

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Gabriel Maia Bezerra

**Uso de Aprendizado de Máquina para a Predição de Densidade
Mineral Óssea**

Fortaleza – CE

24 de fevereiro de 2023

Gabriel Maia Bezerra

Uso de Aprendizado de Máquina para a Predição de Densidade Mineral Óssea

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Sistemas Inteligentes
Aplicados à Saúde

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

24 de fevereiro de 2023

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

B574u Bezerra, Gabriel Maia.
Uso de Aprendizado de Máquina para a Predição de Densidade Mineral Óssea / Gabriel Maia Bezerra. -
2023.
125 f. : il.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus
Fortaleza, 2023.
Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.

1. Densidade Mineral Óssea. 2. Regressão. 3. SHAP. I. Título.

CDD 004



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Gabriel Maia Bezerra

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

**Orientador: Prof. Dr. Pedro Pedrosa Rebouças
Filho**

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Prof. Dr. Amauri Holanda de Souza Júnior
Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Profa. Dra. Atslands Rego da Rocha
Universidade Federal do Ceará (UFC)

Fortaleza – CE
24 de fevereiro de 2023

Agradecimentos

Primeiramente, à minha família, por seu amor, apoio, incentivo, esforço e compreensão incondicionais. Palavras nunca serão suficientes para manifestar meus agradecimentos a vocês.

A todos meus amigos, pelas conversas, amizade e apoio. Em especial, aos amigos do Laboratório de Processamento de Imagens, Sinais e Computação Aplicada (LAPISCO), pelo compartilhamento de conhecimentos, experiências e bons momentos.

Ao meu orientador, Prof. Dr. Pedro Pedrosa Rebouças Filho, pela orientação, confiança, ensinamentos e por contribuir com minha formação.

Aos membros da banca examinadora, por aceitarem o convite e cederem parte dos seus tempos para contribuírem com esta pesquisa.

A todos professores que já passaram pela minha vida, especialmente aos professores do Programa de Pós-Graduação em Ciência da Computação (PPGCC), pela dedicação, ensinamentos e por acreditarem na educação.

À minha psicóloga, por me fornecer ferramentas que auxiliaram minha caminhada nesta árdua jornada.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), por mais uma vez permitir meu crescimento acadêmico e por ser um lugar de aprendizados para milhares de alunos.

E a todos aqueles que contribuíram, direta ou indiretamente, para a realização desta pesquisa ou que tiveram um impacto positivo na minha formação.

Resumo

A osteoporose é um problema de saúde global caracterizado por baixa densidade óssea e deterioração do tecido ósseo, aumentando o risco de fraturas. A identificação precoce de baixa densidade mineral óssea (DMO) é crucial para reduzir os riscos, fornecendo métodos corretos de tratamento ou prevenção. A absorciometria por raios-X com dupla energia (DXA) é frequentemente usada para medir a DMO. Entretanto, é cara ou inacessível para muitos pacientes. Alternativamente, informações como idade, sexo, peso, altura, anamnese e circunferências corporais mostraram uma associação com a DMO e são baratas e fáceis de obter. Assim, esta pesquisa propõe uma abordagem de aprendizado de máquina para prever a DMO através de medidas antropométricas, anamnese, idade e sexo. Também é introduzido um conjunto de dados contendo 905 pacientes com seus respectivos atributos e valores de DMO. A abordagem deste trabalho avalia o desempenho de diferentes tipos de algoritmos de regressão através de validação cruzada aninhada. Por fim, interpretam-se as previsões do melhor modelo com SHAP (*SHapley Additive exPlanations*). Os resultados indicam um bom desempenho da abordagem avaliada, atingindo médias de erro absoluto médio (MAE) e erro percentual absoluto médio (MAPE) de 0,0771 g/cm² e 6,34%, respectivamente. Além disso, identifica-se um alto impacto do atributo “peso” na saída do modelo, o que condiz com a alta associação entre DMO e peso encontrada na literatura. Portanto, conclui-se que o método proposto pode se tornar uma ferramenta para profissionais da saúde predizerem valor de DMO, de forma barata e acessível.

Palavras-chave: densidade mineral óssea. regressão. SHAP.

Abstract

Osteoporosis is a global health problem characterized by low bone density and deterioration of bone tissue that increases the risk of fractures. Early detection of low bone mineral density (BMD) is critical to reducing risks by providing correct methods of treatment or prevention. Dual-energy x-ray absorptiometry (DXA) is often used to measure BMD. However, it is not affordable or accessible to many patients. Alternatively, information such as age, sex, weight, height, anamnesis, and body circumferences have shown an association with BMD and are inexpensive and easy to obtain. Thus, this research proposes a machine learning approach to predict BMD through anthropometric measurements, anamnesis, age, and sex. We also introduce a dataset containing 905 patients with their respective features and BMD values. Our approach evaluates the performance of different types of regression algorithms through nested cross-validation. Lastly, we interpret the model predictions with SHAP (SHapley Additive exPlanations). The results indicate a good performance of the evaluated approach, reaching average mean absolute error (MAE) and mean absolute percentage error (MAPE) of 0.0771 g/cm² and 6.34%, respectively. Furthermore, we identified a high impact of the attribute “weight” on the model output, which is consistent with the high association between BMD and weight found in the literature. Therefore, we conclude that the proposed method can become a tool for health professionals to predict BMD value in a cheap and accessible way.

Keywords: bone mineral density. regression. SHAP.

Lista de ilustrações

Figura 1	– Exemplo de contribuições e seus impactos na predição esperada do modelo. Elas explicam como o modelo sai do valor base $E[f(z)]$ para a predição $f(x)$. A contribuição de um atributo x_i é representada por ϕ_i , com exceção de ϕ_0 que simboliza o valor inicial sem a informação dos atributos. Além disso, $E[f(z) z_i = x_i]$ representa o valor esperado para a saída do modelo ao adicionar o atributo x_i	42
Figura 2	– Quantidade de respostas para o atributo Padrão Alimentar.	46
Figura 3	– Distribuição das respostas dos atributos categóricos com base no sexo.	49
Figura 4	– Histograma dos valores de DMO presentes no <i>dataset</i>	49
Figura 5	– Matriz de correlação de Pearson das variáveis finais do <i>dataset</i>	51
Figura 6	– Diagrama da metodologia do trabalho proposto. Na região azul, detalha-se o Processo de Aprendizado.	54
Figura 7	– Representação visual das divisões de dados realizadas na <i>nested cross-validation</i> , com três <i>outer folds</i> e dois <i>inner folds</i>	58
Figura 8	– Exemplo genérico do Diagrama de Diferença Crítica. ($CD = 0,9$) . . .	68
Figura 9	– Gráfico <i>beeswarm</i> genérico representando o comportamento dos valores SHAP de acordo com os valores de quatro atributos.	69
Figura 10	– Gráfico genérico representando o comportamento dos valores SHAP do atributo B de acordo com sua interação com o atributo C.	70
Figura 11	– Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-A. ($CD = 16,4$)	76
Figura 12	– Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-F. ($CD = 16,4$)	80
Figura 13	– Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-M. ($CD = 16,4$)	83
Figura 14	– Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-A). . . .	87
Figura 15	– Comportamento dos valores SHAP de acordo com a relação entre frequência de atividade e peso (referentes à abordagem selecionada no experimento E1-A).	88
Figura 16	– Comportamento dos valores SHAP de acordo com a relação entre idade e sexo de nascimento (referentes à abordagem selecionada no experimento E1-A).	89

Figura 17 – Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-F). . . .	90
Figura 18 – Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-M). . . .	91
Figura 19 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-A. (CD = 16,4)	109
Figura 20 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-A. (CD = 16,4)	111
Figura 21 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-A. (CD = 16,4)	113
Figura 22 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-F. (CD = 16,4)	115
Figura 23 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-F. (CD = 16,4)	117
Figura 24 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-F. (CD = 16,4)	119
Figura 25 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-M. (CD = 16,4)	121
Figura 26 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-M. (CD = 16,4)	123
Figura 27 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-M. (CD = 16,4)	125

Lista de tabelas

Tabela 1	– As cinco Unidades Federativas com menos estabelecimentos do SUS com equipamento de densitometria óssea.	26
Tabela 2	– Resumo das características do presente trabalho e dos trabalhos relacionados	27
Tabela 3	– Atributos originais do conjunto de dados e suas categorias, subcategorias e respostas encontradas.	44
Tabela 4	– Atributos finais do conjunto de dados e suas categorias, subcategorias e possíveis valores.	47
Tabela 5	– Média, desvio padrão (DP) e intervalo dos atributos quantitativos presentes na base de dados. As informações são apresentadas para todo o conjunto e para os conjuntos de pacientes masculinos e femininos, separadamente.	48
Tabela 6	– Siglas dos diferentes experimentos realizados e especificações do conjunto de atributos utilizado e do grupo de pacientes selecionado, em relação ao sexo de nascimento.	56
Tabela 7	– Hiperparâmetros ajustados por <i>Random Search</i> e seus respectivos modelos e espaços de busca.	60
Tabela 8	– Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	74
Tabela 9	– Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	75
Tabela 10	– Métricas obtidas pela abordagem selecionada no E1-A ao avaliar as amostras de acordo com intervalos de DMO. As amostras de teste são ordenadas pelo valor de DMO e então divididas em quatro subconjuntos de mesmo tamanho.	77
Tabela 11	– Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	78

Tabela 12 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	79
Tabela 13 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	81
Tabela 14 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.	82
Tabela 15 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.	84
Tabela 16 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-F e de cada experimento EX-A, analisando apenas amostras do sexo feminino. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.	85
Tabela 17 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-M e de cada experimento EX-A, analisando apenas amostras do sexo masculino. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.	85
Tabela 18 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	107
Tabela 19 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	108

Tabela 20 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	109
Tabela 21 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	110
Tabela 22 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	111
Tabela 23 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	112
Tabela 24 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	113
Tabela 25 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	114
Tabela 26 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	115

Tabela 27 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	116
Tabela 28 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	117
Tabela 29 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	118
Tabela 30 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	119
Tabela 31 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	120
Tabela 32 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	121
Tabela 33 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	122

