KopenótIA: Classificação de padrões de incidência de Diabetes Mellitus na Aldeia Limão Verde - Aquidauana/MS

Geovana Figueiredo Silva [1], Kaue Malacrida Dias[1], Marcia Ferreira Cristaldo[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Mato Grosso do Sul(IFMS) - Campus Aquidauana

Aquidauana – MS – Brasil

{geovanaf1912, kauemalacrida64}@gmail.com,
marcia.cristaldo@ifms.edu.br

Abstract. *Diabetes Mellitus affects at least 425 million people worldwide, including indigenous people living in protected areas. For more information on this disease, the aim of this study was to use machine learning, with samples composed of data from indigenous residents in the village of Limão Verde village in Aquidauana/MS. A standardized questionnaire was used to collect the information, then used this database for the Decision Tree algorithm, revealing 93% accuracy in the prediction of diabetes, also verified that the glycemic rate above 127mg/dl is one of the determining factors of diabetes, along with, the pre-existence of other diseases of the patient and sedentary lifestyle.*

Resumo. *A doença Diabetes Mellitus atinge pelo menos 425 milhões de pessoas em todo mundo, incluindo indígenas que residem em áreas protegidas. Para obter mais informações sobre esta doença, o objetivo deste estudo foi utilizar a aprendizagem de máquina, com amostras composta por dados de indígenas residentes na aldeia Aldeia Limão Verde em Aquidauana/MS. Foi utilizado um questionário para a coleta das informações, depois utilizou-se desta base de dados para o algoritmo de Árvore de Decisão, revelando 93% de precisão na previsão de diabetes, também verificou que a taxa glicêmica acima de 127 mg/dl é um dos fatores determinantes da diabetes, junto com, a pré-existência de outras doenças do paciente e o sedentarismo.*

1. Introdução

A Diabetes Mellitus, atinge pelo menos 425 milhões de pessoas em todo mundo, sendo considerada um dos grandes problemas da saúde pública do século XXI. Fatores como envelhecimento, sedentarismo e hábitos alimentares inadequados estão contribuindo para o aumento do número de casos (OLIVEIRA *et al*., 2011).

Segundo a Sociedade Brasileira de Diabetes (SBD, 2016), cerca de 6,9% da população brasileira é diagnosticada com esta doença crônica, deste modo, colocando o Brasil em 4º lugar na lista de países com maiores números de casos de incidência.

Segundo Tavares *et al*. (1999), os Karipúna e Palikur da região do Amapá foram as primeiras etnias indígenas a apresentarem registros oficiais de Diabetes Mellitus no Brasil. Após isso, outros casos foram ocorrendo e sendo registrados no país, nas mais diversas regiões, desde a Amazônia até o Centro-Oeste, incluindo a etnia Terena, que também se tomou alvo deste agravo.

Na população indígena, a Diabetes Mellitus é considerada uma doença emergente. Além dos fatores de risco tradicionais, problemas sociais relacionados às alterações da economia de subsistência, ao consumo de alimentos industrializados e ao contato cada vez mais frequente com a população urbana, contribuem para o aparecimento dessa e de outras doenças (COIMBRA, 2003).

O acesso à saúde pelos povos indígenas ainda é precária e, apesar de constante investimento, como o programa Políticas de Atenção à Saúde dos Povos Indígenas, diversos fatores tornam tardio o progresso de melhora no setor. Além de relatos da falta de formação de profissionais aptos a atuação em áreas interétnicas, grandes problemas relacionados a locomoção para acesso aos territórios, ainda são persistentes (MENDES *et al.,* 2018).

Visando atender a carência de profissionais da saúde em aldeias de difícil acesso, o projeto pretende reconhecer os padrões na incidência de Diabetes Mellitus na aldeia Limão Verde em Aquidauana/MS, ajudando a evitar complicações e facilitando o gerenciamento da doença na região. Para isso, foram utilizadas técnicas de Inteligência Artificial e conceitos de aprendizado de máquina.

2. Fundamentação teórica

2.1. Diabetes Mellitus

A Diabetes Mellitus(DM) é identificada por uma série de distúrbios metabólicos, caracterizada pelo excesso de glicemia (hiperglicemia) no organismo, resultante de defeitos na secreção de insulina, na ação da insulina ou em ambas (SBD, 2016).

