamente igual, dado que apenas a sensibilidade foi afetada pelos demais atributos.

Tabela 11 - Resultados para o desfecho de AQP na amostra específica

Técnica	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta
	Acurácia		Kappa		Sensibilidade		Especificidade	
Base	96,17%	-	86,54	-	80,24%	-	99,89%	-
RNA	96,23%	96,23%	86,76	86,76	80,55%	80,55%	99,89%	99,89%
AD	96,17%	96,26%	86,54	86,87	80,24%	80,70%	99,89%	99,89%
KNN	96,17%	96,17%	86,54	86,54	80,24%	80,24%	99,89%	99,89%
XGB	96,17%	96,26%	86,54	86,87	80,24%	80,70%	99,89%	99,89%

4.1.4 AMOSTRA COMPLETA

A última amostra foi uma tentativa de avaliar a temporalidade junto com a variação no contexto dos países, consistindo em uma amostra muito maior que as demais. Esse conjunto de inquéritos totalizou 1.603.282 crianças ao incluir

todos os 177 inquéritos com informação disponível para AQP, com o inquérito mais antigo datado em 1993 e o mais recente em 2015.

A utilização exclusiva do preditor principal obteve 92,86% de acurácia, uma estimativa um pouco abaixo das estimativas de base das outras amostras. Como pode ser observado na Tabela 12, os resultados dessa amostra apontaram as árvores de decisão com o auxílio do atributo de renda absoluta como o melhor modelo. Seu desempenho foi superior em acurácia e no valor Kappa, visto que na sensibilidade ficou apenas 0,25% a menos que KNN, e a especificidade teve o desempenho do preditor principal superior a todos os modelos gerados pelas demais técnicas. O ganho de acurácia foi em torno de um ponto percentual e a sensibilidade chegou a 3,3 pontos percentuais de diferença. Essa amostra apresentou dois resultados completamente destoantes dos demais, com KNN e XGB produzindo modelos com menos de 60% de acurácia. Curiosamente, um dos resultados destoantes foi apenas com atributos originais, para KNN, que com a inclusão da renda absoluta obteve o segundo melhor modelo gerado, e outro resultado fora dos padrões foi apresentado pela técnica de XGB, dessa vez com a inclusão do atributo de renda absoluta. Todavia, o resultado da técnica XGB obtido apenas com atributos originais atingiu um desempenho próximo do preditor principal, o que não indica um padrão claro na geração destes modelos.

Tabela 12 - Resultados para o desfecho de AQP na amostra completa

Técnica	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta	Originais	Originais & renda absoluta
	Acurácia		Kappa		Sensibilidade		Especificidade	
Base	92,86%	-	85,54	-	89,83%	-	97,11%	-
RNA	93,43%	93,71%	86,55	87,12	92,60%	92,91%	94,60%	94,84%
AD	93,86%	93,97%	87,43	87,65	93,01%	93,13%	95,06%	95,16%
KNN	58,67%	93,95%	0,01	87,6	1,00%	93,38%	0,01%	94,77%
XGB	92,70%	59,71%	85,17	26,3	90,27%	35,34%	96,16%	94,31%

4.2 ANÁLISES POR DÉFICIT DE ALTURA POR IDADE

Diferente do desfecho apresentado anteriormente, o desfecho de Déficit de Altura por Idade não possui um preditor com alta concordância disponível. É importante destacar que todos os atributos presentes no módulo de antropometria do inquérito foram descartados devido a aplicabilidade do problema. Isto é, a ausência de informação para Déficit de Altura ocorre quando o inquérito não coleta o módulo de antropometria. Utilizar atributos que não estarão disponíveis nos conjuntos de dados futuros invalida a aplicação dos modelos preditivos. Desta forma, os dez atributos apresentados a seguir foram identificados como os preditores para o desfecho de interesse, além dos atributos país e ano, e do atributo externo de renda absoluta apresentado separadamente. Além disso, a Tabela 13 mostra os valores de cada atributo utilizado neste desfecho.

- Fonte de água para beber: Atributo que indica o tipo de fonte de água disponível no domicílio para consumo. O número e os tipos das categorias variam de acordo com a amostra, dado que os inquéritos possuem categorizações diferentes.
- Tipo de instalação sanitária: Atributo que indica o sistema de evacuação dos dejetos do domicílio.
- Área de residência: Atributo que indica se o domicílio está situado na zona urbana ou rural.
- Tipo de combustível para cozinhar: Atributo que indica o tipo de combustível utilizado nos domicílios para cozinhar. As categorias “não cozinha”, “não sabe” ou sem informação foram agrupadas.
- Educação materna em anos: Atributo numérico que indica o número de anos de estudo completos da mãe. Não sabe ou informação incompleta foi considerada como 0.

- Educação paterna em anos: Atributo numérico que indica o número de anos de estudo completos do pai. Não sabe ou informação incompleta foi considerada como 0.
- Peso ao nascer em gramas: Atributo contínuo que indica o peso em gramas da criança ao nascer. Existem uma associação na literatura indicando que crianças com baixo peso ao nascer são mais propensas a desnutrição. Valores acima de 10kg foram descartados.
- Alguma vez amamentada: Atributo binário que indica se a criança recebeu leite materno pelo menos uma vez na vida.
- Ainda amamenta: Atributo binário que indica se a criança ainda está sendo amamentada.
- Recebeu 4 ou mais consultas pré-natal: Atributo binário que indica se a mãe recebeu 4 ou mais consultas durante a gravidez de determinada criança.

Tabela 13 - Atributos utilizados nas análises de Déficit de Altura por Idade

Déficit de Altura por Idade	
Atributos	Valores
Fonte de água para beber	Água encanada, torneira pública, fonte protegida, poço, água da chuva, água engarrafada, rio/lago, etc...
Zona de Residência	Urbano, Rural
Tipo de instalação sanitária	Descarga para o sistema de esgoto, descarga para fossa séptica, latrina, arbustos, etc..
Escore do índice de bens	[-25,0:50,0]
Tipo de combustível para cozinhar	Eletricidade, gás natural, madeira, esterco, carvão, etc..