Tabela 34 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	123
Tabela 35 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.	124

Lista de Algoritmos

1	Pseudocódigo para <i>Nested Cross-Validation</i>	62
---	--	----

Lista de abreviaturas e siglas

CD	<i>Critical Difference</i> (Diferença Crítica)
CEP	Comitê de Ética em Pesquisa
CV	<i>Cross-validation</i> (Validação Cruzada)
DMO	Densidade Mineral Óssea
DP	Desvio padrão
DT	<i>Decision Tree</i> (Árvore de Decisão)
DXA	<i>Dual Energy X-ray Absorptiometry</i> (Absorciometria por Raios-X de Dupla Energia)
EN	<i>Elastic Net</i>
IA	Inteligência Artificial
IFCE	Instituto Federal de Educação, Ciência e Tecnologia do Ceará
INJC	Instituto de Nutrição Josué de Castro
ISAK	<i>International Society for the Advancement of Kinanthropometry</i>
k-NN	<i>k-Nearest Neighbors</i> (k-Vizinhos Mais Próximos)
LabDAFEE	Laboratório de Desenvolvimento de Alimentos para Fins Especiais e Educacionais
LIME	<i>Local Interpretable Model-Agnostic Explanations</i>
LR	<i>Linear Regression</i> Regressão Linear
MAE	<i>Mean absolute error</i> (Erro absoluto médio)
MAPE	<i>Mean absolute percentage error</i> (Erro percentual absoluto médio)
ME	<i>Maximum error</i> (Erro máximo)
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
MLP	<i>Multilayer Perceptron</i> (Perceptron de Múltiplas Camada)
MMQ	Método dos Mínimos Quadrados

R ²	Coeficiente de determinação
RBF	<i>Radial Basis Function</i>
RF	<i>Random Forest</i> (Floresta Aleatória)
RMSE	<i>Root mean square error</i> (Raiz do erro quadrático médio)
SHAP	<i>SHapley Additive exPlanations</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte)
TA	Tamanho do arquivo
Tte	Tempo de teste
Ttr	Tempo de treino
UFRJ	Universidade Federal do Rio de Janeiro
XGBoost	<i>eXtreme Gradient Boosting</i>

Sumário

1	INTRODUÇÃO	25
1.1	Trabalhos Relacionados	27
1.2	Objetivos Geral e Específicos	29
1.3	Produção científica	30
1.4	Estrutura da Dissertação	31
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	Métodos de Regressão	33
2.1.1	Regressão Linear	34
2.1.2	<i>Elastic Net</i>	34
2.1.3	k-Vizinhos Mais Próximos	35
2.1.4	Perceptron de Múltiplas Camadas	36
2.1.5	Máquina de Vetores de Suporte	37
2.1.6	Árvore de Decisão	38
2.1.7	Floresta Aleatória	39
2.1.8	<i>eXtreme Gradient Boosting</i>	40
2.2	Interpretabilidade com <i>SHapley Additive exPlanations</i>	41
3	CONJUNTO DE DADOS	43
3.1	Coleta	43
3.2	Atributos	44
3.2.1	Atributos Originais	44
3.2.2	Atributos Finais	45
3.3	Análise do Conjunto de Dados	47
4	METODOLOGIA	53
4.1	Configuração dos experimentos	54
4.2	Pré-processamento	56
4.3	Processo de Aprendizado	57
4.4	Avaliação dos Resultados	62
4.4.1	Métricas	63
4.4.1.1	MAE	63
4.4.1.2	RMSE	64
4.4.1.3	MAPE	64
4.4.1.4	ME	65
4.4.1.5	R^2	65

4.4.1.6	Tempo de treino	65
4.4.1.7	Tempo de teste	66
4.4.1.8	Tamanho do arquivo	66
4.4.2	Testes estatísticos	66
4.5	Interpretabilidade	68
5	RESULTADOS E DISCUSSÃO	73
5.1	Experimentos Principais	73
5.1.1	E1-A	73
5.1.2	E1-F	77
5.1.3	E1-M	80
5.2	Experimentos Adicionais	83
5.3	Interpretação	86
6	CONCLUSÃO	93
6.1	Limitações	94
6.2	Trabalhos Futuros	95
	REFERÊNCIAS	97
	APÊNDICES	105
	APÊNDICE A – RESULTADOS COMPLETOS DOS EXPERIMEN-	
	TOS ADICIONAIS	107

1 Introdução

A osteoporose é uma doença que afeta o esqueleto e é caracterizada pela baixa massa óssea e cálcio, o que leva a uma deterioração da estrutura óssea (Organização Mundial da Saúde, 2003). O efeito clínico primário dessa doença é a ocorrência de fraturas por baixo impacto (RADOMINSKI et al., 2017). Sua incidência vem aumentando nos países desenvolvidos e em desenvolvimento, tornando-se assim uma preocupação de saúde global. Estima-se que mais de 200 milhões de pessoas são afetadas pela osteoporose em todo o mundo, causando mais de 8,9 milhões de fraturas por ano (KANIS et al., 2007). Além disso, as complicações de tais fraturas e os custos relacionados a cirurgias, cuidados prolongados, invalidez e hospitalizações acarretam despesas econômicas consideráveis (Organização Mundial da Saúde, 2003). Um exemplo de tais fraturas são as fraturas de quadril. Elas frequentemente requerem hospitalização e são consideradas um dos tipos mais críticos de fraturas. Na América Latina, mais de 23% dos pacientes morrem no primeiro ano após uma fratura de quadril (AZIZIYEH et al., 2019).

Apesar de ser uma doença comum e com possíveis graves consequências, a osteoporose costuma ser uma doença clinicamente silenciosa (KLING; CLARKE; SANDHU, 2014). Por isso, a baixa massa óssea continua sendo uma condição subdiagnosticada e subtratada (MILLER, 2016). Uma detecção precoce dessa doença e uma avaliação antecipada da densidade mineral óssea (DMO) são úteis para iniciar o tratamento ou as medidas de prevenção adequadas (SÖZEN; ÖZİŞIK; BAŞARAN, 2017). Estudos mostram que existem recomendações que podem retardar a perda de DMO e até prevenir osteoporose. Algumas dessas medidas são: ajuste da dieta (principalmente relacionado a cálcio e vitamina D), prática de atividade física, não fumar, evitar o consumo excessivo de álcool, uso de fármacos que auxiliam a saúde óssea, ou até a interrupção do uso de medicamentos que afetam a DMO (Organização Mundial da Saúde, 2003).

O valor da DMO é uma fonte de informação sobre a saúde óssea do paciente que pode ser usada para diagnosticar a osteoporose e determinar o risco de fraturas (National Institutes of Health, 2018a). Esse dado também identifica a osteopenia, uma condição que antecede a osteoporose, quando há uma baixa massa óssea, mas não tão baixa quanto na osteoporose (KARAGUZEL; HOLICK, 2010). Portanto, a DMO é uma informação importante para detectar essas doenças e também para estimar as chances de desenvolvê-las. Abrahamsen et al. (2006) concluíram que o valor de DMO obtido em mulheres saudáveis após a menopausa é um forte preditor de chances de desenvolver osteoporose nos 10 anos seguintes. Portanto, a densidade mineral óssea é crucial para as tomadas de decisão de um profissional de saúde da área.

Há vários métodos para determinar a DMO; entre eles, um dos mais conhecidos é a absorciometria por raios-X de dupla energia (DXA, do inglês *dual energy X-ray absorptiometry*). Embora o exame de DXA seja considerado o método mais confiável para estimar o DMO (Organização Mundial da Saúde, 2003), ele apresenta desafios em relação à sua falta de portabilidade e custo (HIEN et al., 2005; NAYAK; ROBERTS; GREENSPAN, 2011). Além disso, a sua presença é escassa no sistema público de saúde brasileiro. Dados do SUS (Sistema Único de Saúde), de dezembro de 2022, indicam que há apenas 61 estabelecimentos do SUS na região norte com o equipamento necessário para medir DMO (CNES, 2022), tornando-a a região com a menor presença desse equipamento. Já em relação a todas Unidades Federativas do Brasil, a Tabela 1 apresenta as cinco delas com menor quantidade de estabelecimentos do SUS que possuem um equipamento de medir DMO. Em particular, o estado do Acre contém apenas um estabelecimento com o equipamento. Portanto, pessoas de baixa renda, comunidades rurais, comunidades indígenas, comunidades afastadas de grandes centros urbanos e populações de países subdesenvolvidos podem ter dificuldade de aquisição ou de acesso a este procedimento.

Tabela 1 – As cinco Unidades Federativas com menos estabelecimentos do SUS com equipamento de densitometria óssea.

Unidade Federativa	Estabelecimentos do SUS com o equipamento
Acre	1
Amapá	3
Rondônia	4
Sergipe	6
Roraima	7

Fonte: CNES (2022)

Em razão dos problemas mencionados em obter o valor de DMO, surge a necessidade de desenvolver métodos alternativos para estimar essa informação. Nos últimos anos, os algoritmos de aprendizado de máquina (ML, do inglês *machine learning*) estão se tornando mais populares devido a suas características como versatilidade e capacidade de estabelecer conexões entre atributos e predições (BENKE; BENKE, 2018; DEO, 2015). A habilidade dos métodos de ML de reconhecer padrões os faz ótimas ferramentas de auxílio médico. Eles vêm sendo utilizados para tarefas como diagnóstico de doenças (MYSZCZYNSKA et al., 2020) e predição de resultados de exames (CABITZA; BANFI, 2018).

Portanto, para que um algoritmo de ML seja capaz de realizar uma predição de DMO, ele necessita de outras informações do paciente. O objetivo então do algoritmo é encontrar padrões nesses atributos que estejam relacionados à saúde óssea do indivíduo. Logo, obter esses dados é uma etapa crucial para a construção de um modelo de ML. Pesquisadores já identificaram associações entre DMO e diversos outros atributos, mas esta pesquisa visa utilizar apenas dados que são fáceis e baratos de se obter. Dessa forma, contorna-se alguns dos obstáculos encontrados no uso de DXA.