Segundo SBD (2016) a doença pode ser dividida em 4 grupos distintos: diabetes tipo 1, presente em 5 a 10% dos casos, é predominante em crianças e jovens, onde o pâncreas não produz insulina (insulino dependentes); diabetes tipo 2, presente em 90 a 95% dos casos, se manifesta geralmente em adultos acima de 40 anos, é definida pela insuficiência do corpo em suprir com a demanda de glicose no sangue, baixa produção

de insulina associada a outras condições (falta de atividade física e má alimentação).

Normalmente pessoas com sobrepeso ou obesidade apresentam esse quadro clínico, compondo 60% a 90% dos portadores da doença; diabetes gestacional, similar à tipo 2, no entanto surge no período de gravidez, se encerrando ou não após a mesma; tipos específicos, compostas pela formas raras de diabetes mellitus, onde seus defeitos ou causas podem ser identificadas, inclui defeitos genéticos na função da célula beta, defeitos genéticos na ação da insulina, doenças do pâncreas exócrino e outras complicações (SBD, 2016).

Ainda existem estados clínicos como a pré-diabetes e a tolerância de glicose alterada, que podem ser fatores de risco para o desenvolvimento de diabetes e doenças cardiovasculares (SBD, 2016).

2.2. Descoberta de conhecimento

Com foco na especialização e no menor tempo de aprendizado, a análise de classificadores se torna uma tarefa importante para a descoberta de conhecimento. A quantidade de informações gerada a cada minuto pode ser imensa, e o ser humano torna-se incapaz de assimilar e administrar tais conhecimentos (REZENDE, 1994). Mas, apenas disponibilizar esses dados não é suficiente para um melhor aproveitamento das informações. É necessário ter ferramentas que facilitem a análise desses dados e auxiliem no desenvolvimento de estratégias de ação, isto é, a tomada de decisão (PRIETO *et al*., 2004).

A área médica tem sido uma das áreas mais beneficiadas pela tecnologia, por ser considerada detentora de problemas clássicos, possuidores de todas as peculiaridades necessárias para serem instrumentalizados por tais sistemas (NILSON, 1982). Neste contexto, a avaliação dos classificadores e sua implementação em um sistema que auxilie no diagnóstico de doenças é totalmente possível.

2.3. Mineração de dados

Mineração de dados consiste no processo de exploração e análise de dados com o objetivo de descobrir regras ou padrões previamente desconhecidos.

A KDD (*knowledge-discovery in databases*), ou extração de conhecimento no português, é um processo de extração de informações de base de dados. Estudos de HAND *et al.* (2001) dizem que a mineração de dados é a etapa principal do processo de KDD, sendo responsável pela busca no conjunto de dados, dos padrões que podem originar conhecimento útil .

A mineração pode ser realizada de 3 diferentes formas devido ao problema estudado. Se há pouco conhecimento, faz-se a descoberta não supervisionada, se há suspeita de alguma relação interessante, faz-se a testagem da relação (descoberta supervisionada) ou, se deseja ensinar o computador a realizar determinada ação durante uma situação específica, utiliza-se o aprendizado por reforço.

Na base de dados usada neste trabalho usou-se a descoberta supervisionada, ou seja, pacientes já diagnosticados. Na fase da mineração, dentro da KDD, necessita-se

definir a técnica e o algoritmo a ser utilizado para previsão. Uma vez escolhido o algoritmo a ser utilizado, deve-se implementá-la e adaptá-la ao problema proposto.

A escolha do algoritmo J48 foi devido a capacidade dele de trabalhar com variáveis qualitativas e quantitativas na base de dados. Além de possuir um dos melhores resultados para a confecção de árvores de decisão, ele é capaz de processar os ruídos com alto desempenho e baixo custo computacional, sem prejudicar o resultado final (LIBRELOTTO e MOZZAQUATRO, 2013).

