

## AVALIAÇÃO DO DESEMPENHO DO ALGORITMO *RANDOM FOREST* NA PREDIÇÃO DE DIABETES MELLITUS

Hean Talles Souza Ferreira<sup>1</sup>  
Adriano Carlos Soares<sup>2</sup>

professoradrianosoares@gmail.com

**ÁREA DO CONHECIMENTO:** Ciências da Saúde

### RESUMO

A detecção precoce da Diabetes Mellitus é crucial para possibilitar a adoção de medidas preventivas e evitar prejuízos à saúde. Essa morbidade geralmente decorre da produção insuficiente de insulina pelo organismo ou da sua má absorção. No entanto, identificar a doença em estágios iniciais pode ser desafiador para os especialistas. Com esse objetivo em mente, este estudo buscou avaliar a eficácia de um modelo de classificação *Random Forest* no diagnóstico da Diabetes Mellitus. Para isso, uma pesquisa experimental foi conduzida utilizando uma base de dados pública disponível na plataforma de ciência de dados Kaggle. Após o tratamento e balanceamento dos dados, foi aplicado o algoritmo *Random Forest* e analisado por meio de métricas como acurácia, precisão, sensibilidade e *F1-score*. Notavelmente, o modelo preditivo obteve um desempenho superior ao utilizar a técnica de *oversampling* para balancear os dados, alcançando uma acurácia média de 85,9%. Esse resultado representou um aumento significativo de 9,38% em comparação com o conjunto de dados não balanceado. Os achados desta pesquisa foram devidamente comparados com informações presentes na literatura acadêmica, corroborando o potencial do algoritmo como uma valiosa ferramenta de suporte nas decisões médicas de pré-diagnóstico.

**PALAVRAS-CHAVE:** diabetes mellitus, *random forest*, aprendizado de máquina, pré-diagnóstico.

### INTRODUÇÃO

A diabetes mellitus (DM) é uma doença metabólica crônica que afeta milhões de pessoas em todo o mundo (Muzy *et al.*, 2021). Caracterizada pelo aumento significativo dos níveis glicêmicos, a diabetes exerce um impacto significativo na

---

<sup>1</sup> Acadêmico do curso de Farmácia – Centro Universitário Vértice – UNIVÉRTIX.

<sup>2</sup> Farmacêutico Bioquímico (UFOP); Cirurgião Dentista (UNIVÉRTIX); Doutor em Bioquímica Aplicada (Biotecnologia) (UFV); Mestre em Ciências Naturais e da Saúde (UNEC); Especialista em Docência do Ensino Superior (UCAM, RJ). Professor dos cursos de Farmácia, Psicologia, Enfermagem, Biomedicina e Odontologia do Centro Universitário Vértice – UNIVÉRTIX.

saúde e no bem-estar das pessoas afetadas. Nos países em desenvolvimento, como o Brasil, a diabetes tem emergido como um desafio significativo para a saúde pública, contribuindo com cerca de 30% a 40% dos casos de morbidade. (Castro *et al.*, 2021).

Segundo a Sociedade Brasileira de Diabetes (2019), aproximadamente 50% dos pacientes diagnosticados com DM, desconheciam sua condição. Dado que esse conjunto de indivíduos não possui conhecimento sobre sua situação clínica, eles se encontram impossibilitados de adotar os cuidados essenciais. Conseqüentemente, o diagnóstico para esse grupo ocorre de forma tardia ou não ocorre, acarretando situações de risco para a saúde do paciente. Nesse contexto, os algoritmos de aprendizado de máquina, podem desempenhar um papel fundamental na resolução dessa problemática. Ao analisar grandes volumes de dados, esses algoritmos têm a capacidade de identificar padrões e indicadores precoces da diabetes, permitindo a predição da condição clínica do paciente. Além disso, eles podem ser integrados em sistemas de triagem e monitoramento de saúde, auxiliando os profissionais na identificação de indivíduos em risco e na implementação de medidas preventivas (Moreira *et al.*, 2021).