Dois atributos popularmente associados à saúde óssea são a idade e o sexo designado no nascimento (LAURENT et al., 2019). Outros dados que também podem ser obtidos por um simples questionário e têm associação com DMO são os hábitos de vida do paciente, a presença de doenças crônicas e a condição do ciclo menstrual (National Institutes of Health, 2019). Além disso, com instrumentos simples, é possível mensurar medidas antropométricas, que também apresentam associação com osteoporose (HASANI-RANJBAR et al., 2019; SRIDHARAN et al., 2020). Portanto, formula-se a hipótese que um conjunto de dados, contendo diversos indivíduos com os respectivos atributos das categorias mencionadas, pode ser utilizado no processo de aprendizado de um algoritmo de ML responsável por prever DMO. Dessa forma, há uma possibilidade de se criar uma ferramenta acessível para estimar a condição óssea de um paciente, que por sua vez, pode auxiliar profissionais da saúde na decisão por medidas preventivas ou tratamentos.

Alguns trabalhos da literatura também procuram técnicas para estimar DMO ou classificar a presença de osteoporose. Entretanto, muitos deles utilizam dados mais difíceis de se obter ou restringem o público-alvo da pesquisa. A seção seguinte apresenta alguns desses trabalhos e discute sobre os obstáculos encontrados nos mesmos.

1.1 Trabalhos Relacionados

Esta seção apresenta trabalhos relacionados que utilizam o aprendizado de máquina para estimar o valor de DMO, auxiliar na detecção de osteoporose ou estimar riscos relacionados aos ossos. Não foram encontradas pesquisas de predição de DMO que disponibilizassem publicamente os conjuntos de dados utilizados. A Tabela 2 resume as características da presente pesquisa e dos trabalhos mencionados nesta seção.

Tabela 2 – Resumo das características do presente trabalho e dos trabalhos relacionados

Trabalho	Variável alvo	Entrada	Dataset
Presente trabalho	DMO do corpo inteiro	Idade, sexo, medidas antropométricas e anamnese	905 pessoas
Sridharan et al. (2020)	DMO do fêmur e da coluna lombar	Medidas antropométricas	308 pessoas
Ho et al. (2021)	DMO do fêmur	Imagem de raio-x da pelve	3472 imagens
Iliou, Anagnostopoulos e Anastassopoulos (2014)	Presença de osteoporose	Idade, sexo, altura e peso	3426 pessoas
Kim et al. (2013)	Risco de osteoporose	Idade, medidas antropométricas e anamnese	1674 pessoas
Yang et al. (2021)	Presença de osteoporose	Idade, medidas antropométricas, anamnese e testes hematológicos e bioquímicos	5982 pessoas
Kilic e Hosgormez (2015)	Presença de osteoporose ou osteopenia	Idade e exames de saúde óssea	350 pessoas

Fonte: adaptado de Sridharan et al. (2020), Ho et al. (2021), Iliou, Anagnostopoulos e Anastassopoulos (2014), Kim et al. (2013), Yang et al. (2021), Kilic e Hosgormez (2015).

Sridharan et al. (2020) investigaram a relação entre indicadores antropométricos e DMO do colo do fêmur e da coluna lombar. Esse estudo foi realizado com 308 mulheres

na pós-menopausa, moradoras de comunidades rurais do sul da Índia e com mais de 50 anos. Os autores encontraram uma correlação positiva entre peso e DMO em ambas regiões. Além disso, aplicaram regressão logística multivariada e identificaram que o peso e a idade foram significativos para prever osteoporose no fêmur, enquanto que o peso e a circunferência da cintura foram os melhores preditores para osteoporose na coluna lombar. Porém, os pesquisadores foram rigorosos nos critérios de seleção do *dataset*, pois removeram participantes que tinham doenças renais crônicas, doenças hepáticas crônicas, câncer, insuficiência cardíaca ou histórico de acidente vascular cerebral. Também foram removidas pessoas com diagnóstico prévio de osteoporose, histórico de fraturas, ou histórico de uso de medicamentos que interferem na saúde óssea. Portanto, por focar em um grupo demográfico muito restrito e por impor rigorosas exigências na seleção dos participantes, há uma grande limitação no uso do método proposto pelos autores.

Alguns estudos avaliam o uso de imagens para estimar DMO. Ho et al. (2021) desenvolveram uma rede neural de aprendizado profundo, chamada *DeepDXA*, para prever DMO a partir de um exame de raio-x. Eles utilizaram um conjunto de dados com 3472 imagens de raio-x da pelve e com os valores correspondentes de DMO do fêmur de cada paciente. Com os valores estimados de DMO do fêmur, a abordagem deles classifica o paciente como tendo osteoporose, osteopenia ou sendo saudável. Dessa forma, alcançaram uma acurácia de 88% para a classificação de osteoporose. Entretanto, há obstáculos para o uso dessa abordagem, já que necessita-se de um raio-x da pelve e os autores utilizam um *software* proprietário para segmentar a região do fêmur antes de enviar para a *DeepDXA*.

Iliou, Anagnostopoulos e Anastassopoulos (2014) avaliaram o uso do aprendizado de máquina e seleção de atributos para detectar a osteoporose. Eles utilizaram um conjunto de dados composto de 3426 pacientes. Os autores concluíram que somente a idade e o peso alcançaram resultados satisfatórios na previsão da osteoporose. Entretanto, seu conjunto de dados era composto principalmente de mulheres (98%). Além disso, os autores não forneceram o conjunto de dados e não apresentaram a distribuição dos dados.

Kim et al. (2013) propuseram o uso de aprendizado de máquina para prever o risco de osteoporose. O conjunto de dados utilizado no estudo compreendia 1674 mulheres na pós-menopausa, que incluía dados clínicos e demográficos, tais como idade, peso, circunferência da cintura, histórico de fraturas e osteoartrose, entre outras informações. Atingiram uma precisão de 76,7% e 0,827 de área sob a curva ao utilizar o classificador *Support Vector Machine* (SVM). Entretanto, o escopo desse estudo é direcionado a mulheres na pós-menopausa, o que limita a porção demográfica da população na qual o método pode ser aplicado.

Yang et al. (2021) desenvolveram um modelo de aprendizado de máquina para prever a osteoporose. Os autores avaliaram vários métodos de ML, alcançando 0,843 como a mais alta área sob a curva característica de operação do receptor para homens e 0,811

para mulheres. O conjunto de dados foi composto de 2929 pacientes do sexo feminino e 3053 indivíduos do sexo masculino. Entretanto, este estudo também se concentra em uma faixa etária rigorosa e usa como atributos vários testes hematológicos e bioquímicos.

Kilic e Hosgormez (2015) propuseram usar seis conjuntos de atributos de densitometria óssea e a idade do paciente (resultando em 24 atributos de 350 mulheres na pós-menopausa) para realizar uma classificação de 3 classes: osteoporose, osteopenia e grupo de controle. Os autores concluíram que a seleção de cinco medidas de DMO e cinco valores de T-Score alcançou o melhor resultado na identificação da osteoporose usando Floresta Aleatória. No entanto, eles também concentraram seu estudo em mulheres na pós-menopausa. Além disso, os dados foram compostos de medidas de equipamentos sofisticados que já são usados para o diagnóstico da osteoporose.

Em síntese, percebe-se que há pesquisas que se propõem a utilizar aprendizado de máquina para a predição de fatores relacionados à saúde dos ossos. Entretanto, grande parte delas apresenta critérios rigorosos quanto à criação do conjunto de dados, como uma faixa curta de idade, limitação ao sexo do paciente e análise de condições pré-existentes. Tais especificações limitam a aplicação das abordagens sugeridas a grupos demográficos restritos. Além disso, as pesquisas encontradas relacionadas a DMO não disponibilizam os seus conjuntos de dados, o que dificulta a reprodução dos resultados e o uso em situações reais. Por fim, alguns dos trabalhos investigados alcançam bons resultados, mas fazem uso de informações mais difíceis de se obter, como testes laboratoriais, exames de imagem e análises de *softwares* proprietários. Portanto, a presente pesquisa visa contribuir com avanços contra os pontos negativos identificados no estado da arte. Os objetivos para atingir essas contribuições são mencionados na seção seguinte.

1.2 Objetivos Geral e Específicos

O objetivo geral desta pesquisa é propor uma abordagem que utiliza aprendizado de máquina para predizer a DMO através de informações básicas de um paciente, como dados de identificação, de anamnese e medidas antropométricas.

Em seu sentido mais particular, a pesquisa tem como objetivos específicos as seguintes contribuições:

- Avaliar diferentes métodos de predição de DMO usando atributos acessíveis: idade, sexo designado no nascimento, medidas antropométricas e dados de anamnese;
- Apresentar um novo conjunto de dados para a predição de DMO, consistindo de 905 pacientes dos sexos feminino e masculino (idades entre 18 e 65 anos) com uma grande variedade de medidas corporais e hábitos de vida;

- Analisar se modelos de ML treinados apenas em pacientes de um sexo específico têm melhor desempenho do que modelos treinados na base de dados completa;
- Investigar o efeito que a remoção de grupos de atributos tem na performance das abordagens testadas;
- Interpretar e analisar o impacto dos atributos na saída do modelo, o que pode aumentar a confiança no algoritmo e trazer *insights* para outros pesquisadores.

