TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREVISÕES DE DOENÇAS COM A ORANGE

DATA MINING TECHNIQUES FOR DISEASE PREDICTIONS WITH ORANGE

Erick Hidek Yamamoto Hoga [1]
Robson de Lacerda Zambroti [2]

RESUMO

Este trabalho tem como objetivo explorar a aplicação da mineração de dados na previsão de doenças, utilizando a ferramenta Orange com uma abordagem educativa e informativa. Abordando técnicas como Support Vector Machine (SVM), árvores de decisão, regressão logística e Naive Bayes, oferece informações relevantes para interessados, auxiliando-os a utilizar a ferramenta Orange de forma independente. Além disso, busca evidenciar a eficácia e relevância dessa abordagem, explicando conceitos e processos da mineração de dados e demonstrando como os padrões e informações mineradas podem contribuir para o diagnóstico médico.

Palavras-chave: mineração de dados; previsão de doenças; saúde; técnicas de análise; banco de dados.

ABSTRACT

This work aims to explore the application of data mining in disease prediction, using the Orange tool with an educational and informative approach. Techniques such as Support Vector Machine (SVM), decision trees, logistic regression, and Naive Bayes will be addressed, providing relevant information to assist those interested in the subject and helping them to use the Orange tool independently. Furthermore, the goal is to obtain results that highlight the effectiveness and significance of this approach by explaining data mining concepts and processes, as well as demonstrating how mined patterns and information can contribute to medical diagnosis.

Keywords: data mining; disease prediction; health; analysis techniques; database.

1 INTRODUÇÃO

A mineração de dados é uma subárea da Inteligência Artificial, que se concentra em descobrir padrões, informações e conhecimentos úteis em conjuntos

[1] Centro Universitário Filadélfia de Londrina - UniFil
[2] Centro Universitário Filadélfia de Londrina - UniFil

grandes de dados. De acordo com Noemi Galvão (2009), é um processo que envolve a aplicação de técnicas para extrair informações importantes e tomar decisões com base nos dados analisados. é uma das alternativas mais eficazes para extrair conhecimento a partir de grandes volumes de dados, que podem ajudar as instituições nas tomadas de decisões mais rápidas ou, até mesmo, a atingir um maior grau de confiança.

A pesquisa em saúde é fundamental para a tomada de decisões e para a melhoria da qualidade de vida das pessoas. A mineração de dados é uma técnica poderosa que pode ajudar a encontrar informações relevantes sobre questões de saúde, mas é importante saber onde encontrar as fontes de dados mais relevantes e como usá-las de forma eficiente.

Neste contexto, este estudo tem como objetivo explicar as técnicas de mineração de dados e demonstrar o uso da ferramenta Orange para a visualização e mineração desses dados. No decorrer deste trabalho, foram realizados testes utilizando a ferramenta Orange com um conjunto de dados contendo informações de indivíduos com ou sem diabetes, obtido a partir do site Kaggle, que é uma plataforma reconhecida por hospedar conjuntos de dados e competições relacionadas à ciência de dados e aprendizado de máquina. O objetivo é explorar o potencial da mineração de dados para obter *insights* a partir de dados de saúde, especialmente no contexto do diagnóstico e gerenciamento de diabetes.

A mineração de dados pode ser benéfica para o setor de saúde em muitas perspectivas, por exemplo, medicina preditiva, gerenciamento de relacionamento com o cliente, fraude, descoberta de abuso, monitoramento de saúde e também acompanhamento da eficiência de certas terapias (Richa Jain, 2021). Isso significa extrair informações úteis de um grande volume de dados. Nesta área, esses dados podem vir de prontuários eletrônicos, resultados de exames, pesquisas clínicas entre outras fontes.

Essas técnicas podem ajudar os profissionais de saúde a tomar decisões mais precisas e eficazes e com o uso de algoritmos de aprendizado de máquina, é possível analisar padrões e relacionamentos nos dados que podem ser utilizados para prever o desenvolvimento de doenças em pacientes.