O *Random Forest* (Floresta aleatória) é um algoritmo de aprendizado supervisionado que cria múltiplas árvores de decisão aleatórias e realiza a combinação destas para obter previsões estáveis e precisas. Graças à sua capacidade de classificar e reconhecer padrões com precisão, pesquisadores passaram a empregar esse algoritmo no campo da bioinformática visando solucionar desafios clínicos, como a realização de diagnósticos e previsões prognósticas para diferentes quadros clínicos (Pinheiro, 2022).

Em vista dos fatos mencionados, o presente estudo tem como objetivo avaliar o desempenho de um modelo de classificação *Random Forest* no diagnóstico de Diabetes Mellitus.

## FUNDAMENTAÇÃO TEÓRICA

A DM é uma síndrome que compromete o metabolismo dos carboidratos, gorduras e proteínas. Ela ocorre devido à ausência de secreção de insulina ou à diminuição da sensibilidade dos tecidos a esse hormônio. Como resultado, há um

aumento nos níveis de glicose no sangue, uma redução na utilização da glicose pelas células e um aumento na utilização de gorduras e proteínas como fonte de energia (LADEIA *et al.*, 2020).

Em pesquisas recentes, o Brasil foi classificado como o sexto país com maior número de adultos diabéticos do mundo, como pode-se visualizar na Tabela 1, contando com 15,7 milhões de registros (IDF, 2021).

Tabela 1: Os 10 países com maior número de adultos diabéticos no mundo em 2021.

Posição	País	Nº de pessoas com diabetes (em milhões)
1º	China	140,9
2º	Índia	74,2
3º	Paquistão	33,0
4º	Estados Unidos da América	32,2
5º	Indonésia	19,5
6º	Brasil	15,7
7º	México	14,1
8º	Bangladesh	13,1
9º	Japão	11,0
10º	Egito	10,9

Fonte: IDF (2021).

A classificação recomendada pela Sociedade Brasileira de Diabetes (SBD) e proposta pela Organização Mundial da Saúde (OMS), inclui na síndrome da DM quatro classes clínicas: a DM tipo 1 (DM1), a DM tipo 2 (DM2), DM gestacional (DMG) e outros tipos específicos. Além disso, existe o estado pré-diabético, que pode ser caracterizado quando os níveis de glicemia em jejum são superiores a 100 mg/dL e menor que 126 mg/dL e por uma diminuição na tolerância à glicose (SBD, 2019).

Os dois principais tipos de diabetes, DM1 e DM2, apresentam diferenças significativas. A DM1, ocasionada por uma reação em que o próprio sistema imunológico do indivíduo ataca as células  $\beta$ -pancreáticas responsáveis pela produção de insulina, afeta principalmente crianças e adolescentes. Por outro lado, a DM2, é mais comum em adultos, estando normalmente associada a obesidade e sedentarismo, que provocam uma gradual redução da produção de insulina pelas células  $\beta$ -pancreáticas e resistência a ação deste hormônio em diversos tecidos do corpo (GENUTH, PALMER, NATHAN, 2018).

Na criação de inteligências artificiais (IA), são empregados dois métodos que emulam o raciocínio humano, o indutivo, que consistem em tirar conclusões ou estabelecer regras gerais a partir de premissas, e o dedutivo, que opera em contraposto, permitindo a obtenção de diversas premissas a partir de regras já consolidadas. No aprendizado de máquina (AM), apenas o método indutivo é aplicado. Dessa maneira, as máquinas adquirem a capacidade de aprender e se aprimorar por meio da experiência adquiridas ao longo de um período de treinamento e validação. Utilizando um conjunto de dados como ponto de partida, os algoritmos são capazes de extrair informações cruciais e estabelecer relações entre elas, permitindo que alcancem um determinado objetivo de forma eficiente (LUDERMIR, 2021).