Educação materna em anos	[0:22]
Educação paterna em anos	[0:22]
Peso ao nascer em gramas	[100:8.800]
Alguma vez amamentada	Sim, Não
Ainda amamenta	Sim, Não
Recebeu quatro ou mais consultas pré-natal	Sim, Não
País	Albania, Bangladesh, Egito, Etiópia, Guiana, Honduras, Malawi, Senegal, Tanzania, etc...
Ano	[1993:2015]

Outra particularidade deste desfecho está no fato de que para todas as amostras, o Déficit de Altura por Idade ser uma classe desbalanceada. A proporção de cada classe do desfecho pode ser visualizada na Tabela 14. Esta ocorrência dificulta o aprendizado da classe minoritária pelas técnicas de aprendizado de máquina, visto que existirão muito mais exemplos negativos disponíveis para extrair os padrões. A diferença foi minimizada ao utilizar a técnica SMOTE que expande a classe minoritária para obter uma proporção mais equivalente. Durante as análises iniciais com a amostra exploratória foram avaliados diferentes parâmetros para a técnica SMOTE, cujo resultados são apresentados na Tabela 15. O SMOTE com valor 100 foi descartado pois apesar de obter a melhor sensibilidade para ambas técnicas, sua acurácia foi muito inferior as demais. Já o SMOTE 200, apesar de sensibilidade também superior, apenas foi destaque no valor Kappa para redes neurais e obteve acurácias menores, mesmo que com diferenças menos significantes. A principal dúvida foi a decisão entre os valores 300 e 400, na qual 300 foi escolhido por ter um Kappa superior em ambas técnicas, dado que a melhor acurácia dependia da técnica escolhida.

Tabela 14 – Distribuição das classes do desfecho de Déficit de Altura por Idade

Déficit de Altura por Idade		
Amostra	Proporção 'não'	Proporção 'sim'
Exploratória	71,84%	28,16%
Principal	66,40%	33,60%
Específica	52,34%	47,66%
Completa	65,07%	34,93%

Tabela 15 – Parâmetros de SMOTE testados na amostra exploratória para Déficit de Altura por Idade

Métrica	Técnica	SMOTE	SMOTE	SMOTE	SMOTE
		100	200	300	400
Acurácia	RNA	63,18	69,68	70,38	68,8
	AD	62,72	67,1	71,13	72,07
Kappa	RNA	23,77	25,18	24,59	24,33
	AD	17,29	14,69	15,29	14,69
Sensibilidade	RNA	69,18	50,91	47,28	51,98
	AD	55,37	37,98	27,99	24,45
Especificidade	RNA	61,14	76,05	78,23	74,52
	AD	65,22	76,99	85,78	88,25

Diferente do primeiro desfecho, os resultados para Déficit de Altura por Idade foram estimados de três maneiras diferentes: apenas com os atributos originais, atributos originais e renda absoluta e todos os atributos com a técnica de SMOTE aplicada.

4.2.1 AMOSTRA EXPLORATÓRIA

Apenas um dos 25 inquéritos selecionados aleatoriamente não possuía informação disponível para o desfecho, resultando em uma amostra com 119.218 crianças. Entretanto, como este desfecho possui diversos preditores e os mesmos contém observações incompletas, o tamanho amostral é reduzido. Por exemplo, o banco de dados de treinamento consiste de 96.567 crianças com informação disponível para o desfecho. Deste número, apenas 68.841 crianças, ou 70% da amostra, foram utilizadas no processo de treinamento dos modelos devido a ausência de informação para algum dos preditores.

A primeira amostra avaliada com o segundo desfecho teve um desempenho superior a todas as outras amostras. O modelo obtido através de árvores de decisão utilizando atributos originais e renda absoluta foi consideravelmente superior aos demais. Sua acurácia atingiu 79.54%, mais de três pontos percentuais acima da mesma técnica sem renda absoluta, que foi a segunda melhor acurácia obtida. O valor Kappa foi mais do que o dobro do que todas as outras técnicas, com exceção de redes neurais com SMOTE que a diferença foi de dez pontos percentuais. A sensibilidade foi bem superior nas redes neurais com SMOTE, a única técnica e conjunto de atributos que chegou próximo de 50%. A aplicação da técnica XGB com SMOTE construiu um modelo extremo, cujo resultado foi classificar todas as observações como negativas, atingindo 100% de especificidade e 0 de sensibilidade e Kappa, conseqüentemente. É importante salientar que ao observar apenas a acurácia desse modelo, que foi de 74,64%, o mesmo poderia não ser descartado devido a um desempenho aparentemente próximo dos demais. A Tabela 16 apresenta todos os resultados produzidos para todas as técnicas e combinações de atributos.

Tabela 16 - Resultados para o desfecho de Déficit de Altura por Idade na amostra exploratória

Técnica	Originals	Originals	Originals	Originals	Originals	Originals	Originals	Originals	Originals	Originals	Originals	Originals
	& renda	& renda	& renda	& renda	& renda	& renda	& renda	& renda	& renda	& renda	& renda	& renda
	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta	absoluta
Acurácia			Kappa			Sensibilidade			Especificidade			
RNA	74,97%	75,17%	70,38%	12,98	13,31	24,59	14,27%	14,18%	47,28%	95,59%	95,89%	78,23%
AD	76,34%	79,54%	71,13%	16,83	34,33	15,29	15,71%	32,82%	27,99%	96,94%	95,42%	85,78%
KNN	74,29%	74,16%	70,34%	15,23	15,27	16,03	19,20%	19,62%	31,38%	93,00%	92,69%	83,57%
XGB	75,43%	75,51%	74,64%	11,08	14,21	0	10,65%	14,46%	0,00%	97,44%	96,26%	100%

4.2.2 AMOSTRA PRINCIPAL

A amostra principal reuniu 60 inquéritos diferentes conduzidos entre 2010 e 2015 que possuíam informação sobre antropometria coletada. Isto resultou em 357.083 crianças distribuídas em 33,6% com a condição de Déficit de Altura por Idade presente e 66,4% sem ela.

Em comparação com a amostra exploratória, o desempenho dessa amostra foi significativamente inferior, dado que ambas utilizaram os mesmos critérios, mas incluíram números diferentes de inquéritos. A melhor acurácia foi através da técnica XGB utilizando apenas atributos originais com 70,85%, quase dez pontos percentuais abaixo da amostra anterior. As redes neurais com SMOTE obtiveram o melhor Kappa e sensibilidade com 25,4 e 43,54%, respectivamente. Novamente pôde ser encontrado um novo modelo extremo, desta vez com KNN, reproduzindo todas as observações como negativas, e atingindo uma acurácia apenas 1,7 pontos percentuais abaixo da melhor. Desconsiderando a especificidade máxima obtida pelo KNN, a técnica XGB novamente ficou na frente com 93,95%. Todos esses resultados podem ser observados na Tabela 17.