2 REVISÃO DA LITERATURA

Nesta seção são citados dois trabalhos que foram relevantes para este artigo que aborda previsões utilizando técnicas de mineração de dados. O primeiro artigo, intitulado "Previsão de Diabetes usando Técnicas de Mineração de Dados", escrito por Rashi Rastogi e Mamta Bansal, explora a importância da previsão precoce da diabetes e destaca como a mineração de dados pode contribuir nesse contexto. Os autores propõem um modelo de previsão de diabetes que utiliza quatro técnicas: Random Forest, Support Vector Machine (SVM), Regressão Logística e Naive Bayes. Utilizando dados reais coletados do Kaggle e treinando o modelo em Python, eles analisam a performance do modelo através de métricas como matriz de confusão, sensibilidade e acurácia. A regressão logística obteve a maior acurácia, alcançando 82,46%, em comparação com as outras técnicas.

O segundo artigo destaca a relevância do estudo de dados, a autora Solange Oliveira Rezende, em seu artigo "Mineração de Dados. Sistemas inteligentes: fundamentos e aplicações", enfatiza a importância de não usar os dados da forma em que se encontram. O processo envolve etapas como a limpeza, seleção e padronização dos dados antes de aplicar as técnicas de mineração de dados. Este artigo salienta a necessidade de preparar adequadamente os dados antes de tirar o melhor proveito das técnicas de mineração de dados.

3 METODOLOGIA

Este artigo adotou uma abordagem metodológica que combina elementos explicativos e exploratórios. A explicativa devido a instruções como um guia detalhado que visa auxiliar o uso da ferramenta Orange na mineração de dados. Este guia explica o funcionamento básico de vários componentes e como aplicá-las na ferramenta.

Por outro lado, a abordagem exploratória foi fundamental na busca de compreender a operação da ferramenta Orange e o processo geral de mineração de dados. Neste contexto, foram realizados testes práticos e experimentos para ilustrar os conceitos e demonstrar a aplicação da ferramenta. Além disso, o estudo também

explorou a coleta e análise de dados, incluindo amostragem, análise quantitativa e análise qualitativa. Antes da escolha desta ferramenta, também foram analisadas outras opções. Contudo, devido ao custo associado às alternativas, foi optado por esta como a melhor para o trabalho.

4 PRÉ-PROCESSAMENTO DE DADOS

Antes de falarmos sobre as técnicas de mineração de dados na previsão de doenças, é necessário preparar e pré-processar os dados. Assim como diz a autora Solange Oliveira Rezende(2003), o pré-processamento de dados inclui três etapas principais:

- Seleção de dados relevantes.
- Limpeza de dados.
- Normalização de dados.

A seleção de dados relevantes é o primeiro passo na preparação dos dados. É importante escolher os atributos que mais influenciam a ocorrência da doença em questão. Por exemplo, no caso da previsão de diabetes, os atributos que foram levados em conta foram a idade, Gravidez, Glicose, Pressão Sanguínea, Espessura da pele, Insulina, IMC e o Histórico Familiar de diabetes que foram importantes para a análise, conseguindo filtrar os atributos mais importante para o seu objetivo e assim reduzindo o tamanho do conjunto de dados além de aumentar a eficiência da análise.

O segundo passo no pré-processamento de dados é a limpeza dos dados, esse processo envolve a detecção e correção de erros e valores ausentes nos dados. Erros e valores ausentes podem ocorrer devido a falhas na coleta, armazenamento e processamento dos dados. É importante detectar e corrigir esses erros, pois podem influenciar significativamente a análise e levar a resultados não muito precisos. A detecção e correção de erros podem ser feitas manualmente ou com o uso de ferramentas de software de limpeza de dados.

O terceiro passo no pré-processamento de dados é a normalização dos dados, esse processo envolve a transformação dos dados em um formato padronizado. Isso é importante porque diferentes atributos podem estar em escalas

diferentes, o que pode levar a resultados imprecisos na análise. Existem várias técnicas de normalização de dados, incluindo a normalização por escala mínima e máxima (MinMax), a normalização Z-score e a normalização decimal. A escolha da técnica de normalização depende dos dados específicos e do problema em questão.

5 TÉCNICAS DE MINERAÇÃO DE DADOS

Após o pré-processamento dos dados, é possível utilizar técnicas de mineração de dados para a previsão de doenças. Existem diversas técnicas disponíveis, neste artigo iremos falar de algumas delas, como árvores de decisão e Naive Bayes ,regressão logística e SVM.