A área de Aprendizado de Máquina conta com quatro métodos principais de treinamento: supervisionado, semissupervisionado, não supervisionado e por reforço. Cada um desses métodos exerce uma eficácia distinta, influenciada pelos dados fornecidos e pela do problema a ser solucionado. Na aprendizagem supervisionada, é fornecido para o algoritmo um conjunto de variáveis, e por meio da relação existente entre elas, é realizado a predição de uma variável alvo. Durante o processo de aprendizagem, os dados a serem fornecidos são separados em conjuntos de treino e teste. Após realizar o treinamento utilizando o conjunto de treino o desempenho do modelo será medido a partir da sua capacidade de identificar e relacionar os dados presentes no conjunto de teste, que até então nunca foram vistos pelo algoritmo (SILVA *et al.*, 2019).

O *Random Forest* é caracterizado como um modelo de aprendizagem supervisionada, sendo também definido como um *Ensemble* da Árvore de decisão. Os algoritmos do tipo *Ensemble* combinam diversos resultados de modelos preditivos isolados de forma a aprimorar a capacidade preditiva. Em outras palavras, o *Random Forest* é composto por um conjunto de Árvores de decisão, e o valor de saída é o resultado escolhido por esse conjunto de árvores. Nesse modelo, a quantidade de árvores a serem utilizadas é configurável, e quanto maior o número de árvores, mais estável e confiável se torna o algoritmo. No entanto, o tempo de processamento e o poder computacional necessário aumenta proporcionalmente (SILVA, 2023).

## METODOLOGIA

Trata-se de um estudo experimental, com abordagem quantitativa. Segundo Yoshida, Grotto e Gonçalves (2019), o método experimental visa investigar relações de causa e efeito entre variáveis, manipulando as condições para identificar os impactos produzidos. Ademais, em pesquisas quantitativas, são utilizadas técnicas estatísticas para quantificar numericamente opiniões e informações para classificá-las e analisá-las.

Os dados utilizados para a realização deste estudo foram obtidos a partir de um banco de dados de domínio público disponível no Kaggle (<https://www.kaggle.com/datasets/mathchi/diabetes-data-set>). Este conjunto de dados é originalmente do Instituto Nacional de Diabetes Doenças Digestivas e Renais (NIDDK), e todos os pacientes incluídos são do sexo feminino, com idade mínima de 21 anos e de origem indígena Pima, apresentando um total de 768 pacientes. O banco de dados possui 9 tipos de atributos por indivíduos, os quais estão descritos na Tabela 2.

Tabela 2: Atributos presentes no banco de dados.

Atributos	Definição
Gestações	Número de gestações.
Glicose	Concentração plasmática de glicose em 2 horas em um teste oral de tolerância à glicose.
Pressão Sanguínea	Pressão arterial diastólica (mmHg).
Espessura da pele	Dobra cutânea tricipital (mm).
Insulina	Nível sérico de insulina.
IMC	Índice de massa corporal.
Pedigree da diabetes	Probabilidade de diabetes com base no histórico familiar.
Idade	Idade do paciente em anos.
Resultado	Valor binário indicando se o paciente é diabético (1 = sim, 0 = não).

Fonte: Chang *et al.* (2022).

Inicialmente foi realizado o tratamento dos dados, verificando e descartando informações duplicadas, assim como, os valores faltantes em uma determinada variável foram substituídos pela média dos valores não-nulos dessa variável. Ademais, foi utilizada uma matriz de correlação para identificar as variáveis mais

relevantes na predição da DM. (ARAUJO, SANTOS, GOMES, 2019; PINHEIRO, 2022).