O algoritmo J48 utiliza o método "dividir para conquistar" com a intenção de resolver problemas complexos encontrados durante o processamento da árvore. Essa técnica consiste em isolar o problema em subproblemas mais simples, aplicando o mesmo processo até que se consiga resolvê-los. Este método, utiliza a redução de entropia, que acaba por gerar um dado sobre o quão informativo um atributo é em relação ao outro. Dessa forma, organizando e distribuindo-os em cada nó da árvore.

Lidando com diferentes tipos de variáveis e os valores relacionados a elas, o J48 possui uma forma especial no tratamento das mesmas, arranjando-as de forma crescente. Para selecionar qual variável ficará no topo da raiz, o algoritmo utiliza equações de redução de entropia para verificar o tipo do conjunto de treinamento e o número de casos apresentados por ele. Depois é calculado o valor de informação esperado para cada variável do conjunto, a partir de um numero específicos de "nós-filhos" que a variável pode assumir.

A equação de ganho de informação é importante para definir qual atributo irá dividir melhor o conjunto de dados, favorecendo atributos com maiores variações em seus valores. Dessa forma, a razão de ganho de informação possui um denominador responsável por regularizar as amostras de dados que apresentarem grandes variações, podendo superar a limitação ao atenuar favorecimentos que possivelmente venham a ocorrer e comprovar que a melhor escolha foi feita (VIEIRA *et al.,* 2018).

Durante o processo de construção da árvore, vários ruídos, ou erros, são identificados, devido às muitas subárvores criadas, isso faz com que o algoritmo entre em um recurso de sobreajuste. Este procedimento traz consigo uma exclusão do modo generalizado de treinamento, provocando um aprendizado mais específico do conjunto.

Para tratar esses ruídos, o algoritmo utiliza o método de poda da árvore, *pruning*, no qual a taxa de acerto é melhorada e a árvore se torna mais simples, facilitando seu entendimento. O J48 apresenta o sistema pós-poda, na qual a árvore é analisada após a sua conclusão, de modo baixo-cima, transformando as informações sem ganhos significativos em nós-folhas.

3. Metodologia

Neste estudo, a técnica foi aplicada à população indígena de origem Terena, Limão Verde, localizada no Pantanal, distante 24 quilômetros da cidade de Aquidauana, no Mato Grosso do Sul, Brasil.

Com base nos objetivos a serem conquistados, essa pesquisa pode ser definida como exploratória. Segundo Quinlan (1993), a pesquisa exploratória visa possibilitar

maior proximidade com o problema proposto, tornando-o mais explícito de maneira que a construção de hipóteses seja possibilitada, tendo o aprimoramento de ideias ou a descoberta de intuições como objetivo principal. Para a execução do projeto foi definido um plano, Figura 1, no qual possui etapas a serem seguidas com o intuito de chegar ao objetivo da pesquisa.

Figura 1. Etapas do desenvolvimento do projeto.

Fonte. Própria (2020).

3.1. Planejamento do projeto

Elaboração de metas e objetivos a serem alcançados, bem como a determinação de como será feita a execução da pesquisa. As metas e objetivos foram listados a partir do auxílio dos orientadores e, seguindo o objetivo proposto no início do projeto.

3.2. Estudo bibliográfico

Foi realizado o estudo sobre a Diabetes Mellitus, conceitos de Inteligência Artificial e aplicações, após, foi separado estudos sobre atributos considerados gerais para a coleta de dados. Esses atributos se relacionam como sintomas e fatores de risco para o desenvolvimento do estado clínico e contribuíram para a formação de um questionário que seria aplicado na coleta de dados.

3.3. Coleta de Dados

Foram escolhidas 24 variáveis qualitativas e quantitativas, que permitam uma avaliação mais ampla do modo de vida do entrevistado. Assim, foram divididas as variáveis em 2 grupos, dados sociais e dados clínicos, para facilitar uma avaliação mais detalhada sobre como se relaciona cada atributo, de maneira geral, com o grupo de entrevistados.

Dentre as variáveis dos Dados Sociais, encontram-se em valores demográficos e de nível socioeconômico: idade(data de nascimento), sexo, escolaridade, estado civil; e hábitos de vida e doenças: praticante de atividade física, vícios, doenças e medicamentos.