Tabela 17 - Resultados para o desfecho de Déficit de Altura por Idade na amostra principal

Técnica	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE
	Acurácia			Kappa			Sensibilidade			Especificidade		
	RNA	70,81%	70,78%	69,37%	17,53	18,91	25,4	21,87%	24,56%	43,54%	92,67%	91,41%
AD	70,60%	70,56%	66,51%	18,9	19,85	13,87	25,16%	27,13%	29,70%	90,89%	89,95%	82,94%
KNN	69,13%	69,76%	66,83%	0	18,24	16,62	0,00%	26,93%	33,99%	100%	88,88%	81,49%
XGB	70,85%	70,60%	69,90%	16,88	15,09	14,84	20,57%	18,30%	20,19%	93,29%	93,95%	92,10%

4.2.3 AMOSTRA ESPECÍFICA

O estudo de caso realizado com os inquéritos de Bangladesh para este desfecho contou com 6 pontos no tempo. O inquérito realizado em 1993, cuja informação para AQP estava indisponível, foi incluído nesta análise por ter coletado informações antropométricas. Ao todo, 29.869 crianças compuseram a amostra que foi submetida aos processos de mineração de dados. O atributo que indica baixo peso ao nascer, um atributo considerado como um importante preditor de Déficit de Altura por Idade, não foi coletado em nenhum dos inquéritos de Bangladesh, o que pode implicar em um menor desempenho.

O desempenho dessa amostra foi inferior as demais, principalmente se avaliarmos a acurácia. Com 62,35% de acurácia, a maior estimativa dessa amostra ficou 8,5 pontos percentuais abaixo da amostra principal e 17 pontos percentuais abaixo da exploratória. Em compensação, os valores Kappa atingiram valores próximos das demais com seu pico em 23,42. Isto se deve ao fato de esta amostra ter obtido alta sensibilidade na maior parte de seus resultados. Com ápice em 56,14% utilizando árvores de decisão, esta amostra obteve desempenho superior as demais nesta métrica. A especificidade, em contrapartida, foi bem abaixo do

encontrado anteriormente, com exceção da técnica XGB com SMOTE, que atingiu 90.98%, as demais ficaram em torno de 65 a 75%. A Tabela 18 apresenta todos os resultados da amostra específica.

Tabela 18 - Resultados para o desfecho de Déficit de Altura por Idade na amostra específica

Técnica	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE
	Acurácia			Kappa			Sensibilidade			Especificidade		
	RNA	60,53%	61,34%	59,81%	19,41	21,49	16,34	49,95%	53,98%	38,52%	69,24%	67,39%
AD	60,41%	60,58%	58,71%	19,06	20,37	14,43	49,20%	56,14%	39,74%	69,62%	64,23%	74,33%
KNN	59,86%	59,73%	57,36%	18,4	18,05	11,01	51,73%	50,98%	34,40%	66,54%	66,92%	76,25%
XGB	61,13%	62,35%	58,12%	21,23	23,42	9,78	54,92%	54,17%	18,18%	66,23%	69,08%	90,98%

4.2.4 AMOSTRA COMPLETA

A maior amostra desse desfecho foi composta de 910.173 crianças oriundas de 156 inquéritos diferentes, representando 64 países.

Os resultados dessa amostra foram similares aos resultados da amostra principal, apresentando uma acurácia em torno de 70%. As duas técnicas que se destacaram nessa amostra foram XGB e redes neurais, com vantagem para XGB na acurácia e especificidade e redes neurais com melhor Kappa e sensibilidade. A diferença mais significativa foi da própria sensibilidade, com 53,21% para redes neurais utilizando SMOTE, que foi muito superior as demais. Esta diferença também elevou o Kappa desta combinação para 26,09, que foi o mais alto da amostra. Todas as estimativas estão presentes na Tabela 19.

Tabela 19 - Resultados para o desfecho de Déficit de Altura por Idade na amostra completa

Técnica	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE	Originals	Originals & renda absoluta	SMOTE
	Acurácia			Kappa			Sensibilidade			Especificidade		
RNA	70,06%	70,59%	67,06%	18,21	22,25	26,09	23,36%	29,57%	53,21%	91,97%	89,81%	73,55%
AD	69,84%	69,91%	67,50%	20,32	22,06	10,32	28,39%	31,86%	17,85%	89,30%	87,74%	90,76%
KNN	69,22%	69,21%	66,81%	20,64	20,78	13,98	31,47%	31,90%	26,90%	86,94%	86,69%	85,51%
XGB	70,28%	70,90%	69,15%	18,71	22,35	10,29	23,50%	28,50%	12,42%	92,23%	90,77%	95,73%

4.3 VALIDAÇÃO DOS MELHORES MODELOS

Os resultados apresentados durante as análises de ambos os desfechos são estimativas extraídas do processo de mineração de dados aplicado aos conjuntos de dados de treinamento cujo modelos produzidos foram aplicados nos conjuntos de dados de validação. Entretanto, mesmo mantendo este conjunto de dados de validação separado das observações utilizadas para treinar os modelos, ainda existe o risco de sobreajuste por conta da grande quantidade de técnicas com diferentes parâmetros aplicados sobre os mesmos. Para avaliar o desempenho real destes modelos em dados ainda não vistos, um terceiro conjunto de dados foi mantido separado durante todo o processo. Apenas os melhores modelos de cada amostra foram aplicados a este conjunto de dados, cujo resultados podem ser visualizados na Tabela 20 para o desfecho de AQP e na Tabela 21 para o desfecho de Déficit de Altura por Idade. Destaca-se que escolha dos melhores modelos foi conduzida de forma arbitrária, dado que cada usuário pode priorizar diferentes métricas com base em suas necessidades.

Devido a variação do desempenho do principal no conjunto de validação e no conjunto final de testes, os resultados para o desfecho de AQP foram apresentados junto com o desempenho do preditor principal nesta amostra.