Para fins comparativos, o banco de dados utilizado foi balanceado utilizando duas técnicas de balanceamento aleatório, conhecidas como *oversampling* e *undersampling* (ALMEIDA, 2021). A técnica de *oversampling* envolve a escolha aleatória de registros pertencentes à classe minoritária e a replicação desses dados para expandir a amostra, visando atingir o equilíbrio numérico entre ambas as classes. Essa abordagem é recomendada para amostras de dados reduzidas, pois resulta em um aumento significativo no número de registros (MAIONE, 2020). Já a técnica de *undersampling* envolve a redução aleatória do número de registros pertencentes à classe majoritária de forma que o conjunto de dados seja equilibrado com a classe minoritária (PINHEIRO, 2022).

Após o cuidadoso tratamento e balanceamento dos dados, as amostras foram divididas aleatoriamente em conjuntos de treino e teste, com 88% dos dados destinados ao treinamento do algoritmo e 22% reservados para a realização do teste. Após esse processo, utilizando a variável binária “Resultado” como alvo, foi possível criar um modelo preditivo com a entrada dos 8 atributos remanescentes.

Todos os testes foram conduzidos utilizando a técnica de Validação Cruzada K-Fold com um total de 5 *folds*, buscando evitar avaliações tendenciosas e fornecer uma estimativa mais realista do desempenho do modelo. Nessa abordagem, o conjunto de dados é dividido em K subconjuntos (chamados de "*folds*") de tamanhos aproximadamente iguais. O modelo é treinado K vezes, usando K-1 *folds* como conjunto de treinamento e o *fold* restante é utilizado como conjunto de teste em cada iteração. O desempenho é avaliado pela média dos resultados obtidos em cada uma das iterações (CUNHA, 2019).

A fim de avaliar adequadamente o desempenho de um algoritmo que gera saída para duas classes ou mais, é comum empregar a matriz de confusão, que explicita as taxas de verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN). A partir dessa matriz, podemos calcular diversas métricas de avaliação para o modelo ou algoritmo em questão (SILVA, 2023).

As métricas utilizadas para a avaliação do algoritmo *Random Forest* estão descritas na Tabela 3. Ao empregar essas métricas de avaliação de desempenho, é possível obter uma visão abrangente sobre o desempenho do algoritmo na tarefa em questão (PINHEIRO, 2022; SILVA, 2023).

Tabela 3: Métricas utilizadas para avaliação do algoritmo preditivo.

<b>Métrica</b>	<b>Definição</b>
Acurácia	Proporção de previsões corretas em relação ao total de previsões realizadas.
Precisão	Proporção de verdadeiros positivos em relação ao total de previsões positivas.
Sensibilidade	Proporção de verdadeiros positivos em relação ao total real de positivos.
F1-Score	Média harmônica entre precisão e sensibilidade.

Fonte: Pinheiro (2022).

## RESULTADOS E DISCUSSÕES

Em uma análise exploratória dos dados, foi gerada uma matriz de correlação capaz de interrelacionar as variáveis presentes no conjunto de dados, a qual está representada na Figura 1.