Os dados clínicos são compostos por medidas antropométricas e clínicas: IMC(peso, altura), circunferência da cintura, circunferência abdominal, taxa glicêmica e pressão arterial;

Os dados foram colhidos com o auxílio de profissionais da saúde, que atuam na aldeia, e profissionais de educação física do IFMS campus Aquidauana. Os profissionais mediram em centímetros(cm) as circunferências do abdômen e da cintura, com o auxílio da fita métrica. Também foram medidas a pressão arterial em mmHg e a taxa glicêmica, que foi obtida por meio do medidor de glicose G-Tech Free da marca Accumed-Glicomed.

A coleta de dados permitiu gerar uma grande demanda de dados que necessitavam de ferramentas capazes de separá-los e fazer as análises necessárias. As informações foram dispostos em uma tabela, na qual foram organizados, conferidos e transferidos para linhas de código de banco de dados.

3.4. Mineração de Dados

Devido a grande demanda de dados, foi utilizado o processo de mineração de dados, que melhora a performance dos algoritmos de aprendizado de máquina. Simplifica os modelos de predição e reduz o custo computacional para "rodar" esses modelos. Ainda fornece um melhor entendimento sobre os resultados encontrados, uma vez que existe um estudo prévio sobre o relacionamento entre os atributos.

O software utilizado foi o WEKA 3.8, aplicando o algoritmo *Correlation-based Feature Selection* (CFS), baseado no método de relação de atributos (WAIKATO, 2019). Esse processo ajudou a eliminar atributos redundantes e irrelevantes, pois ele possui um método em que um conjunto de atributos é considerado bom se; contiver atributos altamente correlacionados com a classe e atributos não correlacionados entre si.

Após a aplicação do algoritmo CFS, em quais dos 24 atributos, foram selecionados apenas 3: glicose no sangue; outras doenças e atividades físicas, como atributos altamente relevantes para confirmação da diabetes para 124 pacientes analisados.

A Figura 2 mostra o resultado do CFS. A configuração do CFS foi: **weka.attributeSelection.CfsSubsetEval -P 1 -E 1**, sendo o método de busca: **weka.attributeSelection.BestFirst -D 1 -N 5**

Attribute Subset Evaluator (supervised, Class (nominal): 23 class):

CFS Subset Evaluator

Including locally predictive attributes

Selected attributes: 12,18,21 : 3

glicemia

atividades_fisicas

outras_doencas

Figura 2. Resultado do algoritmo CFS.

Fonte. Própria (2020).

CFS é um algoritmo que cria um ranking de subconjuntos de atributos, de acordo com uma função heurística de avaliação. A função prioriza subconjuntos que contém atributos altamente correlacionados aos atributos classe, e não-correlacionados entre si. Dessa forma atributos irrelevantes são eliminados pois possuem baixa correlação com a classe, e atributos redundantes são eliminados pois são altamente correlacionados com um ou mais dos demais atributos.

O valor de um atributo é medido a partir do ganho de informação que ele proporciona em relação ao atributo classe. É um método de seleção de atributos amplamente utilizado, porém possui a desvantagem de não levar em consideração a relação entre os atributos, apenas a relação entre cada atributo e a classe. O cálculo leva em consideração a razão de ganho de informação, que vai de 0 a 1, entre o ganho de informação e o valor intrínseco do atributo (entropia). Em geral, atributos com um maior número de valores distintos são selecionados por terem uma alta razão de ganho de informação. Esta propriedade explica a seleção dos atributos glicemia, outras atividades e outras doenças.

3.5. Aplicação da árvore de decisão

O algoritmo J48 é uma implementação da Árvore de Decisão C4.5 proposta por Ross Quinlan em 1993. Para aplicação do algoritmo foi utilizado o *software* WEKA, e para ter acesso ao banco de dados utilizados na etapa de mineração, deve-se criar arquivos do tipo ARFF, nessa criação o usuário precisa realizar a consulta das informações no banco de dados, extrair os resultados para um arquivo TXT ou CSV e utilizar um *software* para converter o arquivo para ARFF.

Para a proposta de criação de um modelo de árvore de decisão utilizou-se como estratégia de avaliação (teste) parte da base, ou seja, foi utilizado uma parte para a criação (indução) do modelo de árvore e a outra parte para avaliar o desempenho do modelo criado.