1.3 Produção científica

Durante este projeto de mestrado, os seguintes trabalhos científicos foram aceitos e publicados ou apresentados:

- BEZERRA, G. M. et al. Estimating bone mineral density based on age, sex, and anthropometric measurements. In: BRAZILIAN INSTITUTE OF DATA SCIENCE. *11th Brazilian Conference on Intelligent Systems (BRACIS 2022)*. Campinas, Brasil, 2022. **(Primeiro autor. Artigo referente à dissertação. Qualis A4)**
- OHATA, E. F. et al. Automatic detection of covid-19 infection using chest x-ray images through transfer learning. *IEEE/CAA Journal of Automatica Sinica*, IEEE, v. 8, n. 1, p. 239–248, 2020. **(Segundo autor. Qualis A1. Fator de impacto 7.847)**
- IVO, R. F. et al. Non-grain oriented electrical steel photomicrograph classification using transfer learning. *Journal of Materials Research and Technology*, Elsevier, v. 9, n. 4, p. 8580–8591, 2020. **(Terceiro autor. Qualis A1. Fator de impacto 6.267)**
- SILVA, F. H. dos S. et al. A novel feature extractor for human action recognition in visual question answering. *Pattern Recognition Letters*, Elsevier, v. 147, p. 41–47, 2021. **(Segundo autor. Qualis A2. Fator de impacto 4.757)**
- OHATA, E. F. et al. A novel transfer learning approach for the classification of histological images of colorectal cancer. *The Journal of Supercomputing*, Springer, v. 77, n. 9, p. 9494–9519, 2021. **(Terceiro autor. Qualis A2. Fator de impacto 2.557)**
- NUNES, V. X. et al. Adaptive level set with region analysis via mask r-cnn: A comparison against classical methods. In: IEEE. *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, Escócia, 2020. p. 1–8. **(Quarto autor. Qualis A1)**
- ALMEIDA, J. S. et al. An optimized approach to covid-19 detection via cough audio signals and dimensionality reduction. In: SPRINGER. *The 28th International Conference on Systems, Signals and Image Processing (IWSSIP 2021)*. Bratislava, Eslováquia, 2021. **(Segundo autor. Qualis A3)**

- ANDRADE, J. P. B. et al. Sistema baseado em microsserviços para identificação automática de impressões digitais. In: BRAZILIAN COMPUTER SOCIETY (SBC). *XII Brazilian Symposium on Computing Systems Engineering (SBESC)*. Fortaleza, Brasil, 2022. **(Terceiro autor. Qualis A4)**

1.4 Estrutura da Dissertação

Esta dissertação está estruturada em seis capítulos, seguidos por uma seção de apêndices. Neste Capítulo 1, apresentou-se a contextualização do problema, junto com as motivações da pesquisa, trabalhos relacionados e objetivos. Já no Capítulo 2, listam-se os algoritmos de ML utilizados e discute-se sobre a técnica de interpretação. Em seguida, no Capítulo 3, descreve-se o conjunto de dados utilizado no trabalho, detalhando a sua coleta e seus atributos. Logo então, o Capítulo 4 apresenta a metodologia desta pesquisa, explicando os experimentos investigados e as etapas de pré-processamento, aprendizado, avaliação e interpretação. Na sequência, os resultados e as discussões são debatidos no Capítulo 5. Já no Capítulo 6, expõem-se as conclusões obtidas na pesquisa, destacando as limitações da mesma e sugerindo trabalhos futuros. Por fim, no Apêndice A, apresentam-se tabelas e figuras com resultados adicionais deste trabalho.

2 Fundamentação Teórica

Este capítulo destina-se a discutir sobre alguns conceitos relevantes para o desenvolvimento desta pesquisa. Primeiramente, descreve-se brevemente o conceito de regressão. Logo então, são apresentados os algoritmos de regressão que foram selecionados para os testes de predição de DMO. Por último, discute-se sobre a definição de interpretabilidade, a sua importância e a abordagem utilizada para essa etapa.

2.1 Métodos de Regressão

O objetivo da regressão é prever o valor de uma variável alvo contínua y , dado o valor de um vetor D -dimensional x de variáveis de entrada (BISHOP; NASRABADI, 2006). Nesta pesquisa, a variável alvo y é o valor da DMO. Já o vetor de entrada x é o conjunto de valores dos outros atributos do paciente. A dimensão D é dada pela quantidade de atributos usados na entrada.

Mais especificamente, a regressão é uma técnica de aprendizado supervisionado. Portanto, a partir de um conjunto de treino contendo N amostras, onde cada amostra possui seu vetor x de atributos e sua variável alvo y , a regressão visa estimar uma função $f(x)$ que melhor descreve a relação entre x e y (BISHOP; NASRABADI, 2006). Para avaliar o quão bem a função modelada consegue prever a variável alvo para um vetor de atributos correspondente, geralmente utiliza-se um conjunto de teste, que contém amostras que não estão presentes no conjunto de treino, mas que possuem as mesmas variáveis dele.

Existem diversos métodos que podem ser aplicados em problemas de regressão. Esta seção apresenta alguns desses métodos, os quais foram selecionados para os experimentos de predição de DMO. Nomeadamente, os algoritmos utilizados nesta pesquisa são:

- Regressão Linear pelo Método dos Mínimos Quadrados;
- *Elastic Net*;
- k-Vizinhos Mais Próximos;
- Perceptron de Múltiplas Camadas;
- Máquina de Vetores de Suporte (com os *kernels* Linear, Polinomial, Sigmoidal e RBF);
- Árvore de Decisão;
- Floresta Aleatória;
- *eXtreme Gradient Boosting*.

Os critérios de seleção desses algoritmos foram baseados na frequência que eles são encontrados na literatura ou em aplicações reais e, principalmente, nos pontos fortes que

cada um deles possui. Alguns desses pontos fortes, junto com alguns pontos fracos e breves descrições de cada método, são discutidos nas subseções seguintes.

2.1.1 Regressão Linear

A Regressão Linear (LR, do inglês *Linear Regression*) é o método de regressão mais simples. Ela tem como objetivo ajustar uma reta ao conjunto de dados. Nesta pesquisa, a abordagem utilizada para a LR faz uso do Método dos Mínimos Quadrados (MMQ). Essa técnica visa encontrar os parâmetros da reta a ser ajustada, de forma a minimizar a soma dos quadrados dos resíduos (HASTIE et al., 2009). O resíduo, por sua vez, é a diferença entre o valor estimado e o valor observado da variável alvo.

A LR possui algumas vantagens e desvantagens. Em particular, algumas vantagens são:

- *Rápida no treino e na predição.* Por usar MMQ, os parâmetros são estimados através de uma equação com complexidade $O(ND^2)$, logo a etapa de treino costuma ser rápida. Já para a predição, basta utilizar a função da reta estimada, que tem complexidade $O(D)$.
- *Fácil interpretação.* Como a LR estima um coeficiente para cada atributo (além do coeficiente linear da reta), torna-se simples identificar o impacto que cada atributo tem na saída do modelo (HASTIE et al., 2009). Por ser mais compreensível, torna-se um método mais popular e mais facilmente aceito por especialistas.
- *Baixo uso de armazenamento.* Já que o modelo treinado armazena apenas os coeficientes da reta, a LR não toma muito espaço de armazenamento, o que facilita a sua aplicação.

Por outro lado, a LR possui desvantagens como:

- *Linearidade.* Por ajustar uma reta aos dados, a LR tende a ter uma baixo poder preditivo para dados que têm um comportamento não-linear.
- *Sensível a outliers.* Como o MMQ tenta minimizar os resíduos, a presença de *outliers* pode causar um alto impacto nos parâmetros estimados.
- *Colinearidade.* Na presença de *features* perfeitamente colineares ou próximas disso, o MMQ pode falhar no cálculo dos parâmetros ou ficar muito sensível a ruídos, causando a explosão de alguns parâmetros (BONACCORSO, 2017).

2.1.2 Elastic Net

Elastic Net (EN) é um método de regressão linear que combina as regularizações L2 (*Ridge*) e L1 (*Lasso*) (ZOU; HASTIE, 2005). Essa técnica tem como objetivo ajustar uma reta ao conjunto de dados, mas aplicando um termo de regularização, evitando valores

altos para os coeficientes da reta. Esse termo aplicado pela EN é uma combinação linear das penalidades L1 e L2. É necessário definir um hiperparâmetro que controla se a EN tem um comportamento mais semelhante a uma regressão *Ridge*, uma regressão *Lasso*, ou um equilíbrio entre as duas.

A regressão Ridge (HOERL; KENNARD, 1970) adiciona uma penalidade L2 que é igual ao quadrado da magnitude dos coeficientes. Logo, esse termo penaliza valores altos para os coeficientes, o que desencoraja o modelo a se ajustar ao ruído encontrado nos dados. Assim, as magnitudes dos coeficientes são reduzidas para próximo de zero, mas sem chegar a zero. Já na regressão Lasso (TIBSHIRANI, 1996), adiciona-se uma penalidade L1 que é igual ao valor absoluto do coeficiente. Essa técnica permite reduzir a magnitude de alguns coeficientes até zero, conseguindo funcionar como um método de seleção de atributos.

Algumas das características positivas de se utilizar EN para tarefas de regressão são:

- *Combinação de dois métodos.* Ao combinar as regressões Ridge e Lasso, por muitas vezes a EN consegue pegar os benefícios de cada uma dessas técnicas e performar melhor que elas por conta própria (ZOU; HASTIE, 2005).
- *Atributos correlacionados.* A EN tem a vantagem de lidar melhor em casos em que há atributos altamente correlacionados (FRIEDMAN; HASTIE; TIBSHIRANI, 2010).
- *Eficiente.* Assim como a LR, a EN tem uma rápida etapa de treinamento, consegue prever novas amostras em um curto período e ocupa pouco espaço de armazenamento.

Já algumas desvantagens encontradas no uso da EN são:

- *Linearidade.* Por ser um método linear, a EN pode ter um baixo poder preditivo em conjuntos de dados com comportamento não-linear.
- *Sensível à escala dos atributos.* Por aplicar termos de regularização, a EN é sensível à escala dos atributos. A regularização acaba tendo um efeito menor em *features* com alta escala. Logo, é recomendado pré-processar os dados para que estejam em escalas semelhantes.

2.1.3 k-Vizinhos Mais Próximos

k-Vizinhos Mais Próximos (k-NN, do inglês *k-Nearest Neighbors*) é um algoritmo não-paramétrico que pode ser utilizado para problemas de regressão (BISHOP; NASRABADI, 2006). Primeiramente, o algoritmo calcula a distância entre uma nova amostra e todas as amostras de treino. As k amostras mais próximas são consideradas a vizinhança do novo dado (BISHOP; NASRABADI, 2006). Então, o rótulo da nova amostra é dado pela média

dos rótulos da vizinhança. Alternativamente, o cálculo dessa média pode ser feito de forma ponderada, onde cada amostra vizinha contribui de forma inversamente proporcional à distância da nova amostra (CUNNINGHAM; DELANY, 2021).