A análise dos resultados dos melhores modelos para o primeiro desfecho apresentou resultados muito similares aos encontrados sobre o conjunto de validação. As diferenças se devem a variação amostral cuja ocorrência também pode ser observada no desempenho do preditor principal.

Tabela 20 – Resultados dos melhores modelos para o desfecho de AQP

Amostra	Técnica	Composição	Acurácia	Kappa	Sensibilidade	Especificidade
Exploratória	Base	-	95,53%	89,38	94,89%	97,19%
	XGB	Originais	96,23%	90,79	97,20%	93,82%
Principal	Base	-	94,63%	88,48	94,02%	95,70%
	XGB	Originais e renda absoluta	95,21%	89,61	95,8%	94,15%
Específica	Base	-	95,69%	84,77	78,54%	99,68%
	XGB	Originais e renda absoluta	95,69%	84,81	78,89%	99,61%
Completa	Base	-	92,91%	85,65	89,94%	97,09%
	AD	Originais e renda absoluta	93,93%	87,58	93,15%	95,05%

Na avaliação do segundo desfecho, os resultados de todas as amostras tiveram uma queda no desempenho quando validados sobre o conjunto de testes. Isso indica que houve algum tipo de sobreajuste mesmo utilizando um conjunto de validação separado durante as análises. A queda mais drástica ficou por conta do algoritmo J48 na amostra exploratória. O mesmo já apresentava um desempenho significativamente superior aos demais e conseqüentemente apresentou uma redução de 4,5 pontos percentuais de acurácia. As demais amostras tiveram uma redução entre 1,5 e 2,5 pontos percentuais, em média, mensurando pela acurácia.

Tabela 21 – Resultados dos melhores modelos para o desfecho de Déficit de Altura por Idade

Amostra	Técnica	Composição	Acurácia	Kappa	Sensibilidade	Especificidade
Exploratória	AD	Originais e Renda Absoluta	75%	15.67	16.95%	95.18%
Principal	RNA	SMOTE	67.26%	25.15	52.16%	73.84%
Específica	XGB	Originais e Renda Absoluta	60.47%	19.95	54.53%	65.36%
Completa	RNA	SMOTE	68.47%	26.64	49.64%	77.02%

4.4 DISCUSSÃO DOS RESULTADOS

Como já previsto, o desfecho de Atendimento Qualificado ao Parto teve seu desempenho guiado pelo preditor principal. O adicional obtido pelos demais atributos foi pouco significativo, com uma média de um ponto percentual a mais de acurácia e entre uma e duas unidades do valor Kappa. A amostra exploratória obteve o melhor desempenho se comparado com as demais, visto que apesar da amostra específica apresentar uma mínima vantagem na acurácia e praticamente 100% de especificidade, esta conclusão pode ser obtida ao avaliar o valor Kappa de ambas amostras e principalmente o desempenho sobre o preditor principal. Em contrapartida, as métricas elevadas apresentadas pela amostra específica são apenas um reflexo da significância deste preditor, que dominou o desempenho e resultou em modelos de pouca ou nenhuma expressão. Apesar disso, a amostra específica apresentou a maior concordância esperada, o que implica em uma maior chance de acerto ao acaso. Isso explica o valor do Kappa ser inferior se comparado com as outras amostras, mesmo com uma acurácia equivalente ou superior. Essa diferença na concordância esperada, parcialmente explicada pelo desbalanceamento entre as classes do desfecho, pode ter afetado a influência dos demais atributos na predição dos modelos. Todavia, testes foram realizados aplicando a técnica de SMOTE para a amostra específica e não apresentaram

resultados promissores. Estes testes, contudo, foram conduzidos de forma simplificada e sua expansão é discutida como uma possibilidade para trabalhos futuros. Já a amostra completa, apesar do desempenho inferior as demais, apresentou um ganho sobre o preditor principal levemente superior até ao da amostra exploratória. Entretanto, isso pode ser devido ao fato de que com um desempenho médio inferior, esta amostra possui maior margem para melhoria.

Os resultados obtidos para AQP também mostram que, apesar de poucas exceções, o atributo de renda absoluta teve um impacto positivo na produção dos modelos. Entretanto, ao comparar as estimativas que incluíram este atributo, as estimativas que não o incluíram e as estimativas do preditor principal é possível observar que o maior ganho de desempenho não foi proveniente deste atributo.

As análises para o desfecho de Déficit de Altura por Idade também apresentaram melhores resultados para a amostra exploratória, com seu ápice de acurácia atingindo cerca de 80%. Diferentemente do primeiro desfecho, os resultados foram conduzidos com três abordagens diferentes e não possuíram uma base de desempenho para comparação. Com exceção de algumas métricas, a abordagem utilizando apenas os atributos originais não foi superior as demais, apesar de que em muitos casos a significância do atributo de renda absoluta foi muito baixa. A técnica SMOTE, apesar de aumentar significativamente a sensibilidade dos modelos em algumas amostras, impactou pouco no geral, apresentando melhores resultados com a técnica de redes neurais artificiais.

Dentre os possíveis motivos encontrados para o baixo desempenho do segundo desfecho estão a disponibilidade de determinados atributos preditores. Por exemplo, baixo peso ao nascer e o número de visitas pré-natal contribuíram para uma grande redução no tamanho de amostra.

5 TRABALHOS RELACIONADOS

O potencial da inteligência artificial, mais especificamente a subárea de aprendizado de máquina, vem sendo requerido em diversos campos de estudo, entre eles na área da saúde, em aplicações com alto volume de dados e que carecem de sistemas inteligentes, tais como a identificação de causas de doenças, auxílio no diagnóstico de doenças crônicas, detecção de fraudes em seguros de saúde, e muitos outros (Kaur e Wasan, 2006; Tomar e Agarwal, 2013; Yoo *et al.*, 2012).

Muitas das aplicações buscam nessas ferramentas um método resolutivo, que entregue uma resposta para determinada situação, como um diagnóstico, e o aplicam instantaneamente a medida que novos dados são coletados. Alternativamente, existem casos onde a predição visa produzir novos dados e, apesar de utilizar a mesma estratégia determinística, tem um propósito descritivo e de uso não imediato. Um exemplo desses casos é a produção de informações não coletadas em inquéritos populacionais de saúde. Visto que os dados são representativos de um determinado país ou região, a distinção entre indivíduos serve o propósito de informar os subgrupos com necessidades de maior atenção ao formular políticas de saúde.