Para realizar esta configuração no experimento, foi utilizado a opção *Percentage split* com o percentual de 70%. Assim, tendo 70% (89 amostras) para induzir o modelo e 30% (35 amostras) para avaliar o modelo criado. Sendo a configuração do algoritmo: **weka.classifiers.trees.J48 -C 0.25 -M 2.**

3.6. Análise dos Resultados

Quando as RNAs são utilizadas para correlacionar diferentes grandezas físicas e estabelecer um modelo de medição, o valor gerado na saída também não está salvo de erros, para isso é feito a teoria de propagação da incerteza. Deste modo, o resultado dessas medições pode ser metrologicamente contestado. Isso porque, quando se relata o valor de uma medição é obrigatório que seja apresentado alguma indicação quantitativa da qualidade desse resultado.

Para avaliação do desempenho de cada modelo foram utilizados os seguintes critérios de avaliação: Curva de ROC (*Receiver Operating Characteristic*); TP - taxa de verdadeiros positivos; estatística de Kappa e MAE (Erro Absoluto Médio).

A curva ROC mostra a qualidade do modelo criado podendo diferenciar entre 0 e 1, ou positivo e negativo. Os melhores modelos conseguem distinguir com precisão o binômio. Um modelo cujas previsões estão 100% erradas tem uma ROC de 0, enquanto um modelo cujas previsões são 100% corretas tem uma ROC de 1.

A Estatística de Kappa é considerada como uma medida de concordância interobservador que permite avaliar tanto se a concordância está além do esperado tão somente pelo acaso, quanto o grau dessa concordância. Essa medida tem como valor máximo o valor unitário, que representa total concordância. Os valores próximos e até mesmo abaixo de zero indicam nenhuma concordância, ou a presença de uma eventual discordância entre os juízes.

A métrica MAE não leva em conta se o erro foi superestimado ou subestimado, caracterizando-se por ser a média dos erros cometidos pelo modelo de previsão durante uma série de execuções. Para calcular, subtraia o valor da previsão do valor verdadeiro em cada período de execução. O resultado deverá sempre ser positivo, sempre em módulo, soma-se e divide-se pelo número de valores que é usado para obter a soma, representado formalmente pela equação:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{O}_i - O_i \right| \qquad (1)$$

4. Resultados

Entre os 124 indígenas que realizaram o procedimento de coleta de dados, a média de idade foi de 35 anos. A amostra é constituída por 74 mulheres (59,6%) e 50 homens (40,4%).

Inicialmente foi relacionado a idade do entrevistado com o nível de glicose no sangue(mg/dl), vide tabela 1, que permitiu verificar que 12% dos adolescentes de até 15 anos, podem ser classificados como diabéticos. Outra observação importante é que entre os adultos acima de 40 anos, somam-se 48% deles classificados com valores para diabetes.

Tabela 1 - Relação Nível de Glicose no Sangue por faixa etária.

Idade	Normal (< 99mg/dl)	Pré-Diabetes (100 – 125mg/dl)	Diabetes (126mg/dl<)
<15	20,4%(10)	18,7%(9)	12%(3)
15 - 20	24,4%(12)	12,5%(6)	8%(2)
21 - 30	10,2%(5)	12,5%(6)	20%(5)
31 - 40	16,3%(8)	14,5%(7)	12%(3)
41 - 50	4%(2)	16,6%(8)	24%(6)
>51	24,4%(12)	25%(12)	24%(6)
Total	100%(49)	100%(48)	100%(25)

4.1. Desempenho dos algoritmos

Na Tabela 2 pode-se verificar o resultado do algoritmo J48 com a curva de ROC, um dos valores mais importantes produzidos pela WEKA, mostra uma ótima classificação com valores próximos de 1.

Outra análise é da Taxa de TP *Rate*: taxa de verdadeiros positivos (instâncias classificadas corretamente como uma determinada classe), na qual também obteve valores de desempenho classificados como desejáveis, próximo de 1.

Utilizando como métrica o Kappa obteve-se 74% de confiabilidade com o J48. A acurácia foi de 93%, com 116 instância corretamente classificadas. Na MAE que apresenta taxas de erro usada na previsão, o algoritmo apresentou um valor de 0,0752.