Existem pontos positivos e negativos em se utilizar o método k-NN. Alguns dos positivos são:

- *Não-linearidade.* O k-NN não faz suposições sobre a distribuição dos dados e consegue modelar funções não-lineares (HASTIE et al., 2009).
- *Sem treinamento.* O k-NN praticamente não tem uma etapa de treinamento, pois apenas carrega os dados do conjunto de treino. Isso faz com que não seja necessário esperar por uma etapa de treino, tornando a integração de novos dados no modelo muito mais simples e rápida.
- *Simples e intuitivo.* Assim como a LR, é um método mais intuitivo (HASTIE et al., 2009), o que pode torná-lo mais facilmente aceito em aplicações na área da saúde.

Porém, o k-NN também apresenta pontos negativos como:

- *Predição custosa.* Principalmente em conjuntos de dados maiores, a etapa de predição do k-NN é demorada (SONG et al., 2017). Isso ocorre pois é necessário calcular a distância do vetor de *features* da nova amostra para todos os vetores de *features* do conjunto de treino.
- *Armazenamento proporcional ao tamanho do dataset.* Como o k-NN precisa de todos os dados de treino para as predições, o conjunto de treino precisa ficar armazenado mesmo depois da etapa de treinamento. Logo, isso pode ser proibitivo para máquinas com poucos recursos computacionais e problemas com grandes *datasets*.
- *Sensível à escala dos atributos.* Como é um algoritmo baseado em distâncias, a escala dos atributos impacta diretamente nas predições. Caso alguma *feature* tenha uma escala maior que a outra, essa *feature* terá um impacto maior na seleção da vizinhança e, portanto, será mais importante na predição.

2.1.4 Perceptron de Múltiplas Camadas

O Perceptron de Múltiplas Camadas (MLP, do inglês *Multilayer Perceptron*) é um algoritmo de ML que pode ser utilizado para tarefas de regressão. A MLP é uma rede neural artificial que utiliza o conceito de camadas de neurônios para modelar uma função. Ela possui uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (HAYKIN, 2001). Os atributos de uma amostra são enviados aos neurônios da camada de entrada, que então se propagam pela rede por meio de pesos sinápticos que conectam cada um dos neurônios de uma camada aos neurônios da próxima. A saída da

rede, que é o rótulo da amostra, é dada pelo neurônio da camada de saída (GÉRON, 2022).

Na etapa de treinamento, os pesos da MLP são inicializados aleatoriamente. Logo então, inicia-se um processo iterativo em que os dados de treino são apresentados para o modelo e suas previsões são comparadas com os rótulos reais. A diferença entre esses dois valores, chamada de erro, é retropropagada para atualizar os pesos através de um algoritmo de otimização (BISHOP; NASRABADI, 2006).

A aplicação dessa rede neural artificial possui vantagens tais quais:

- *Não-linearidade.* A MLP consegue modelar comportamentos não-lineares, pois ela introduz não-linearidades ao utilizar funções de ativação não-lineares nas saídas dos neurônios das camadas ocultas (HAYKIN, 2001).
- *Lida bem com alta dimensionalidade.* A MLP é competente em problemas de alta dimensionalidade (grande quantidade de atributos) devido aos neurônios das camadas ocultas e suas funções de ativação, que conseguem extrair padrões complexos dos atributos.
- *Predições rápidas.* Após as definições do peso da MLP, predizer o rótulo de uma nova amostra costuma ser uma operação rápida.

Porém, a MLP também apresenta algumas desvantagens, como:

- *Quantidade de dados.* A MLP costuma necessitar de uma grande quantidade de dados para conseguir modelar o problema com precisão, o que pode ser uma limitação em certas aplicações.
- *Treinamento custoso.* Por ser um processo iterativo que envolve diversas operações, o treinamento de uma MLP geralmente é demorado.
- *Complexo e de difícil interpretação.* Por causa da sua estrutura complexa de camadas de neurônio, torna-se difícil identificar o impacto que cada atributo causa na saída do modelo. Essa complexidade e dificuldade de interpretação também pode desencorajar o seu uso em aplicações na saúde (RUDIN, 2019).

2.1.5 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*) é um algoritmo de ML que pode ser utilizado para tarefas de regressão, onde também é conhecido como SVR (do inglês *Support Vector Regression*) (AWAD et al., 2015). De acordo com Smola e Schölkopf (2004), o SVM tem como objetivo ajustar um hiperplano aos dados, de forma que grande parte das amostras de treino não apresente um erro maior que ϵ ao serem comparadas com o hiperplano. Dessa forma, cria-se um tubo não sensível ao ϵ , onde amostras fora desse tubo afetam a função de custo. Já as amostras na fronteira do tubo são consideradas os vetores de suporte.

Uma característica importante do SVM é a sua capacidade de lidar com dados de

comportamento não-linear. Isso é possível ao utilizar funções *kernel* não-lineares, que permitem mapear os vetores de atributos para um espaço de alta dimensionalidade (GÉRON, 2022). Alguns dos *kernels* que podem ser aplicados no SVM são o polinomial (SVM_Pol), o sigmoide (SVM_Sig), o linear (SVM_Lin) e o RBF (SVM_RBF), onde RBF significa *Radial Basis Function*, ou Função de Base Radial.

O SVM possui algumas vantagens e desvantagens. Por exemplo, algumas de suas vantagens são:

- *Não-linearidade*. O SVM consegue se ajustar a conjuntos de dados com comportamento não-linear com a ajuda das funções *kernel*.
- *Robusto contra ruídos*. Outro ponto positivo do SVM é que ele consegue ser robusto contra ruídos, dependendo dos hiperparâmetros definidos (CHUANG et al., 2002).
- *Predições rápidas*. Depois da etapa de treinamento, o SVM consegue predizer o rótulo de uma nova amostra rapidamente.

Já alguns exemplos de desvantagens encontradas no uso do SVM são:

- *Sensível à escala dos atributos*. Por ser um método que se baseia em distâncias, a escala dos atributos impacta no SVM (GÉRON, 2022). Atributos com escalas maiores tendem a ter uma importância maior para o algoritmo.
- *Treinamento custoso*. Em tarefas com grandes *datasets*, o SVM apresenta uma etapa de treino demorada.
- *Difícil interpretação*. O SVM é um algoritmo complexo e é difícil de se interpretar, principalmente quando *kernels* não-lineares são utilizados. Essa dificuldade pode desencorajar o seu uso em aplicações na saúde (RUDIN, 2019).

2.1.6 Árvore de Decisão

A Árvore de Decisão (DT, do inglês *Decision Tree*) é um método de ML que pode ser utilizado para tarefas de regressão (BREIMAN et al., 2017). Esse algoritmo cria um modelo em forma de árvore. Durante o treinamento, a árvore é construída dividindo os dados em subconjuntos com base nos valores dos atributos de entrada. Esse processo é feito pela seleção do atributo que resulta na maior redução do critério escolhido, como o erro quadrático médio, por exemplo (GÉRON, 2022). As divisões continuam até que um critério de parada seja cumprido, tal como uma profundidade máxima da árvore ou um número mínimo de amostras em um nó-folha. Ao realizar uma predição, cada nó da árvore, com exceção dos nós-folha, representa uma decisão baseada no valor de uma *feature*. Já o nó-folha é responsável por definir o valor predito para a variável alvo, com base nos valores das amostras na folha (BREIMAN et al., 2017).

Há vantagens e desvantagens em se utilizar a DT. Algumas das vantagens são:

- *Lida bem com dados diversos.* Por ser um algoritmo que se baseia em regras de “se-então”, a DT consegue lidar com dados de diferentes tipos (numérico e categórico) e de diferentes escalas, sem necessitar de um pré-processamento.
- *Simples e interpretável.* Após a construção da árvore, ela pode ser visualizada. Dessa forma, torna-se mais fácil interpretar como ela está atribuindo o valor a uma nova amostra (GÉRON, 2022).
- *Predições rápidas.* Com a árvore construída, novas amostras precisam apenas passar pelas decisões dos nós até chegar em um nó-folha. Essa baixa quantidade de operações torna a DT um dos algoritmos mais rápidos na predição.

Já algumas das desvantagens são:

- *Overfitting.* Esse algoritmo pode gerar uma árvore muito complexa que não generaliza bem os dados, principalmente se não for podada (HASTIE et al., 2009).
- *Instabilidade.* Pequenas variações no conjunto de treino podem gerar árvores completamente diferentes (GÉRON, 2022).

2.1.7 Floresta Aleatória

A Floresta Aleatória (RF, do inglês *Random Forest*) é um método *ensemble* de ML que pode ser utilizado para tarefas de regressão (BREIMAN, 2001). A RF é composta por um conjunto de árvores de decisão, visando diminuir os problemas encontrados ao utilizar apenas uma árvore. De acordo com Hastie et al. (2009), na etapa de treinamento da RF, várias árvores de decisão são construídas utilizando subconjuntos aleatoriamente selecionados do conjunto de treino. Além disso, na divisão de cada nó das árvores, também utiliza-se apenas um subconjunto aleatório dos atributos. A combinação dessas árvores diversas geralmente resulta em uma menor variância no modelo final. Após todas as árvores terem sido construídas, a predição para uma nova amostra é dada pela média da predição de todas as árvores da RF.

Alguns dos pontos positivos apresentados pela RF são:

- *Menor overfitting.* Por combinar as predições de diversas árvores, a RF costuma resolver alguns problemas encontrados ao usar somente a DT. Especificamente, tem um menor *overfitting*, é menos instável e tem uma melhor generalização dos dados (BREIMAN, 2001).
- *Lida bem com dados diversos.* Por ser um modelo construído por árvores de decisão, a RF acaba apresentando essa mesma capacidade de lidar com dados de diferentes tipos e escalas, sem necessitar de um pré-processamento.
- *Não-linearidade.* A DT já é um modelo não-linear, então consequentemente a RF também consegue modelar funções não-lineares.