O inquérito nacional da Etiópia, conduzido no ano de 2011 pela DHS, serviu como base para alguns estudos desenvolvidos nessa área. (Abegaz e Atomssa, 2017) investigaram a possibilidade de prever mulheres em idade reprodutiva

que receberam a vacina de toxóide tetânico. Com uma amostra de aproximadamente sete mil mulheres, o desempenho dos algoritmos foi entre 60 e 65% de acurácia. Já o estudo de (Setotaw, 2013) avaliou a presença de anemia em mulheres de 15 a 49 anos. Com modelos baseados em árvores de decisão e conjunto de regras, a acurácia alcançada pelos melhores modelos foi superior a 90%. Ainda no contexto dos etíopes, o estudo de (Markos, 2014) avaliou o estado nutricional das crianças com uma proposta de classificação em 4 categorias: normal e três tipos diferentes de desnutrição. Três algoritmos diferentes foram aplicados, onde o algoritmo baseado em regras PART obteve acurácia de 92,6%. Apesar do estudo abordar os indicadores de estado nutricional que incluem o Déficit de Altura por Idade como desfecho, este estudo não é comparável com este trabalho pela utilização dos atributos de peso, altura e derivados como preditores. Visto que a coleta das medidas antropométricas permite a classificação direta das variantes de desnutrição, a aplicação dos modelos em novos inquéritos onde a coleta não foi realizada é impossibilitada.

Outro país que a literatura também demonstrou interesse foi a Índia. (Khare *et al.*, 2017) apresentou análises baseadas no inquérito DHS da Índia conduzido no ano de 2005-2006 também com foco nos quatro indicadores de estado nutricional. Apesar de utilizar uma metodologia que permite pouca generalização dos modelos gerados, os resultados de Déficit de Altura por Idade podem ser comparados com este trabalho. 43.734 crianças constavam no banco de dados das crianças com informações para os desfechos estudados. Os resultados foram apresentados para os 4 desfechos em 2 abordagens diferentes para cada: atributos com base na literatura e atributos selecionados por ganho de informação. As acurácias de cada desfecho variaram entre 76,7% a 82% para Índice de Massa Corporal (IMC) por idade, 60,4% e 63,4% para Déficit de Altura por Idade, 70,2% e 96,8% para Déficit de Peso por Idade e 76,8% e 82,6% para Déficit de Peso por Altura. A Índia, através de um inquérito nacional de familiar conduzido no ano de 2007, foi alvo de mais um estudo sobre o estado nutricional

das crianças. Com base em 254 crianças, (Thangamani e Sudha, 2014) aplicou as técnicas de árvores de decisão, florestas aleatórias e redes neurais artificiais com a melhor acurácia atingindo 77,2%.

O grande diferencial deste trabalho em relação aos demais está na utilização de amostras formadas por diversos inquéritos nacionais. Além do aumento significativo do tamanho amostral, combinar diferentes países ou períodos de tempo minimiza a possível especialização em atributos específicos de determinado inquérito, forçando a obtenção de um modelo mais generalizado. Outro ponto importante está na exploração de diferentes métricas de avaliação dos resultados obtidos. Como foi possível observar em alguns resultados, principalmente para o desfecho de Déficit de Altura por Idade, existe a possibilidade de obter um modelo com uma acurácia alta, superior a diversos modelos concorrentes em alguns casos, que apenas replicou a classe majoritária. A detecção destes modelos extremos pode não ocorrer se o estudo não avaliar a matriz de confusão gerada, as métricas de sensibilidade e especificidade ou até mesmo o valor Kappa.

6 CONCLUSÕES

Indiscutivelmente, ter dados relevantes, disponíveis e de boa qualidade é fundamental no monitoramento do progresso dos países. Visto que nos países de baixa e média renda, devido a precariedade dos sistemas de registro civil e estatísticas vitais, os inquéritos populacionais de saúde são a principal fonte de informação, e extrair o máximo de informação de tais fontes é um desafio constante. Através de desfechos padronizados, as organizações internacionais, em parceria com os governos dos países, avaliam a situação das intervenções e cuidados de saúde, indicando onde novos investimentos são necessários.

O aprendizado de máquina surge como uma alternativa para agregar novos dados com base nas características e informações já presentes nestes inquéritos. Uma vez que tais desfechos de saúde já são consolidados e aceitos na literatura, garantir a máxima disponibilidade dos mesmos foi a base para a condução das análises deste trabalho.

Os dois desfechos abordados apresentaram resultados com comportamentos diferentes. Os modelos para Atendimento Qualificado ao Parto foram construídos com forte dependência do preditor principal de parto institucional. O foco, exceder o ganho sobre este preditor, foi cumprido, mas com aumentos limitados. Ciente da presença de ruídos nos inquéritos, e possuindo uma pequena margem de ganho sobre o determinado preditor, a significância

deste ganho é difícil de ser mensurada. De qualquer forma, o uso de tais estimativas para este desfecho em aplicações é recomendado, dado que os modelos produzidos foram estáveis e obtiveram algum ganho sobre o preditor principal. Quanto as amostras, os modelos mais generalizados, com um maior número de países, apresentaram melhores resultados se comparados com as amostras que apostaram na temporalidade. Tanto o estudo de caso quanto a amostra completa, com múltiplos pontos no tempo, apresentaram resultados inferiores. O motivo da amostra exploratória ter atingido um desempenho superior a amostra principal, visto que ambas utilizaram critérios de construção similares, pode estar no fato de que a amostra exploratória foi composta por mais países ricos. Se analisarmos tanto a cobertura de AQP quanto a prevalência de Déficit de Altura por Idade, é possível observar que a cobertura é maior e a prevalência é menor na amostra exploratória, o que coincide com os padrões encontrados de acordo com a riqueza em análises globais de ambos os desfechos (Victoria *et al.*, 2012).

Já as análises conduzidas para o segundo desfecho, Déficit de Altura por Idade, obtiveram resultados bem inferiores se comparados com o primeiro. As informações incompletas de outros atributos utilizados como preditores são uma possível razão para esta ocorrência. A ausência de um preditor correlacionado pode ser considerada a principal diferença entre o desempenho de ambos desfechos.

Quanto as técnicas de aprendizado de máquina aplicadas no processo, não foi possível identificar nenhum desempenho notável entre elas. O desempenho do KNN, em média, manteve-se abaixo das demais, o que indica que não é uma técnica apropriada para a resolução deste tipo de problema.