Considerando que são técnicas de aprendizado de máquina, primeiro eles vão coletar um vasto volume de dados, criar seus padrões e realizar a mineração de dados, apenas de não ser necessário conhecer das técnicas para aplicação na ferramenta Orange, é importante saber como elas funcionam caso precise ajustar os parâmetros da técnica na ferramenta e não utilizar do seu padrão(*default*). A seguir, descreve-se um pouco de cada uma delas:

- Árvores de decisão é um modelo de previsão que representa uma série de decisões e seus possíveis resultados em uma árvore. Cada nó da árvore representa uma decisão ou uma folha que indica o resultado final da previsão.
- O Support Vector Machine (SVM) é um algoritmo de aprendizado de máquina que encontra uma linha ou superfície que separa pontos de dados em diferentes grupos. Ele é usado em tarefas como classificar e prever coisas, garantindo que a linha de separação esteja o mais distante possível dos pontos de dados, especialmente dos mais próximos, chamados de "vetores de suporte".
- Regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias.
- Naive Bayes é um método de aprendizado de máquina supervisionado que se baseia no teorema de Bayes. Ele é amplamente utilizado em tarefas de classificação, incluindo classificação de texto, detecção de spam e análise de

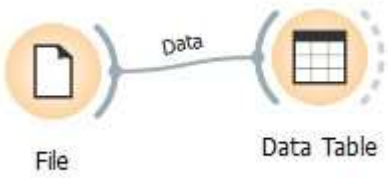
sentimentos. Naive Bayes é especialmente adequado para problemas nos quais a probabilidade de ocorrência de eventos é fundamental.

6 ORANGE

A construção deste trabalho teve como sua escolha predominante a ferramenta Orange, que foi explorada para as análises de mineração de dados. Este software de código aberto tem a reputação de oferecer uma experiência notavelmente eficaz pela sua simplicidade. Sua interface intuitiva, aliada à habilidade de apresentação visual de dados, facilita a extração de informações valiosas e *insights* a partir de conjuntos de dados complexos. Para começar, basta arrastar e soltar componentes no ambiente de trabalho. É como se a própria ferramenta estivesse guiando você ao longo do processo, nesta seção será apresentada algumas delas que ajudaram a realizar a mineração de dados:

- **File e Data Table:** O componente 'File' é onde você vai importar seus arquivos CSV, Excel, o seu arquivo que contém os dados (Dataset). Sem necessidade de mergulhar em configurações complexas, basta selecionar seu arquivo e pronto. Enquanto o componente 'Data Table' gera uma tabela que permite explorar e compreender a estrutura de seus dados, tudo isso com apenas uma conexão simples entre os 2 componentes.

Figura 1 – File e Data Table



Fonte: Captura de tela na ferramenta Orange

- **Select Columns e Select Rows:** O componente 'Select Columns' oferece uma maneira prática de escolher quais atributos ou colunas dos seus dados você deseja incluir em sua análise, enquanto o componente 'Select Rows' permite que

você refine seus dados selecionando as linhas ou instâncias que atendem a critérios específicos, adicionando filtros ao seu conjunto de dados para se concentrar nas observações relevantes.