Porém, a RF também apresenta pontos negativos, tais como:

- *Treinamento custoso.* Por precisar realizar a construção de diversas árvores, o treinamento da RF costuma ser uma etapa demorada, comparando-se aos outros modelos (GÉRON, 2022).
- *Maior armazenamento.* Todas as árvores construídas para a RF precisam ser guardadas. Logo, a RF necessita de um maior armazenamento, que é proporcional ao número de árvores utilizadas.
- *Predição mais lenta.* Uma única DT tem uma predição rápida, mas como a RF precisa antes realizar a predição de diversas árvores, o tempo total gasto na predição de uma nova amostra costuma ser bem maior que nos outros modelos e também é proporcional ao número de árvores utilizadas.

2.1.8 eXtreme Gradient Boosting

O XGBoost (*eXtreme Gradient Boosting*) também é um método *ensemble* de ML que pode ser utilizado para tarefas de regressão (CHEN; GUESTIN, 2016). Esse algoritmo é uma adaptação do método *Gradient Boosting* original. Assim como a RF, o XGBoost utiliza um conjunto de árvores para modelar uma função. Porém, em vez de treinar as árvores paralelamente, o XGBoost é estruturado de forma sequencial, onde a saída de uma árvore é utilizada na construção da árvore seguinte. As árvores são modeladas de forma a se ajustar aos erros residuais cometidos pelo preditor anterior (GÉRON, 2022). O valor final predito pelo XGBoost é dado pela soma das predições de cada árvore sua.

O XGBoost demonstra algumas vantagens e desvantagens. Algumas de suas vantagens são:

- *Menor overfitting.* Além de combinar as predições de várias árvores, o XGBoost também aplica métodos que reduzem o *overfitting*, como a subamostragem de colunas e a redução da contribuição de cada árvore (CHEN; GUESTIN, 2016).
- *Não-linearidade.* Assim como a RF, o XGBoost é composto por árvores e também consegue modelar funções não-lineares.
- *Otimização para grandes datasets.* O algoritmo do XGBoost lida melhor com grandes conjuntos de dados, comparado ao *Gradient Boosting* simples (CHEN; GUESTIN, 2016).

Algumas das desvantagens do XGBoost são semelhantes às encontradas na RF. Por exemplo:

- *Treinamento custoso.* Apesar de ser um algoritmo mais otimizado e geralmente mais rápido que a RF, também necessita treinar diversas árvores.
- *Maior armazenamento.* Assim como a RF, o XGBoost necessita de um maior armazenamento, devido ao número de árvores utilizadas.

- *Predição mais lenta.* Como uma nova amostra deve passar por diversas árvores, a predição de seu valor também é mais lenta que a maioria dos outros algoritmos.

2.2 Interpretabilidade com *SHapley Additive exPlanations*

A interpretabilidade, uma demanda cada vez mais comum em Inteligência Artificial (IA), pode ser definida como o grau em que um humano pode compreender a causa de uma decisão (MOLNAR, 2022). No contexto de ML, a interpretabilidade permite explicar as predições feitas por um modelo. Essa característica possibilita quantificar a importância de cada atributo para uma predição.

Para o pesquisador, a interpretação pode gerar ideias de melhorias para o modelo, auxiliar na identificação de vieses, destacar perturbações que causem uma grande mudança na predição e apoiar o entendimento científico do processo que está sendo modelado (LUNDBERG; LEE, 2017; ARRIETA et al., 2020). Além disso, a interpretabilidade ajuda a derrubar uma das maiores barreiras para a implementação de sistemas inteligentes aplicados à saúde: a falta de transparência do modelo. Profissionais da saúde, que são os responsáveis por analisar a informação do paciente e recomendar tratamentos ou ações preventivas adequadas, sentem a necessidade de entender o comportamento do modelo para ter confiança nele e utilizá-lo (MARKUS; KORS; RIJNBEEK, 2021).

Portanto, esta pesquisa realiza a etapa de interpretabilidade, em virtude dos benefícios agregados por ela. Caso o modelo a ser interpretado seja um dos métodos lineares, a análise da importância de cada atributo é adquirida diretamente dos coeficientes da reta. Já para os modelos não-lineares, a interpretação é realizada através do método *SHapley Additive exPlanations* (SHAP) (LUNDBERG; LEE, 2017).

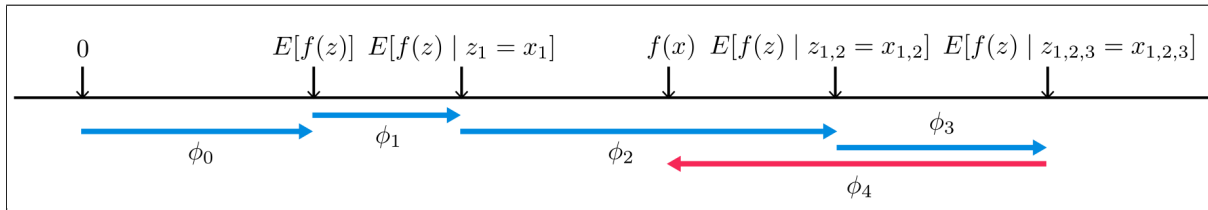
O SHAP é uma ferramenta que permite medir as contribuições de cada atributo para a predição do modelo. Essa técnica tem como base o valor de Shapley da teoria de jogos. Nessa teoria, em um jogo cooperativo com recompensa, as contribuições dos jogadores podem ser calculadas ao analisar a recompensa em todas possíveis combinações de jogadores. O SHAP realiza um processo semelhante, onde, analogamente, os jogadores são os atributos, o jogo é a função modelada e a recompensa é a predição (MOLNAR, 2022).

O SHAP propõe duas abordagens principais para estimar o valor de Shapley: KernelSHAP e TreeSHAP. O KernelSHAP é uma abordagem modelo-agnóstica, ou seja, pode ser usada para qualquer modelo. Essa abordagem utiliza uma associação de valor de Shapley e de LIME (*Local Interpretable Model-Agnostic Explanations*) linear (RIBEIRO; SINGH; GUESTRIN, 2016) e, em combinações que não têm todos atributos, simula esses faltantes substituindo-os por valores de outras amostras aleatoriamente. Já o TreeSHAP é uma abordagem específica para modelos baseados em árvores (LUNDBERG; ERION;

LEE, 2018). Esse método tem a vantagem de ser mais rápido que o KernelSHAP por utilizar a própria estrutura das árvores para simular atributos faltantes.

Para quantificar os efeitos de cada atributo em uma predição (também chamados de valores de SHAP), primeiro define-se um valor base, que representa a saída esperada sem o conhecimento de qualquer atributo. Esse valor é dado pela média de todas as predições realizadas pelo modelo. Depois, adiciona-se um atributo por vez e verifica-se o novo valor esperado, até que todos tenham sido adicionados. A contribuição de todos atributos juntos deve levar a saída esperada do valor base para o valor que realmente foi predito. Esse processo é representado na Figura 1. A figura apresenta apenas uma das possíveis ordens dos atributos (x_1 , x_2 , x_3 e depois x_4). Para obter o valor SHAP de um atributo x_i , é necessário calcular a média das contribuições ϕ_i em ordenações variadas.

Figura 1 – Exemplo de contribuições e seus impactos na predição esperada do modelo. Elas explicam como o modelo sai do valor base $E[f(z)]$ para a predição $f(x)$. A contribuição de um atributo x_i é representada por ϕ_i , com exceção de ϕ_0 que simboliza o valor inicial sem a informação dos atributos. Além disso, $E[f(z)|z_i = x_i]$ representa o valor esperado para a saída do modelo ao adicionar o atributo x_i .



Fonte: Lundberg e Lee (2017)

Finalmente, a ferramenta SHAP foi escolhida por três motivos. Primeiro, atualmente ela é a ferramenta de interpretabilidade mais popular (HOLZINGER et al., 2022), com mais de 18 mil estrelas na plataforma *GitHub*. Outro motivo é que, além de poder ser utilizado para explicar predições individuais, o SHAP também consegue obter a importância dos atributos globalmente. Para calcular essa importância, basta calcular a média dos valores de SHAP para todas as amostras. Por último, o SHAP é baseado em uma sólida teoria e mantém propriedades como: aditividade (a soma das contribuições é igual a diferença entre o valor base e a saída do modelo) e consistência (se o modelo muda de forma que a contribuição de um atributo específico aumenta ou permanece a mesma, o valor SHAP daquele atributo não deve diminuir) (LUNDBERG; LEE, 2017).

3 Conjunto de Dados

Este capítulo é dedicado à descrição do conjunto de dados utilizado nesta pesquisa. O *dataset* é formado por dados de 905 pessoas, contendo informações relacionadas à identificação, a medidas antropométricas, a respostas de anamnese e ao valor de DMO do corpo inteiro. Inicialmente, o capítulo especifica o processo de coleta dos dados. Em seguida, apresentam-se os atributos contidos no conjunto de dados e as transformações aplicadas a eles. Por fim, realizam-se análises estatísticas do *dataset* final.

3.1 Coleta

O projeto de pesquisa foi analisado e aprovado pelo Comitê de Ética em Pesquisa (CEP) do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), com número de protocolo 3593367. Todos os participantes da pesquisa concordaram com a coleta de dados e aceitaram que esses dados poderiam ser usados para fins de pesquisa.

Todos os dados foram coletados por membros do LabDAFEE (Laboratório de Desenvolvimento de Alimentos para Fins Especiais e Educacionais), do Instituto de Nutrição Josué de Castro (INJC) da Universidade Federal do Rio de Janeiro (UFRJ). A equipe de coleta é formada por nutricionistas e educadores físicos, entre alunos de mestrado, doutorado e pós-doutorado.

A coleta de dados antropométricos foi feita por um profissional com certificação nível 2 da ISAK (*International Society for the Advancement of Kinanthropometry*, traduzida livremente como Sociedade Internacional para o Avanço da Cineantropometria). A coleta de circunferências foi realizada de acordo com o protocolo definido pela ISAK. Já a coleta de DMO do corpo inteiro foi obtida utilizando-se um equipamento de DXA, modelo *Prodigy Advance* da marca *GE Healthcare*.