Tabela 2 - Desempenho dos algoritmos

Algoritmos	Acurácia	ROC Area negativo	TP Rate negativo	ROC Area positivo	TP Rate positivo	MAE	Kappa
J48	93%	0,940	0,975	0,940	0,900	0,0752	0,74
MLP	92%	0,935	0,975	0,600	0,935	0,1378	0,68
Naives Bayes	92%	0,905	0,952	0,904	0,789	0,0886	0,7262
KNN	79%	0,699	0,905	0,699	0,211	0,2026	0,1292

4.2. Matriz de confusão

Analisando o algoritmo J48 que apresentou melhores resultados de classificação, sua matriz de confusão está apresentada na Tabela 3.

A matriz de confusão retrata quais foram os casos em que o algoritmo errou a previsão. Ela apresenta as instâncias classificadas como corretas ou incorretas para cada classe, que é representada por **a** e **b**.

Tabela 3 - Matriz de confusão J48

a	b	Classificação
26	1	**a - negativo**
1	9	**b - positivo**

Aqui houve 37 instâncias teste, então as porcentagens e os números brutos se somam, aa + bb = 26+ 9 = 35, ab + ba = 1 +1 = 2. Na matriz verifica-se que dos 26 indígenas classificados como negativos para diabetes, apenas 1 foi classificado como positivo. Já dos 9 indígenas classificados como positivos para diabetes, somente 1 foi classificado como negativo.

A porcentagem de instâncias classificadas corretamente é frequentemente considerada com exatidão ou precisão na amostra. Ela tem algumas desvantagens, como a estimativa de desempenho, por isso deve-se avaliar a curva de ROC e a medida de Kappa mostrada acima, pois possuem valores mais estáveis e maior veracidade.

4.3 Árvore de decisão

Na árvore de decisão do algoritmo J48, Figura 3, mostra que os valores glicêmicos assumidos para não diabéticos apresentam como glicemia abaixo de 127 mg/dl, contrariando os valores de normalidade para população de 99 mg/dl.

Para valores acima de 127 mg/dl verifica-se a incidência de outras doenças, confirmando que a diabetes se associa a outras doenças. Caso não possua outras doenças, ainda é verificado se existe prática de atividades físicas, demonstrando a importância da manutenção diária da saúde, prevendo assim problemas futuros.

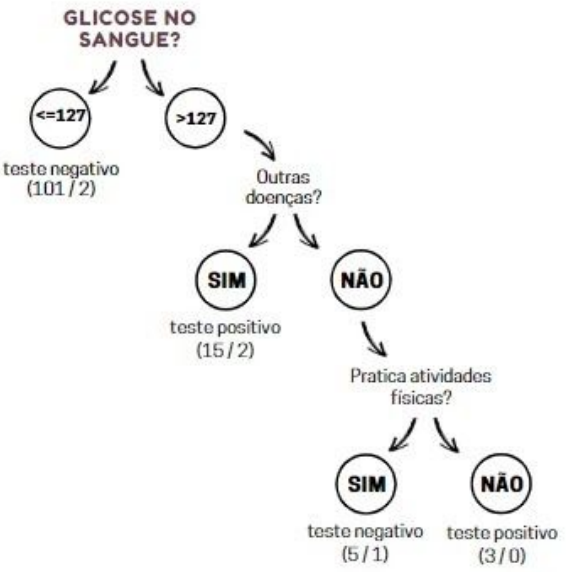

Figura 3. Árvore de decisão
Fonte: Própria (2020).

5. Conclusão

Após análise dos resultados, verificou-se que o melhor algoritmo foi o J48, que obteve 93% de acerto nas suas previsões. Os padrões da incidência da Diabetes Mellitus nos indígenas da aldeia Limão Verde foram: taxa glicêmica acima de 127 mg/dl, se possuía doenças pré-existentes e se praticava atividades físicas.

Os altos valores de diabetes encontrados para essa população são oriundos de uma alimentação pobre em fibras e rica em alimentos processados. Outro dado interessante apresentado pelos resultados é a associação de outras doenças com a diabetes, confirmando a diabetes como síndrome metabólica com associações a diversas enfermidades.