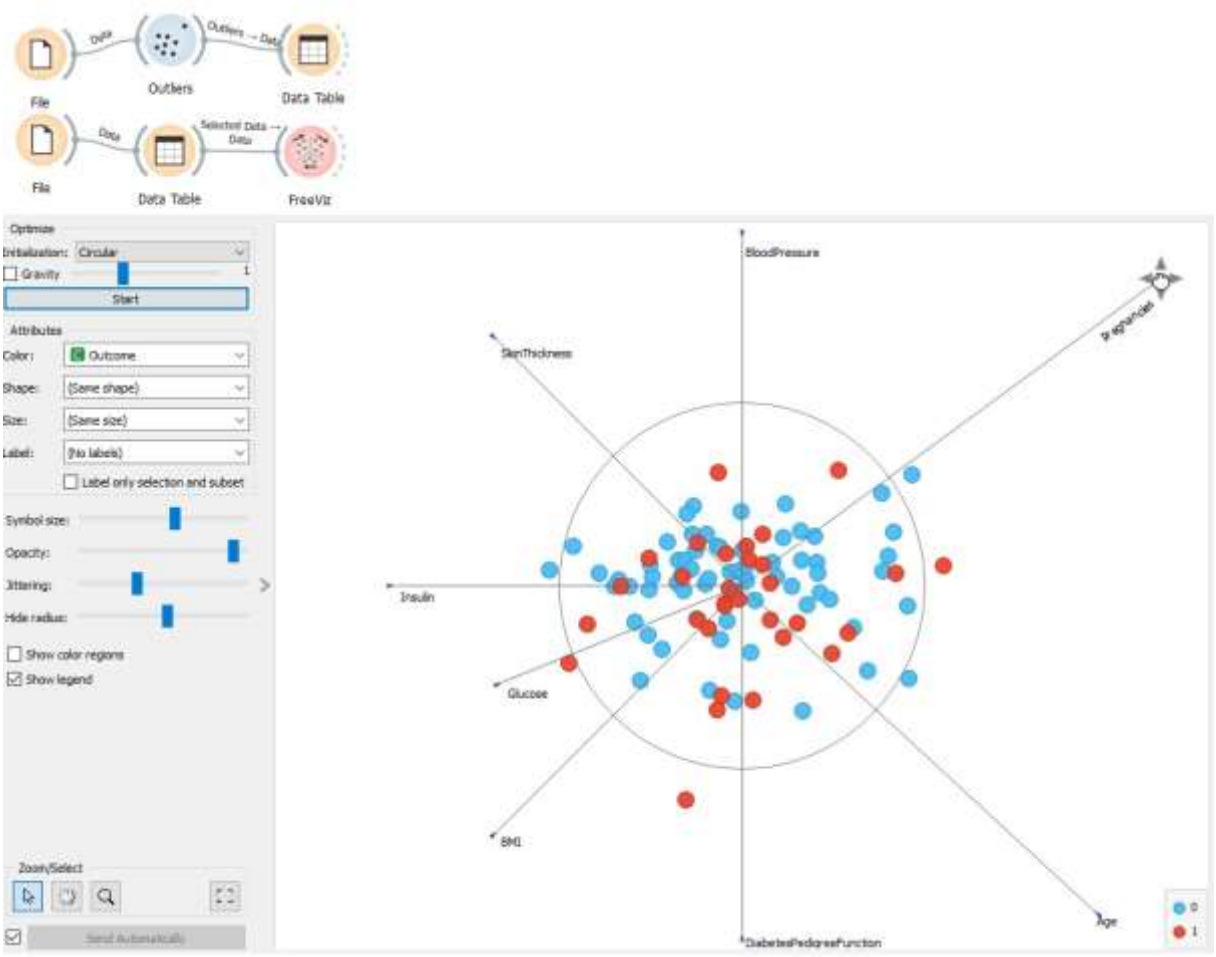
Esses dois componentes são cruciais para a fase de pré-processamento de dados em que são feitos a seleção de dados relevantes diminuindo assim o volume de dados.

Figura 2 – Select Columns e Rows

Fonte: Captura de tela na ferramenta Orange

- **Outliers e FreeViz:** Esses são outros componentes que ajudam na etapa de pré-processamento de dados, o 'Outliers' uma opção já programada pela Orange, ela automaticamente retira dados fora dos padrões, você apenas precisará definir parâmetros (porcentagem ou quantidade) para funcionar, já o 'FreeViz' é uma adição valiosa para visualização multidimensional, permitindo a identificação e remoção manual de dados 'fora da caixinha', ambos componentes podem ajudar nos resultados da mineração de dados.