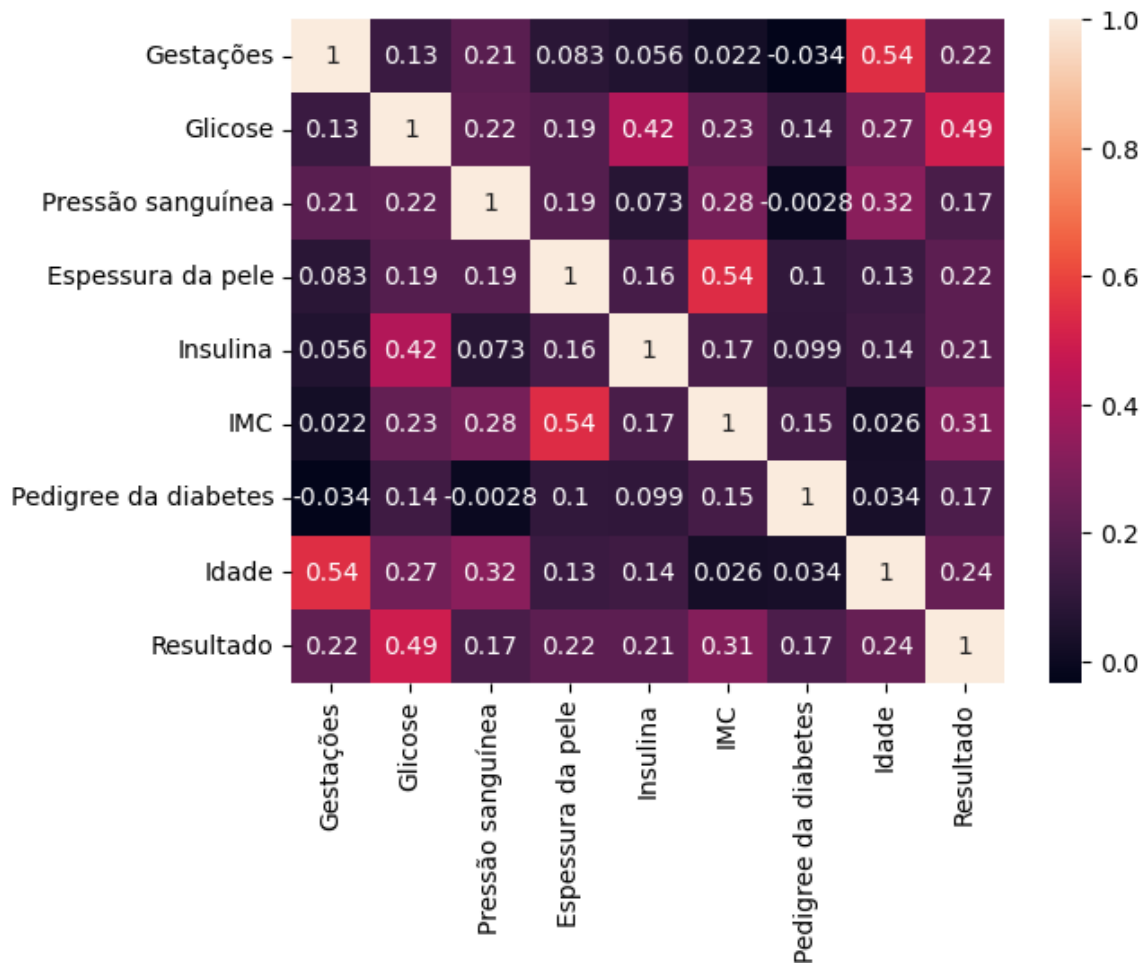


Figura 1: Matriz de correlação

Em relação a figura 1, vale ressaltar que, quanto mais próximo de 1, maior a correlação positiva existente entre as variáveis. No entanto, quando mais próximo de -1, maior a correlação negativa entre essas variáveis. Sendo assim, é possível observar que a concentração plasmática de glicose apresenta a maior relação com o resultado dos pacientes, seguido pelo Índice de Massa Corporal (IMC), demonstrando uma grande relevância dessas variáveis para a predição de diabetes (ARAÚJO, SANTOS, GOMES, 2019).

Ao consultar os registros no banco de dados, observou-se um desbalanceamento quanto ao quadro clínico dos pacientes. Enquanto haviam 500 registros de pacientes sem diabetes, o número de registro de paciente com a doença era igual a 268. Um conjunto de dados é considerado balanceado quando todas as

classes têm uma quantidade semelhante de dados, evitando que o modelo se torne tendencioso para classe majoritária ou apresente um desempenho inferior na classificação da classe minoritária. Nesse sentido, o balanceamento do conjunto de dados é uma consideração essencial ao utilizar algoritmos de aprendizado de máquina. (ALMEIDA, 2021).

Para a realização da avaliação do algoritmo, foram utilizadas três abordagens investigativas, cada uma delas envolvendo diferentes conjuntos de dados: um sem balanceamento, outro com aplicação de *oversampling* e um terceiro com utilização de *undersampling*. Em cada um desses cenários, o modelo preditivo foi submetido a um total de 30 execuções, com a aplicação da técnica de validação cruzada em cada uma delas. Após realizar as execuções, procedeu-se ao cálculo da média dos resultados, que foram compilados na Tabela 4.