O conjunto de dados foi construído com alguns critérios de seleção. Só foram incluídos no *dataset* aqueles participantes que:

- possuíam no mínimo 18 anos e no máximo 65 anos;
- não possuíam alguma doença debilitante;
- tinham condições de chegar ao local de coleta de dados;
- e não haviam sofrido amputação.

3.2 Atributos

Esta seção apresenta e categoriza os atributos presentes no conjunto de dados. Inicialmente, examinam-se os atributos originais obtidos na coleta e discute-se sobre seus problemas. Após essa análise, apresentam-se as transformações aplicadas e os atributos finais resultantes.

3.2.1 Atributos Originais

O *dataset* inicialmente obtido pelo grupo de coleta contém a variável-alvo (DMO, medida em g/cm^2) e os atributos listados na Tabela 3. Como apresentado na tabela, essas informações obtidas dos participantes foram divididas em três categorias principais: Informações de identificação; Medidas antropométricas; e Informações obtidas na anamnese. Além disso, foi feita uma subdivisão em subcategorias para melhor agrupar o conjunto de atributos. Por último, apresentam-se na tabela os conjuntos de respostas ou valores encontrados no conjunto de dados.

Tabela 3 – Atributos originais do conjunto de dados e suas categorias, subcategorias e respostas encontradas.

Categoria	Subcategoria	Atributo	Respostas obtidas
Identificação	Sexo	Sexo designado no nascimento	“Feminino” ou “Masculino”
	Idade	Idade (em anos)	Número inteiro de 18 a 65
Medidas Antropométricas	Medidas Básicas	Altura (cm)	Número real positivo
		Peso (kg)	Número real positivo
	Medidas de Circunferência	Circunferência do braço (cm)	Número real positivo
		Circunferência do antebraço (cm)	Número real positivo
		Circunferência da cintura (cm)	Número real positivo
		Circunferência do quadril (cm)	Número real positivo
		Circunferência da coxa (cm)	Número real positivo
		Circunferência da panturrilha (cm)	Número real positivo
Anamnese	Hábitos de vida	Faz atividade física?	“Sim” ou “Não”
		Frequência da atividade (dias/semana)	Número inteiro positivo e o zero
		Duração total dos treinos (horas/semana)	Número real positivo e o zero
		Padrão Alimentar	“Onívoro”, “Pescovegetariano”, “Ovolactovegetariano”, “Lactovegetariano”, “Ovovegetariano” ou “Vegano”
	Patologias	Possui diagnóstico para alguma doença ou algum tipo de prótese?	“Sim” ou “Não”
		Qual doença ou prótese?	Resposta aberta (exemplo: “Hipertensão, hipotireoidismo”)
	Ginecologia	Ciclo menstrual é regular?	Resposta aberta (exemplo: “Pós menopausa”)
		Usa anticoncepcional?	“Sim” ou “Não”
		Quantas gestações de mais de 3 meses?	Número inteiro positivo e o zero

Fonte: elaborada pelo autor.

Ao analisar a coluna de “Respostas encontradas” da Tabela 3, percebe-se que muitos atributos possuem respostas no formato de *string*. Esse formato de dado não é aceito como entrada por grande parte dos algoritmos de ML. Por isso, para contornar esse problema, é necessário codificar essas informações para valores numéricos.

Além disso, as informações que permitem respostas abertas são ainda mais desafiadoras. Por não ter um conjunto de respostas definido e por permitir um grande número de respostas diferentes, esse tipo de atributo dificulta a aplicação de técnicas de codificação, como a *one-hot-encoding*. Outro desafio é a escrita das respostas. Erros de digitação, uso de siglas e equívocos na ortografia tornam o agrupamento das respostas complexo.

3.2.2 Atributos Finais

Esta subseção apresenta as transformações aplicadas ao conjunto original de dados para contornar os problemas mencionados na subseção 3.2.1. Também são mostrados os novos atributos gerados a partir dos atributos originais. Por fim, apresenta-se a tabela com os atributos finais que são usados no restante da pesquisa.

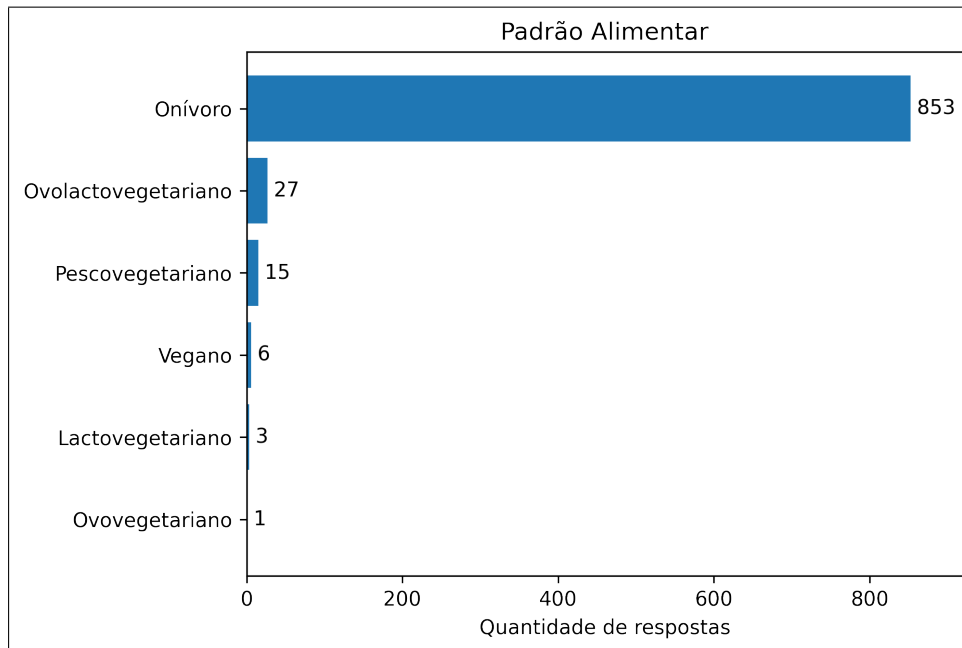
Para os atributos que permitem apenas respostas binárias (“Sim ou Não” e “Feminino ou Masculino”), seus valores foram convertidos em 1 ou 0. Especificamente, as respostas “Sim” e “Feminino” foram convertidas para 1, enquanto que as respostas “Não” e “Masculino” foram convertidas para 0. Dessa forma, o valor deixa de ser uma *string* e torna-se um valor numérico que é aceito por todos algoritmos de ML usados.

No atributo Padrão Alimentar, há um conjunto fixo de seis valores possíveis. Uma opção para essa *feature* poderia ser aplicar a técnica de *one-hot-encoding*, transformando esse atributo em seis colunas binárias, cada uma representando a presença ou não de um padrão alimentar. Entretanto, ao verificar no *dataset* a quantidade de respostas de cada tipo, percebe-se que a grande maioria das respostas são do tipo “Onívoro”, com poucos exemplos dos tipos “Pescovegetariano”, “Ovolactovegetariano”, “Lactovegetariano”, “Ovovegetariano” e “Vegano”, como apresentado na Figura 2. Utilizar *one-hot-encoding* nessa situação geraria cinco atributos esparsos, que dificultam a etapa de aprendizado dos algoritmos. Por isso, optou-se por agrupar as respostas que são variações de “vegetariano” (todas as respostas menos “Onívoro”) e transformar esse atributo na pergunta “É vegetariano?”. Assim, essa coluna foi codificada como 0 para a resposta “Onívoro” e 1 para as variações de “vegetariano” (todas as outras respostas).

Para a pergunta “Ciclo menstrual é regular?”, foi feito um mapeamento das respostas. Primeiramente, os caracteres de todas as respostas foram alterados para caixa baixa, facilitando o mapeamento. Então, todas as respostas que consistem da *string* “sim” foram consideradas como ciclo regular. Qualquer outra resposta foi considerada como ciclo irregular (por exemplo: “não”, “menopausa”, “não menstrua” e “histerectomia”). Por fim, essa informação foi transformada no atributo binário “tem_ciclo_irregular”, onde 1 indica que a pessoa tem ciclo irregular e 0 indica que tem ciclo regular ou que não houve resposta (o que se aplica a todos pacientes do sexo masculino).

Em relação à pergunta “Faz atividade física?”, optou-se por não incluir esse atributo

Figura 2 – Quantidade de respostas para o atributo Padrão Alimentar.



Fonte: elaborada pelo autor.

por essa informação já estar contida nos atributos “Frequência da atividade” e “Duração total dos treinos”.

Por último, os dois atributos de Patologias foram transformados na pergunta “Tem alguma doença de interesse?”. Inicialmente, as respostas foram analisadas e a partir delas foi criada uma lista de “doenças de interesse”: Hipertireoidismo, Hipotireoidismo, Tireoidite de Hashimoto, Linfoma, Artrite Reumatoide, Lúpus, Doença Celíaca, Síndrome do Intestino Irritável, Intolerância à Lactose, Asma, Diabetes e Câncer. Tais doenças foram escolhidas por seus portadores terem maiores chances de possuir valores de DMO alterados (Healthdirect, 2020; National Institutes of Health, 2019). Algumas dessas patologias podem ser fatores de risco para baixo DMO devido aos medicamentos comumente usados por seus portadores. Por exemplo, o uso prolongado de uma classe de esteroides (frequentemente utilizada para tratar a asma) pode aumentar a perda de massa óssea (National Institutes of Health, 2019). Outras doenças afetam diretamente o DMO, como o hipertireoidismo, que afeta a produção de hormônios, e a doença celíaca, que afeta a absorção de nutrientes como o cálcio (National Institutes of Health, 2018c). Após a definição da lista de doenças de interesse, foi feita uma busca por essas patologias nas respostas do atributo original “Qual doença ou prótese?”. Para facilitar a busca, todos caracteres foram alterados para caixa baixa e foram removidos os acentos. Aquelas respostas que possuíam pelo menos uma das doenças mencionadas, foram codificadas como 1 e todas as outras como 0. Ressalta-se que ambos casos, pessoas com ou sem “doenças de interesse”, foram mantidos no *dataset*.