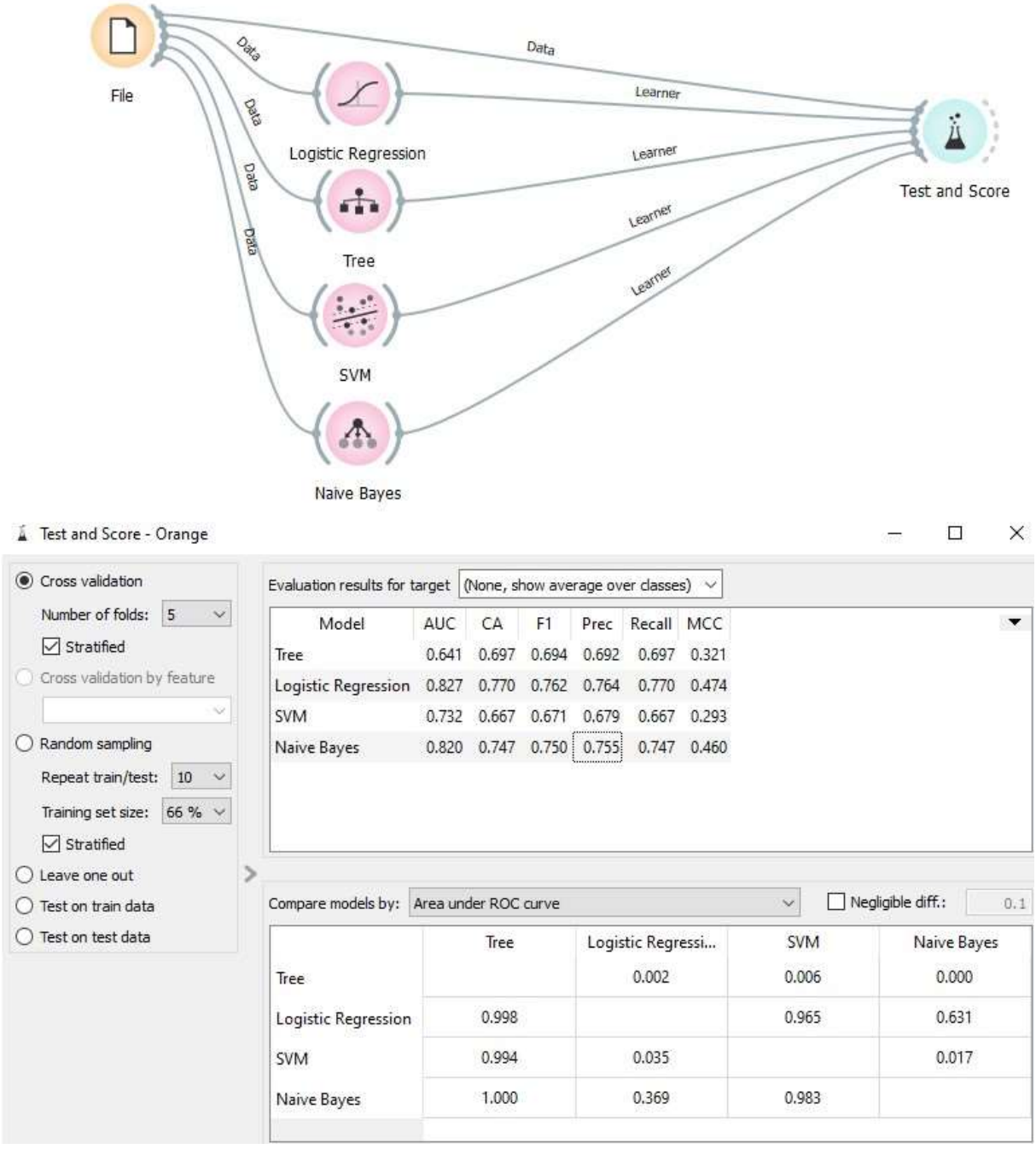
Figura 3 – Outliers e FreeViz



Fonte: Captura de tela na ferramenta Orange

- **Técnicas de mineração de dados na Orange:** Agora, com os dados prontos para uso, a ferramenta Orange torna a análise de mineração de dados uma tarefa acessível. Antes de fazer qualquer previsão, podemos avaliar a eficiência das técnicas usando o componente 'Test and Score' para obter dados valiosos da mineração como a taxa de acurácia e precisão de acertos dos resultados, lembrando que pode variar muito dependendo dos parâmetros que você utiliza/coloca para filtro. Essa facilidade de uso da Orange permite que você explore diferentes técnicas de mineração de dados de maneira intuitiva e eficiente sem ter tanto conhecimento técnico.

Figura 4 – Testes e resultados



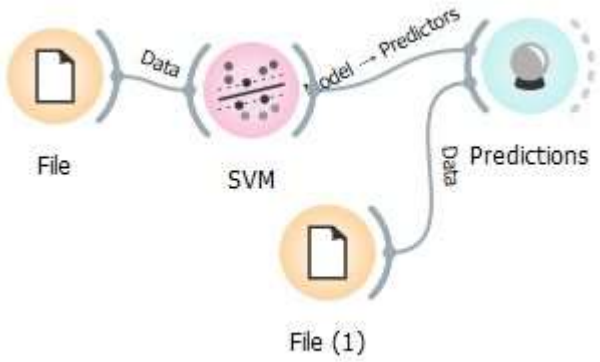
Fonte: Captura de tela na ferramenta Orange

- **Realizando previsões utilizando as técnicas:** Agora que compreendemos a eficácia de cada técnica de mineração de dados, vamos explorar um componente fundamental chamado 'Predictions'. Para utilizá-lo, conectamos um arquivo que contenha pelo menos uma coluna com dados categóricos a outro componente de

modelo de mineração, o qual pode ser qualquer uma das técnicas mencionadas anteriormente. E assim o 'Predictions' realiza um treinamento interno com base nos dados fornecidos.