Tabela 4: Resultados obtidos em diferentes métodos de balanceamento de dados.

<b>Método de Balanceamento</b>	<b>F1-Score</b>	<b>Acurácia</b>	<b>Sensibilidade</b>	<b>Precisão</b>
Sem Balanceamento	0,6179	0,7652	0,5446	0,7017
<i>Oversampling</i>	0,8623	0,8590	0,9360	0,8174
<i>Undersampling</i>	0,7952	0,7879	0,8194	0,7647

Ao analisarmos os resultados, percebe-se que o conjunto sem balanceamento obteve um baixo desempenho, com apenas 54,46% de sensibilidade média, assim como 61,79% de *F1-score*. Além disso, apresentou os piores resultados em todas as métricas, demonstrando a importância da utilização de métodos de balanceamento de dados ao realizar a utilização de um modelo preditivo (ALMEIDA, 2021).

Em termos gerais, observou-se que os conjuntos sem balanceamento, com *oversampling* e com *undersampling* apresentaram, respectivamente, uma acurácia média de 76,52%, 85,9% e 78,79%. Essa constatação encontra respaldo em um estudo conduzido por Chang *et al.* (2022), o qual avaliou três algoritmos de aprendizado de máquina para a classificação do diabetes mellitus em índios Pima. Nessa pesquisa, o algoritmo de Random Forest obteve os melhores resultados em

termos de acurácia, alcançando uma taxa de 79,57%. Essa evidência reforça a eficácia do algoritmo ao lidar com a problemática proposta.

Assim como demonstrado na Figura 2, o conjunto de dados balanceado utilizando *oversampling* obteve o melhor desempenho em todas as métricas, demonstrando um aumento na acurácia média de 9,38% quando comparado com o conjunto sem balanceamento, e 7,11% em relação ao conjunto com aplicação de *undersampling*. Contudo, é importante destacar que o *oversampling* consiste em replicar instâncias da classe minoritária, podendo ocasionar distorções nos resultados, uma vez que a duplicação desses registros pode enviesar o processo de treinamento ao introduzir repetições de registros no conjunto de treino e teste, o que pode limitar a representatividade dos dados em relação à realidade (MAIONE, 2020).