Dado que não existe um valor exato na literatura sobre qual seria o desempenho mínimo aceitável para a utilização de dados preditos, esse limite foi o auxílio de especialistas. De acordo com pesquisadores do Centro Internacional

de Equidade em Saúde, da Universidade Federal de Pelotas, estimativas que apresentem acurácia em torno de 90% são suficientes para sua utilização em análises globais e regionais. Utilizar tais dados como estimativas únicas de determinado inquérito ou avaliando padrões entre indivíduos requer análise mais detalhada das demais métricas e do direcionamento dos erros apresentado por cada modelo.

Mesmo com o avanço no desempenho de componentes de hardware nas últimas décadas que viabilizou a utilização de diversos processos computacionalmente custosos, incluindo as técnicas de aprendizado de máquina, a magnitude dos problemas estudados também cresceu. Durante o desenvolvimento deste trabalho foram utilizados recursos de diferentes composições de hardware, visto que alguns algoritmos nas amostras mais pesadas não foram suportados por computadores pessoais. A necessidade de buscar recursos mais poderosos impactou no tempo necessário para produzir todos os resultados apresentados, assim como limitou o número de análises conduzidas.

6.1 TRABALHOS FUTUROS

Durante as análises exploratórias, algoritmos que implementam uma técnica de aprendizado de máquina conhecida como Máquina de Vetor de Suporte, do inglês *Support Vector Machines* (SVM), foram aplicados. Nestes testes, o SVM levou um tempo muito superior aos demais para produzir resultados, e em alguns casos não atingiu a convergência. Visto que SVM é uma técnica consolidada na literatura e sua aplicação costuma ter bons resultados, uma avaliação dos motivos que levaram a estes resultados é desejável.

Uma das razões de utilizar a técnica SMOTE nas análises do segundo desfecho foi o desbalanceamento das classes. Esta diferença também pode ser observada nas análises do primeiro desfecho, principalmente para a amostra

específica. Testes conduzidos nesta amostra indicaram que o SMOTE teve um efeito negativo. Entretanto, a realização de uma análise completa aplicando a mesma metodologia empregada para o segundo desfecho é uma possibilidade de continuação deste trabalho.

REFERÊNCIAS

ABEGAZ, K. H.; ATOMSSA, E. M. Data Mining of Access to Tetanus Toxoid Immunization Among Women of Childbearing Age in Ethiopia. **Machine Learning Research**, v. 2, n. 2, p. 54–60, 2017.

ANALYTICS, R.; WESTON, S. **doSNOW: Foreach Parallel Adaptor for the “snow” Package**, 2015. Disponível em: <<https://cran.r-project.org/package=doSNOW>>

AZEVEDO, A.; SANTOS, M. F. **KDD, SEMMA and CRISP-DM: a parallel overview** | ADIS European Conference on Data Mining. **Anais...** Amsterdam, The Netherlands: 2008

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BROWN, M. L.; KROS, J. F. Data mining and the impact of missing data. **Industrial Management & Data Systems**, v. 103, n. 8, p. 611–621, 2003.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE : Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002.

CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System** 22nd ACM SIGKDD International Conference. **Anais...** 2016 Disponível em: <<http://arxiv.org/abs/1603.02754>> <<http://dx.doi.org/10.1145/2939672.2939785>>

CHEN, T.; HE, T.; BENESTY, M.; KHOTILOVICH, V.; TANG, Y. **xgboost:**

Extreme Gradient Boosting, 2017. Disponível em: <<https://cran.r-project.org/package=xgboost>>

CORSI, D. J.; NEUMAN, M.; FINLAY, J. E.; SUBRAMANIAN, S. V. Demographic and health surveys: A profile. **International Journal of Epidemiology**, v. 41, n. 6, p. 1602–1613, 2012.

DASH, M.; LIU, H. Feature Selection for Classification. **Intelligent Data Analysis**, v. 1, n. 3, p. 131–156, 1997.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, 1996.

FILMER, D.; PRITCHETT, L. Estimating Wealth Effects Without Expenditure Data--Or Tears: An Application to Educational Enrollments in States of India. **Demography**, v. 38, n. 1, p. 115–132, 2001.

FINK, G.; VICTORA, C. G.; HARTTGEN, K.; VOLLMER, S.; VIDALETTI, L. P.; BARROS, A. J. D. Measuring Socioeconomic Inequalities with Predicted Absolute Incomes Rather Than Wealth Quintiles: A Comparative Assessment Using Child Stunting Data from National Surveys. **American Journal of Public Health**, v. 107, n. 4, p. 550–555, 2017.

FRIEDMAN, J. H. GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE. **Statistics**, v. 29, n. 5, p. 1189–1232, 2001.

GAKIDOU, E.; HOGAN, M. Data Collection in Developing Countries. *In: Encyclopedia of social measurement*. First ed. [s.l.] Elsevier Academic Press, 2005. p. 587–592.

GUYON, I.; ELISSEEFF, A. An Introduction to Variable and Feature Selection. **Journal of Machine Learning Research (JMLR)**, v. 3, n. 3, p. 1157–1182, 2003.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. Third ed. [s.l.] Morgan Kaufmann Publishers, 2012.

HOLZINGER, A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? **Brain Informatics**, v. 3, n. 2, p. 119–131, 2016.

HORNIK, K.; BUCHTA, C.; ZEILEIS, A. Open-Source Machine Learning: R Meets Weka. **Computational Statistics**, v. 24, n. 2, p. 225–232, 2009.

KAUR, H.; WASAN, S. K. Empirical Study on Applications of Data Mining Techniques in Healthcare. **(B3) Journal of Computer Sciences**, v. 2, n. 2, p. 194–200, 2006.

KHARE, S.; KAVYASHREE, S.; GUPTA, D.; JYOTISHI, A. **Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data**Procedia Computer Science. **Anais...2017**

KOH, H. C.; TAN, G. Data mining applications in healthcare. **J Healthc Inf Manag**, v. 19, n. 2, p. 64–72, 2005.

KOHAVI, R. **A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection**International Joint Conference on Artificial Intelligence. **Anais...1995**Disponível em: <<http://robotics.stanford.edu/~ronnyk>>

KOHAVI, R.; PROVOST, F. Glossary of Terms. **Machine Learning**, v. 30, p. 271–274, 1998.