Por fim, para obter resultados previstos, conectamos outra tabela que contém a coluna categórica, agora representada em formato de texto. Dessa forma, o 'Predictions' aplicará as técnicas de mineração previamente configuradas aos novos dados e gerará previsões com base nas informações fornecidas.

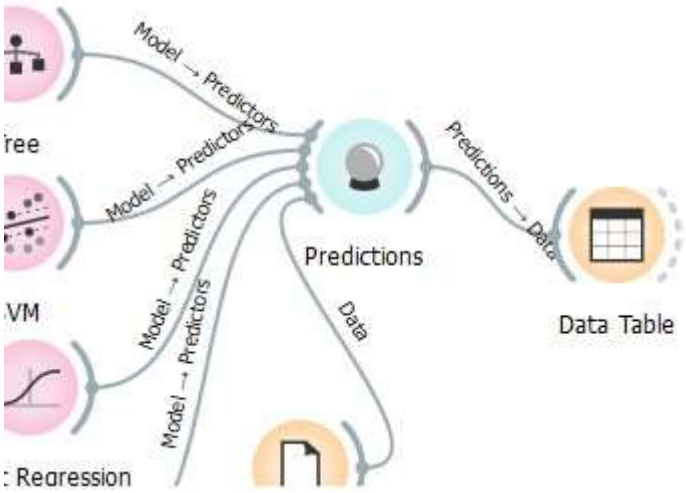
Figura 5 – Previsão da técnica



Fonte: Captura de tela na ferramenta Orange

- **Verificando as previsões das técnicas:** As previsões foram geradas com sucesso. No entanto, para uma análise mais detalhada, podemos conectá-las a uma tabela de dados (Data Table). Isso nos permitirá examinar as previsões em relação às porcentagens associadas a cada linha, indicando se a pessoa foi identificada como portadora ou não de diabetes (0 ou 1) por meio de cada técnica. Às vezes, a diferença entre as previsões e os resultados reais pode ser bastante sutil, e avaliar essas porcentagens de cada linha é essencial para compreender a precisão do modelo em contextos específicos, e na tabela podemos observar isso também.

Figura 6 – Previsão da técnica



Fonte: Captura de tela na ferramenta Orange

Figura 7 – Visualização da escolha da técnica

	Tree	SVM	Logistic Regression	Naive Bayes	Outcome	Pregnancies
1	0.04 : 0.96 → 1	0.18 : 0.82 → 1	0.29 : 0.71 → 1	0.04 : 0.96 → 1	1	6
2	0.97 : 0.03 → 0	0.44 : 0.56 → 0	0.94 : 0.06 → 0	0.98 : 0.02 → 0	0	1
3	0.00 : 1.00 → 1	0.34 : 0.66 → 1	0.21 : 0.79 → 1	0.20 : 0.80 → 1	1	8
4	0.99 : 0.01 → 0	0.49 : 0.51 → 0	0.95 : 0.05 → 0	0.99 : 0.01 → 0	0	1
5	0.00 : 1.00 → 1	0.51 : 0.49 → 0	0.12 : 0.88 → 1	0.17 : 0.83 → 1	1	0
6	0.95 : 0.05 → 0	0.39 : 0.61 → 1	0.84 : 0.16 → 0	0.90 : 0.10 → 0	0	5
7	0.00 : 1.00 → 1	0.52 : 0.48 → 0	0.92 : 0.08 → 0	0.99 : 0.01 → 0	1	3

	Outcome	Tree	SVM	Logistic Regression	Naive Bayes	Pregna
1	1	1	1	1	1	
2	0	0	0	0	0	
3	1	1	1	1	1	
4	0	0	0	0	0	
5	1	1	0	1	1	
6	0	0	1	0	0	
7	1	1	0	0	0	

Fonte: Captura de tela na ferramenta Orange

7 RESULTADOS

Após a conclusão dos testes utilizando a plataforma Orange, conseguimos observar alguns *insights* sobre a eficácia das técnicas de mineração de dados

aplicadas. Entre as quatro técnicas avaliadas, a regressão logística destacou-se como a mais precisa, como podemos visualizar na Figura 4, na coluna “CA”(acurácia) ou “Prec”(precisão) . É importante ressaltar que esses resultados foram obtidos considerando o uso das configurações e filtros padrões da ferramenta e com uma tabela retirada do Kaggle de 768 linhas.