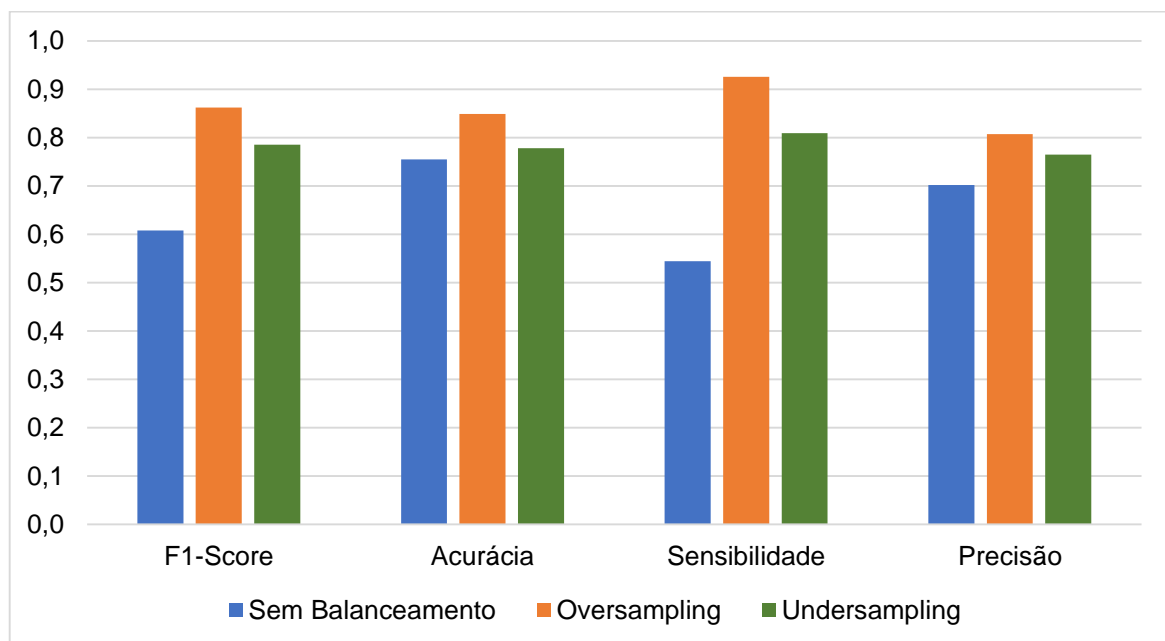


Figura 2: Comparação das métricas de avaliação em relação ao método de balanceamento de dados.

O conjunto de dados balanceado com *undersampling* apresentou um ligeiro aumento de 2,27% na acurácia média em relação ao conjunto sem balanceamento. Entretanto, o banco de dados utilizado apresenta um baixo número de registros, não sendo ideal para a aplicação desta técnica de balanceamento, uma vez que o método *undersampling* deleta dados úteis para o aprendizado do modelo. Todavia, é possível

que em um banco de dados com uma diferença menor entre o número de registros em cada classe, esse método pode causar menos prejuízo se comparado ao *oversampling* (PINHEIRO, 2022).

As evidências encontradas neste estudo reforçam a eficácia do algoritmo *Random Forest* como uma ferramenta promissora para previsão da DM. Além disso, todos os resultados obtidos estão de acordo com os achados descritos por Pinheiro (2022), Silva (2023) e Chang *et al.* (2022).

## CONSIDERAÇÕES FINAIS

Neste estudo, buscou-se avaliar o desempenho do algoritmo *Random Forest* na predição de Diabetes Mellitus, e os resultados obtidos foram promissores. O algoritmo mostrou-se eficiente segundo as métricas utilizadas, apresentando melhores resultados no conjunto de dados balanceados com *oversampling*. Nesse sentido, a utilização deste algoritmo pode servir como uma ferramenta de suporte na tomada de decisões na medicina pré-diagnóstica, identificando doenças em estágios iniciais, evitando assim o agravamento do quadro clínico do paciente.

Entretanto, é importante destacar que qualquer modelo preditivo, incluindo o *Random Forest*, deve ser validado em diferentes populações e ambientes clínicos para garantir sua generalização e aplicabilidade em cenários reais. Ainda assim, ao analisarmos os resultados obtidos, é evidente a proximidade com os achados descritos na literatura acadêmica. Essa consonância reforça a relevância e validade dos dados obtidos neste estudo.

Por fim, é válido ressaltar que o banco de dados utilizado se trata de uma amostra pequena e que abrange somente pacientes do sexo feminino. Sendo assim, futuras pesquisas são necessárias para explorar conjuntos de dados maiores e mais abrangentes, bem como considerar o uso de outras técnicas relacionadas ao aprendizado de máquina para aprimorar o desempenho do algoritmo.

## REFERÊNCIAS

ALMEIDA, H. P. B. **Aprendizagem Computacional Aplicada à Detecção de Intrusões - Efeito das Técnicas de Balanceamento de Dados**. Orientador: Prof. Dr. Carlos Grilo. 2021. 110 f. Dissertação (Mestrado em Cibersegurança e Informática

Forense) - Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Leiria, 2021.

ARAUJO, J. V. G. A.; SANTOS, M.; GOMES, C. F. S. Desenvolvimento de um código em python para geração de matrizes de correlação de pearson com laços a partir de "N" variáveis tomadas duas a duas. **Simpósio de Pesquisa Operacional e Logística da Marinha (SPOLM)**, Rio de Janeiro, v.19, 2019.