KOTSIANTIS, S. B. Supervised Machine Learning: A Review of Classification Techniques. **Informatica**, v. 31, p. 249–268, 2007.

KUHN, M. **Package “caret”: Classification and Regression Training**, 2016. Disponível em: <<https://github.com/topepo/caret/>>

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. **Nature**, v. 521, p. 436–444, 2015.

MANNILA, H. **Data mining: machine learning, statistics, and databases**Proceedings of 8th International Conference on Scientific and Statistical Data Base Management. **Anais...1996**Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=505910>>

MARKOS, Z. Predicting Under Nutrition Status of Under-Five Children Using Data Mining Techniques: The Case of 2011 Ethiopian Demographic and Health Survey. **Journal of Health & Medical Informatics**, 2014.

MITCHELL, T. M. **Machine Learning**. [s.l.] McGraw-Hill Science/Engineering/Math, 1997.

MURRAY, C.; NEWBY, H. Data resource profile: United Nations children’s fund (UNICEF). **International Journal of Epidemiology**, v. 41, n. 6, p. 1595–1601,

2012.

ONIS, M. DE; BLÖSSNER, M. The World Health Organization Global Database on Child Growth and Malnutrition: Methodology and applications. **International Journal of Epidemiology**, v. 32, n. 4, p. 518–526, 2003.

PEDDI, N. **10 most popular Machine Learning algorithms**. Disponível em: <<https://upxacademy.com/10-popular-machine-learning-algorithms/>>.

POWERS, D. M. W. Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. **Journal of Machine Learning Technologies**, v. 2, n. 1, p. 37–63, 2011.

QUINLAN, J. R. Induction of Decision Trees. **Machine Learning**, v. 1, n. 1, p. 81–106, 1986.

REQUEJO, J.; VICTORA, C.; BRYCE, J. Data resource profile: Countdown to 2015: Maternal, newborn and child survival. **International Journal of Epidemiology**, v. 43, n. 2, p. 586–596, 2014.

SETOTAW, G. **PREDICTING THE STATUS OF ANAEMIA IN WOMEN AGED 15-49 BY APPLYING DATA MINING TECHNIQUES USING THE 2011 ETHIOPIA DEMOGRAPHIC AND HEALTH SURVEY (EDHS) DATASETS**. [s.l.] ADDIS ABABA UNIVERSITY, 2013.

SHEARER, C. The CRIS-DM model: The New Blueprint for Data Mining. **Journal of Data Warehousing**, v. 5, n. 4, p. 13–22, 2000.

SOMAN, K. P.; DIWAKAR, S.; AJAY, V. **Insight into Data mining: Theory and practice**. 1. ed. [s.l.] Prentice Hall of India, 2006.

SONG, X.; MITNITSKI, A.; COX, J.; ROCKWOOD, K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. **Studies in Health Technology and Informatics**, 2004.

STATA CORP. **Stata Statistical Software: Release 13** College Station, TX StataCorp LP, , 2013.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. 1. ed. [s.l.] Pearson Education, 2006.

THANGAMANI, D.; SUDHA, P. Identification Of Malnutrition With Use Of

Supervised Datamining Techniques –Decision Trees And Artificial Neural Networks. **International Journal Of Engineering And Computer Science**, v. ISSN, p. 2319–7242, 2014.

The DHS Program. Disponível em: <www.dhsprogram.com>. Acesso em: 18 dez. 2017.

TOMAR, D.; AGARWAL, S. A survey on data mining approaches for healthcare. **International Journal of Bio-Science and Bio-Technology**, 2013.

TORGO, L. **Data Mining with R, learning with case studies** Chapman and Hall/CRC, , 2010. Disponível em: <<http://www.dcc.fc.up.pt/~ltorgo/DataMiningWithR>>

UNITED NATIONS. **Transforming our world: the 2030 Agenda for Sustainable Development** General Assembly 70 session, 2015. Disponível em: <<http://undocs.org/A/RES/70/1>>

VENABLES, W. N.; RIPLEY, B. . **Modern Applied Statistics with S**. Fourth ed. [s.l.] Springer-Verlag New York, 2002.

VICTORA, C. G.; BARROS, A. J. D.; AXELSON, H.; BHUTTA, Z. A.; CHOPRA, M.; FRANÇA, G. V. A.; KERBER, K.; KIRKWOOD, B. R.; NEWBY, H.; RONSMANS, C.; BOERMA, J. T. How changes in coverage affect equity in maternal and child health interventions in 35 Countdown to 2015 countries: An analysis of national surveys. **The Lancet**, v. 380, n. 9848, p. 1149–1156, 2012.

WEISS, S.; ZHANG, T. Performance Analysis and Evaluation. *In: The Handbook of Data Mining*. 1. ed. [s.l.] LAWRENCE ERLBAUM ASSOCIATES, 2003. v. 1542p. 425–440.

WHO. **Making pregnancy safer: the critical role of the skilled attendant: joint statement**. [s.l: s.n.].

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP. WHO Child Growth Standards based on length / height , weight and age. **Acta Paediatrica**, v. Suppl 450, n. Suppl 450, p. 76–85, 2006.

WICKHAM, H.; FRANCOIS, R.; HENRY, L.; MÜLLER, K. **dplyr: A Grammar of Data Manipulation**, 2017. Disponível em: <<https://cran.r-project.org/package=dplyr>>

WICKHAM, H.; MILLER, E. **haven: Import and Export “SPSS”, “Stata” and “SAS” Files**, 2017. Disponível em: <<https://cran.r-project.org/package=haven>>

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. Second ed. [s.l.] Morgan Kaufmann Publishers, 2005.

WORLD HEALTH ORGANIZATION. **100 Core Health Indicators**. Geneva: [s.n.]. Disponível em: <<http://ci.nii.ac.jp/naid/40020358184/>>.

YOO, I.; ALAFAIREET, P.; MARINOV, M.; PENA-HERNANDEZ, K.; GOPIDI, R.; CHANG, J. F.; HUA, L. Data mining in healthcare and biomedicine: A survey of the literature. **Journal of Medical Systems**, 2012.