Foi observado que, ao utilizar o componente 'Predictions', alguns valores apresentaram resultados muito próximos de 1 e 0 com os ajustes de parâmetros padrões da ferramenta. Assim, qualquer alteração nesses parâmetros poderia resultar em valores potencialmente mais discrepantes ou mais certeiros, evidenciando a sensibilidade do modelo às configurações utilizadas.

Então, é relevante mencionar que, para análises mais específicas ou em cenários complexos, é possível ajustar os parâmetros e filtros da ferramenta Orange de acordo com as necessidades do projeto. Essa flexibilidade permite a adaptação das técnicas de mineração de dados para atender aos objetivos e características dos conjuntos de dados específicos.

8 LIMITAÇÕES E DESAFIOS

Embora as técnicas de mineração de dados possam ser extremamente úteis na previsão de doenças, elas apresentam algumas limitações. Uma das principais limitações é que elas dependem da qualidade dos dados. Se os dados não forem precisos e completos, as previsões feitas com base neles podem ser imprecisas. Além disso, algumas técnicas de mineração de dados podem ser sensíveis ao tamanho do conjunto de dados, o que significa que elas podem não funcionar tão bem em conjuntos de dados menores.

A aplicação de técnicas de mineração de dados na área de saúde também apresenta alguns desafios específicos. Um desafio importante é a privacidade dos pacientes e a confidencialidade dos dados médicos. É importante garantir que os dados do paciente sejam protegidos e que a privacidade seja mantida, ao mesmo tempo em que se permita que os pesquisadores acessem os dados necessários para realizar suas análises.

Outro desafio é a falta de padronização nos dados de saúde, por isso temos

que seguir os passos do pré-processamento de dados citados no começo, pois os dados podem ser coletados de diferentes fontes e em diferentes formatos, o que torna difícil compará-los e combiná-los. É importante ter em mente essas diferenças ao selecionar e pré-processar os dados para a análise.

Além disso, as técnicas de mineração de dados podem produzir resultados complexos e difíceis de interpretar. É importante ter especialistas em saúde e estatísticos envolvidos no processo de análise para garantir que os resultados sejam interpretados corretamente e aplicados de forma adequada na prática médica.

9 CONCLUSÃO

Este artigo apresentou a mineração de dados de forma acessível e simplificada, fazendo uso da ferramenta Orange. Esta solução descomplicada tornou a exploração de conjuntos de dados complexos algo acessível mesmo para aqueles sem experiência prévia em mineração de dados ou conhecimentos técnicos aprofundados. Com o Orange, foi mostrado como importar, visualizar, aplicar técnicas de mineração e realizar previsões de maneira simples.

Espera-se que este artigo possa ter contribuído para pessoas que têm interesse neste assunto. Apesar da complexidade dessa área, a ferramenta demonstrada simplifica bastante o seu uso prático, eliminando a necessidade de profundos conhecimentos em programação ou a utilização de softwares complicados. Isso possibilita o aprofundamento no estudo dessa área, uma vez que a mineração de dados é um campo interdisciplinar fundamental, com relações em diversas áreas, como Inteligência Artificial e Aprendizado de Máquina, entre outras.

A resumo novamente volto a citar que o uso da ferramenta Orange abre portas para uma compreensão mais profunda dos seus próprios dados, permitindo-lhe tomar decisões informadas e fazer descobertas valiosas. Dessa forma, este artigo proporciona um ponto de partida para estudos mais aprofundados no campo da ciência de dados, utilizando uma ferramenta de fácil acesso e uso.

REFERÊNCIAS

A. Aldallal, A.A.A. Al-Moosa. Using data mining techniques to predict diabetes and heart diseases. In:2018 4th International Conference on Frontiers Of Signal Processing (ICFSP), IEEE, September 2018, p. 150-154.