Neste trabalho inicial houve a coleta de dados para verificar quais padrões eram relevantes para a pesquisa, sendo a previsão do diagnóstico a ser realizada após aprovação pelo Comitê de Ética. Este artigo mostrou os padrões mais significativos para diagnosticar a diabetes na população estudada.

Portanto conclui-se que os resultados obtidos através do processo de descoberta de conhecimento atendem o objetivo proposto por este estudo, além de ser possível concluir a eficiência e precisão do método de descoberta supervisionada e a utilização da ferramenta WEKA, trazendo benefícios visíveis no processo de apuração da confiabilidade das técnicas de mineração de dados e da visualização dos principais atributos encontrados pela árvore de decisão J48 gerada.

Referências

Coimbra Jr., C. E. A. , Santos, R. V. e Escobar, A. L. (2003) “Epidemiologia e saúde dos povos indígenas no Brasil” em: Scielo Books , Editado por Fiocruz.

Hand, D., Mannila, H. e Smyth, P. (2001) “Principles of Data Mining”. Disponível em:< https://doc.lagout.org/Others/Data%20Mining/Principles%20of%20Data%20Mining%20%5BHand%2C%20Mannila%20%26%20Smyth%202001-08-01%5D.pdf>. Acessado em: 20 Nov. 2019.

IBGE (2010) “População residente, segundo a situação do domicílio e condição de indígena”. Disponível em:<https://indigenas.ibge.gov.br/graficos-e-tabelas-2.html>. Acessado em: 15 de Fev. de 2020.

Librelotto, S. R. e Mozzaquatro, P. M. (2013) “Análise dos algoritmos de mineração J48 e apriori aplicados na detecção de indicadores da qualidade de vida e saúde” em: Revint, página 5.

Mendes, A. M. , Leite, M. S. , Langdon, E. J. e Grisotti, M. (2018) “O desafio da atenção primária na saúde indígena no Brasil”.Disponível em:<https://iris.paho.org/bitstream/handle/10665.2/49563/v42e1842018.pdf?sequence=1&isAllowed=y>. Acessado em: 5 Mar. 2020.

Nilson, N.S. (1982) “Principles of Artificial Inteligence” editado por Springer Verlag, Berlin.

Oliveira, G. F. , Oliveira, T. R. R. , Rodrigues, F. F. , Corrêa, L. F. , Ikejiri, A. T. e Casulari, L. A. (2011) “Prevalência de diabetes melito e tolerância à glicose diminuída nos indígenas da Aldeia Jaguapiru”. Revista Panamericana de Salud Pública, Brasil.

Prieto, R. G. , Linhares, K. C. , Pinto, L. G. , Ortiz, J. R. (2004), “Programa de Mineração de Dados para Análise de Diabetes e Hipertensão”. Universidade do Vale do Itajaí (UNIVALI), Brasil.

Quinlan, J. R. (1993) “C4.5:Programing for machine learning” editado por Morgan Kaumann Publishers.

Rezende, S. O. (1994) “Sistema Inteligentes: fundamentos e aplicações”. Barueri, SP, Editora Manole Ltda.

SBD (2016) “Diretrizes da Sociedade Brasileira de Diabetes”. Disponível em: https://www.diabetes.org.br/profissionais/images/DIRETRIZES-COMPLETA-2019-2020.pdf>. Acessado em: 30 Maio 2020.

Tavares, E. F. , et al. (1999) “Anormalidades de tolerância à glicose e fatores de risco cardiovascular em uma tribo indígena aculturada da região amazônica brasileira”. Revista Brasileira de Endocrinol Metabol, v. 9, n. 43.

Vieira, E. , Neves, N. , Oliveira, A. , Moraes, R. e Nascimento, J. (2018) “Avaliação da performance do algoritmo J48 para construção de modelos baseados em árvores de decisão”. Revista Brasileira de Computação Aplicada, v. 10, n. 2.

Waikato, U. O. (2019) “Weka Data Mining Software in Java”, Disponível em:<http://www.cs.waikato.ac.nz/ml/weka/>. Acessado em: 15 de Outubro 2019.