APÊNDICE A - DEFINIÇÃO DOS INDICADORES

De acordo com as orientações da OMS e UNICEF, os indicadores internacionais padronizados de monitoramento são construídos para garantir a comparabilidade ao longo do tempo e entre países. O Centro Internacional de Equidade em Saúde conduz análises periódicas nos inquéritos populacionais de saúde para extrair estes indicadores de forma padronizada, respeitando as definições existentes. As classificações de instituições e profissionais de saúde realizadas individualmente em cada inquérito oriundas de tais processos de análises foram disponibilizadas exclusivamente para este trabalho e não possuem acesso público. No âmbito deste trabalho foram utilizados quatro indicadores de monitoramento e suas definições operacionais são apresentadas na Tabela 22.

Tabela 22 – Definição dos indicadores

Indicador	Numerador	Denominador
Parto Institucional	Parto conduzido em uma instituição de saúde	Nascidos vivos nos últimos cinco anos
Atendimento Qualificado ao Parto	Parto conduzido por um profissional de saúde. A definição de profissional de saúde adotada é baseada nas categorizações de cada país	Nascidos vivos nos últimos cinco anos
Déficit de Altura por Idade	Altura por idade < 2 desvios padrão	Crianças vivas menores de 5 anos com altura medida e não marcadas como <i>outliers</i>
4 ou mais consultas pré-natal	Recebeu 4 ou mais visitas pré-natal	Último filho nascido vivo nos últimos cinco anos

APÊNDICE B - LISTA DE INQUÉRITOS DISPONÍVEIS

A construção das amostras utilizadas neste trabalho foi baseada em todos os inquéritos disponíveis entre o período de 5 a 15 de fevereiro de 2017. A Tabela abaixo lista todos os inquéritos. A nomenclatura dos países está de acordo com website da DHS (“The DHS Program”, 2017).

Tabela 23 – Lista de inquéritos utilizados na construção das amostras

Afghanistan 2015	Albania 2008	Armenia 2000	Armenia 2005	Armenia 2010
Azerbaijan 2006	Bangladesh 1993	Bangladesh 1996	Bangladesh 1999	Bangladesh 2004
Bangladesh 2007	Bangladesh 2011	Bangladesh 2014	Benin 1996	Benin 2001
Benin 2006	Benin 2011	Bolivia 1994	Bolivia 1998	Bolivia 2003
Bolivia 2008	Brazil 1996	Burkina Faso 1998	Burkina Faso 2003	Burkina Faso 2010
Burundi 2010	Central African Republic 1994	Cambodia 2000	Cambodia 2005	Cambodia 2010
Cambodia 2014	Cameroon 1998	Cameroon 2004	Cameroon 2011	Chad 1996
Chad 2004	Colombia 1995	Colombia 2000	Colombia 2005	Colombia 2010
Colombia 2015	Comoros 1996	Comoros 2012	Congo 2005	Congo 2011
Congo DR 2007	Congo DR 2013	Cote d'Ivoire 1994	Cote d'Ivoire 1998	Cote d'Ivoire 2011
Dominican Republic 1996	Dominican Republic 1999	Dominican Republic 2002	Dominican Republic 2007	Dominican Republic 2013
Egypt 1995	Egypt 2000	Egypt 2005	Egypt 2008	Egypt 2014
Ethiopia 2000	Ethiopia 2005	Ethiopia 2011	Gabon 2000	Gabon 2012
Gambia 2013	Ghana 1993	Ghana 1998	Ghana 2003	Ghana 2008
Ghana 2014	Guatemala 1995	Guatemala 1998	Guatemala 2014	Guinea 1999

Guinea 2005	Guinea 2012	Guyana 2009	Haiti 1994	Haiti 2000
Haiti 2005	Haiti 2012	Honduras 2005	Honduras 2011	India 1998
India 2005	Indonesia 1994	Indonesia 1997	Indonesia 2002	Indonesia 2007
Indonesia 2012	Jordan 1997	Jordan 2002	Jordan 2007	Jordan 2012
Kazakhstan 1995	Kazakhstan 1999	Kenya 1993	Kenya 1998	Kenya 2003
Kenya 2008	Kyrgyzstan 1997	Kyrgyzstan 2012	Lesotho 2004	Lesotho 2009
Liberia 2007	Liberia 2013	Madagascar 1997	Madagascar 2003	Madagascar 2008
Malawi 2000	Malawi 2004	Malawi 2010	Malawi 2015	Maldives 2009
Mali 1995	Mali 2001	Mali 2006	Mali 2012	Moldova 2005
Morocco 2003	Mozambique 1997	Mozambique 2003	Mozambique 2011	Myanmar 2015
Namibia 2000	Namibia 2006	Namibia 2013	Nepal 1996	Nepal 2001
Nepal 2006	Nepal 2011	Nicaragua 1998	Nicaragua 2001	Niger 1998
Niger 2006	Niger 2012	Nigeria 1999	Nigeria 2003	Nigeria 2008
Nigeria 2013	Pakistan 2006	Pakistan 2012	Peru 1996	Peru 2000
Peru 2004	Peru 2005	Peru 2006	Peru 2007	Peru 2008
Peru 2009	Peru 2010	Peru 2011	Peru 2012	Philippines 1993
Philippines 1998	Philippines 2003	Philippines 2008	Philippines 2013	Rwanda 2000
Rwanda 2005	Rwanda 2010	Sao Tome and Principe 2008	Senegal 1997	Senegal 2005
Senegal 2010	Senegal 2012	Senegal 2014	Sierra Leone 2008	Sierra Leone 2013
South Africa 1998	Swaziland 2006	Tajikistan 2012	Tanzania 1996	Tanzania 1999
Tanzania 2004	Tanzania 2010	Tanzania 2015	Timor Leste 2009	Togo 1998
Togo 2013	Turkey 1993	Turkey 2003	Uganda 1995	Uganda 2000
Uganda 2006	Uganda 2011	Ukraine 2007	Uzbekistan 1996	Vietnam 1997
Vietnam 2002	Yemen 2013	Zambia 1996	Zambia 2001	Zambia 2007
Zambia 2013	Zimbabwe 1994	Zimbabwe 1999	Zimbabwe 2005	Zimbabwe 2010
Zimbabwe 2015				

APÊNDICE C – TRABALHO SUBMETIDO AO CONGRESSO INTERNACIONAL DE EPIDEMIOLOGIA



Figura 7 – Pôster apresentado no Congresso Internacional de Epidemiologia