CASTRO, R. M. F. *et al.* Diabetes mellitus e suas complicações - uma revisão sistemática e informativa. **Brazilian Journal of Health Review**, Curitiba, v.4, n.1, p. 3349-3391, 19 fev. 2021.

CHANG, V. *et al.* Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. **Neural computing & applications**, [s. l.], v.35, p.16157-16173, 24 mar. 2022.

Classification and Diagnosis of Diabetes. *In*: GENUTH, S. M.; PALMER, J. P.; NATHAN, D. M. **Diabetes in America**. [S. l.]: National Institute of Diabetes and Digestive and Kidney Diseases, 2018. p. 1-39

CUNHA, J. P. Z. **Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos**. Orientador: Profa. Dr. Viviana Giampaoli. 2019. 59 p. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 2019.

INTERNATIONAL DIABETES FEDERATION (IDF). **IDF Diabetes Atlas**, 10th edn. Brussels, Belgium. 2019. Disponível em: <https://diabetesatlas.org/atlas/tenth-edition/>. Acesso em: 12 jun. 2023.

LADEIA, F. J. M. *et al.* Análise do entendimento do paciente sobre programa de automonitoramento da diabetes. **Brazilian Journal of Health Review**, Curitiba, v.3, n.3, p.6965-6979, 26 jun. 2020.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, [s. l.], v.35, n.101, p.85-94, 2021.

MAIONE, C. **Balanceamento de dados com base em oversampling em dados transformados**. Orientador: Prof. Rommel Melgaço Barbosa. 2020. 141 f. Tese (Doutorado em Computação) - Instituto de Informática da Universidade Federal de Goiás, Goiânia, 2020.

MOREIRA, J. R. H. *et al.* Modelos de Aprendizado de Máquina na Predição de Diabetes Tipo 1 na Gestação usando Dados do Sistema Único de Saúde. **Anais do XXI Simpósio Brasileiro de Computação Aplicada à saúde (SBCAS)**, Porto Alegre, p.392-403, 15 jun. 2021.

MUZY, J. *et al.* Prevalência de diabetes mellitus e suas complicações e caracterização das lacunas na atenção à saúde a partir da triangulação de pesquisas. **Cadernos de Saúde Pública**, Rio de Janeiro, v.37, 28 maio 2021.

PINHEIRO, C. C. **PRÉ-DIAGNÓSTICO DE DIABETES MELLITUS EM MULHERES USANDO TÉCNICAS CLÁSSICAS DE APRENDIZADO DE MÁQUINA**. Orientador: Prof. Dr. Fidelis Zanetti de Castro. 2022. 50 f. Trabalho de Conclusão de Curso (Bacharel em Sistemas de Informação) - Instituto Federal do Espírito Santo, Serra, 2022.

SILVA, J. C. P. *et al.* Aprendizado de Máquina nos Jogos para Medicina: Uma Revisão Sistemática. **Simpósio brasileiro de computação aplicada à saúde (SBCAS)**, Niterói, v.19, p.70-81, 11 jun. 2019.

SILVA, L. F. **Modelo de aprendizado de máquina para predição de diabetes tipo 2 por meio de variáveis de fácil acesso**. Orientador: Danielle Arisa Caranti. 2023. 56 f. Dissertação (Mestrado em Ciências da Saúde) - Universidade Federal de São Paulo, Santos, 2023.

SOCIEDADE BRASILEIRA DE DIABETES. **Dados epidemiológicos do diabetes mellitus no Brasil**. 2019. Disponível em: <https://diabetes.org.br/wp-content/uploads/2021/06/SBD-Dados-Epidemiologicos-do-Diabetes-High-Fidelity.pdf>. Acesso em: 12 jun. 2023.

YOSHIDA, V. M. H.; GROTO, D.; GONÇALVES, D. B. **Delineamento Experimental**. Sorocaba: EDUNISO, 62 p., 2019.