Asmat Ara Shaikh, Amala Nirmal Doss, Muthukumar Subramanian, Vipin Jain, Mohd Naved, Md. Khaja Mohiddin. Major applications of data mining in medical. Materials Today: Proceedings, Volume 56, Part 4, 2022, Pages 2300-2304. ISSN 2214-7853. DOI: 10.1016/j.matpr.2021.11.642.

CARVALHO, D. R.; ESCOBAR, L. F. A.; TSUNODA, D. Pontos de Atenção para o Uso da Mineração de Dados da Saúde. Informação & Informação, [S. l.], v. 19, n. 1, p. 249–272, 2014. DOI: 10.5433/1981-8920.2014v19n1p249. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/16532.

Galvão ND, Marin H de F. Técnica de mineração de dados: uma revisão da literatura. Acta paul enferm [Internet]. 2009Sep;22(5):686–90. Available from: https://doi.org/10.1590/S0103-21002009000500014

GOMES, Christian; ROCHA, Leonardo. Combinando Técnicas de Mineração de Dados para Melhorar o Processo de Detecção Automática de Arritmia Cardíaca. In: Concurso de Trabalhos de Iniciação Científica da SBC (CTIC-SBC), 37., 2018, Natal.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Advanced Pattern Mining. In: HAN, Jiawei; KAMBER, Micheline; PEI, Jian (Eds.). Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2012. pp. 279-325. ISBN 9780123814791. DOI: 10.1016/B978-0-12-381479-1.00007-1.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data Preprocessing. In: HAN, Jiawei; KAMBER, Micheline; PEI, Jian (Eds.). Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2012. pp. 83-124. ISBN 9780123814791. DOI: 10.1016/B978-0-12-381479-1.00003-4.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Introduction. In: HAN, Jiawei; KAMBER, Micheline; PEI, Jian (Eds.). Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2012. pp. 1-38. ISBN 9780123814791. DOI: 10.1016/B978-0-12-381479-1.00001-0.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods. In: HAN, Jiawei; KAMBER, Micheline; PEI, Jian (Eds.). Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2012. pp. 243-278. ISBN 9780123814791. DOI: 10.1016/B978-0-12-381479-1.00006-X.

J. Santos-Pereira, L. Gruenwald, J. Bernardino. Top data mining tools for the healthcare industry. J. King Saud Univ. Comput. Inf. Sci.}, 2021.

JERANT, A. F. et al. Early detection and treatment of skin cancer. American Family Physician, v. 62, n. 2, 2000.

N.P. Jayasri, R. Aruna. Big data analytics in health care by data mining and classification techniques. ICT Express, 8 (2), 2022, pp. 250-257.

NATIONAL CANCER INSTITUTE. Cancer Statistics. 2020. Disponível em: https://www.cancer.gov/about-cancer/understanding/statistics.

R. Jain e D. V, "Data Mining Algorithms in Healthcare: An Extensive Review,"2021 Quinta Conferência Internacional sobre I-SMAC (IoT em redes sociais, dispositivos móveis, análises e nuvem) (I-SMAC), Palladam, Índia, 2021 , pp. 728-733, doi: 10.1109/I-SMAC52330.2021.9640747.

RASTOGI, Rashi; BANSAL, Mamta. Diabetes prediction model using data mining techniques. Diabetes prediction model using data mining techniques2, [S. l.], v. 25, 1 fev. 2023. DOI: 10.1016/j.diabres.2022.108333.

REZENDE, Solange Oliveira et al. Mineração de dados. Sistemas inteligentes: fundamentos e aplicações, v. 1, p. 307-335, 2003.

S.I. Ayon, M.M. Islam. Diabetes prediction: a deep learning approach. Int. J. Inf. Eng. Electron. Bus., 12 (2), 2019, p. 21.

SIEGEL, Rebecca L. et al. Cancer Statistics. CA: A Cancer Journal for Clinicians, v. 71, n. 1, p. 7-33, jan. 2021. DOI: 10.3322/caac.21654.

WENG, Stephen F. et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PloS One, v. 12, n. 4, p. e0174944, Apr. 2017. DOI: 10.1371/journal.pone.0174944.

YALA, Adam et al. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology, v. 292, n. 2, p. 331-338, May 2019. DOI: 10.1148/radiol.2019182716.