Os atributos finais e seus respectivos valores possíveis são apresentados na Tabela 4.

Tabela 4 – Atributos finais do conjunto de dados e suas categorias, subcategorias e possíveis valores.

Categoria	Subcategoria	Atributo	Valores possíveis
Identificação	Sexo	Sexo designado no nascimento	1 ou 0
	Idade	Idade (em anos)	Número inteiro de 18 a 65
Medidas Antropométricas	Medidas Básicas	Altura (cm)	Número real positivo
		Peso (kg)	Número real positivo
	Medidas de Circunferência	Circunferência do braço (cm)	Número real positivo
		Circunferência do antebraço (cm)	Número real positivo
		Circunferência da cintura (cm)	Número real positivo
		Circunferência do quadril (cm)	Número real positivo
		Circunferência da coxa (cm)	Número real positivo
		Circunferência da panturrilha (cm)	Número real positivo
Anamnese	Hábitos de vida	Frequência da atividade (dias/semana)	Número inteiro positivo e o zero
		Duração total dos treinos (horas/semana)	Número real positivo e o zero
		É vegetariano?	1 ou 0
	Patologias	Tem alguma doença de interesse?	1 ou 0
		Tem ciclo irregular?	1 ou 0
	Ginecologia	Usa anticoncepcional?	1 ou 0
		Quantas gestações de mais de 3 meses?	Número inteiro positivo e o zero

Fonte: elaborada pelo autor.

Como mostrado pela tabela, confirma-se que não existem mais valores de atributos em formato de *string* no *dataset* final.

3.3 Análise do Conjunto de Dados

Esta seção contém análises estatísticas do conjunto de dados final. Há no total 905 amostras no *dataset*, sendo 405 delas do sexo masculino e 500 do sexo feminino. Nesta seção, inicialmente são realizadas análises sobre as variáveis contínuas para todas amostras do *dataset* e também para os conjuntos de pacientes do sexo feminino e masculino, separadamente. Em seguida, apresenta-se a distribuição das variáveis categóricas, também analisando os sexos separadamente. Logo então, investiga-se com maior destaque a distribuição de frequência da DMO. Por fim, exibe-se a matriz de correlação das variáveis e realizam-se discussões sobre os valores encontrados.

A Tabela 5 apresenta uma análise de todos atributos não-binários, incluindo a variável-alvo. Apresentam-se a média, o desvio padrão (DP) e o intervalo de cada variável. Além disso, essa mesma análise é feita separadamente para a parcela de pacientes masculinos e a parcela de pacientes femininos. A tabela confirma que há apenas participantes adultos (de 18 a 65 anos) e indica que a idade média é aproximadamente 35 anos. Além disso, os intervalos das medidas antropométricas revelam que há uma grande variedade de tamanhos de corpos no conjunto de dados. Com exceção da circunferência do quadril, todas as médias das medidas antropométricas masculinas são maiores que as médias femininas. Ao analisar

a prática de atividade física, percebe-se que o conjunto contém pessoas sedentárias e outras que praticam exercício com regularidade, sendo os pacientes masculinos em média mais ativos que os femininos. A tabela também mostra que a maioria das participantes teve nenhum ou poucos filhos. Por último, percebe-se que há grandes variações de DMO no *dataset*, com valores de 0,789 g/cm² a 1,689 g/cm², e que os indivíduos masculinos têm a média de DMO maior que os femininos.

Tabela 5 – Média, desvio padrão (DP) e intervalo dos atributos quantitativos presentes na base de dados. As informações são apresentadas para todo o conjunto e para os conjuntos de pacientes masculinos e femininos, separadamente.

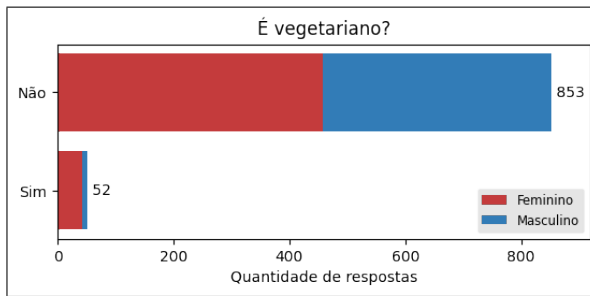
Atributos	Todos (N = 905)		Masculino (N = 405)		Feminino (N = 500)	
	Média (DP)	Intervalo	Média (DP)	Intervalo	Média (DP)	Intervalo
idade (anos)	34.71 (11.92)	[18, 65]	33.70 (11.41)	[18, 65]	35.53 (12.26)	[19, 65]
altura (cm)	168.2 (9.3)	[144.0, 199.0]	175.6 (6.713)	[160.5, 199.0]	162.2 (6.2)	[144.0, 182.0]
peso (kg)	73.3 (16.6)	[30.0, 143.5]	82.3 (14.7)	[53.9, 143.5]	66.1 (14.3)	[30.0, 141.6]
braço (cm)	31.90 (4.61)	[20.65, 48.55]	34.40 (3.70)	[25.30, 48.55]	29.88 (4.26)	[20.65, 46.25]
antebraço (cm)	26.21 (3.29)	[7.85, 50.45]	28.66 (2.58)	[7.85, 50.45]	24.23 (2.32)	[15.95, 34.30]
cintura (cm)	81.27 (12.61)	[37.40, 134.2]	86.57 (11.63)	[37.40, 134.2]	76.98 (11.71)	[55.05, 124.0]
quadril (cm)	100.6 (9.43)	[47.50, 148.5]	99.97 (8.14)	[82.75, 135.1]	101.1 (10.32)	[47.50, 148.5]
coxa (cm)	54.26 (6.38)	[28.00, 98.35]	55.40 (5.72)	[28.00, 79.95]	53.34 (6.73)	[37.90, 98.35]
panturrilha (cm)	37.25 (3.56)	[24.55, 53.70]	38.27 (3.10)	[30.00, 49.50]	36.43 (3.69)	[24.55, 53.70]
freq_atividade (vezes/semana)	3.26 (2.22)	[0, 9]	3.69 (2.17)	[0, 8]	2.902 (2.205)	[0, 9]
tempo_atividade (horas/semana)	4.28 (4.01)	[0, 42]	5.12 (4.34)	[0, 42]	3.6 (3.58)	[0, 28]
número_gestações	0.35 (0.82)	[0, 5]	0 (0)	[0, 0]	0.62 (1.03)	[0, 5]
DMO (g/cm ²)	1.231 (0.136)	[0.789, 1.689]	1.312 (0.123)	[0.944, 1.689]	1.166 (0.107)	[0.789, 1.465]

Fonte: elaborada pelo autor.

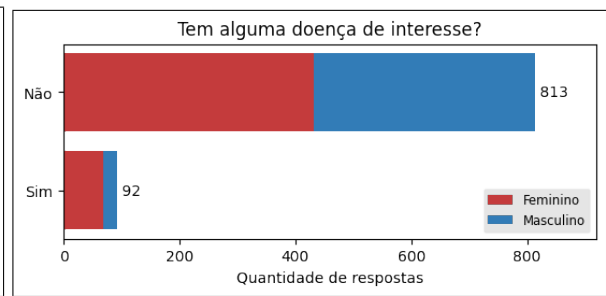
Já para as respostas dos atributos binários, a Figura 3 mostra as suas distribuições, divididas pelo sexo dos pacientes. Percebe-se que em todos casos, a presença de respostas negativas sobressai. Apesar do agrupamento realizado no Padrão Alimentar, o atributo “É vegetariano” continua sendo esparso, com apenas cerca de 6% de respostas positivas. A presença de doenças relevantes nos pacientes avaliados também é incomum: apenas cerca de 10% deles possui uma das doenças listadas. Em relação às pacientes do sexo feminino, verifica-se que cerca de 28% delas usam anticoncepcional e quase 42% possuem ciclo irregular. Em relação aos pacientes do sexo masculino, eles não possuem respostas positivas na subcategoria de Ginecologia e são minoria nas parcelas positivas das outras duas *features*.

Uma ênfase maior é dada na visualização da distribuição da variável alvo. A Figura 4 exibe a distribuição dos valores de DMO em formato de histograma. Visualmente, percebe-se que a distribuição se assemelha à distribuição normal, ocorrendo uma frequência maior de amostras próximo do valor central de DMO e uma pequena quantidade de amostras nas extremidades. Esse comportamento pode prejudicar a capacidade de predição dos modelos em casos de amostras com valores de DMO muito baixos ou altos. Isso pode ocorrer, pois há poucas amostras com esses valores extremos para serem usadas na etapa

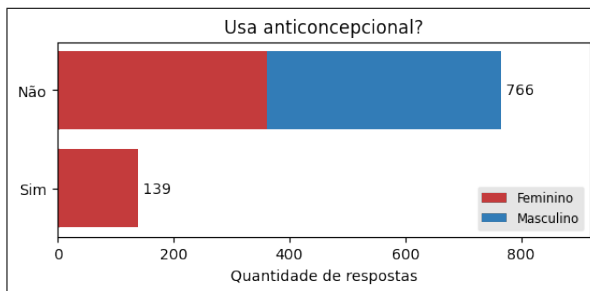
Figura 3 – Distribuição das respostas dos atributos categóricos com base no sexo.



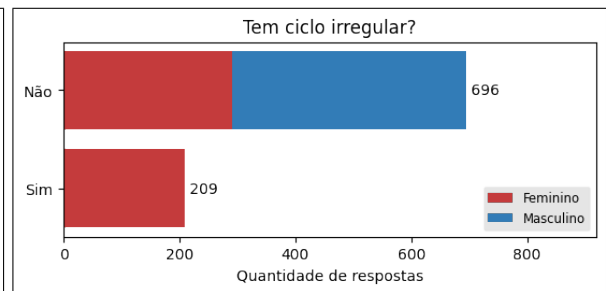
(a) Respostas para “É vegetariano?”



(b) Respostas para “Tem alguma doença de interesse?”



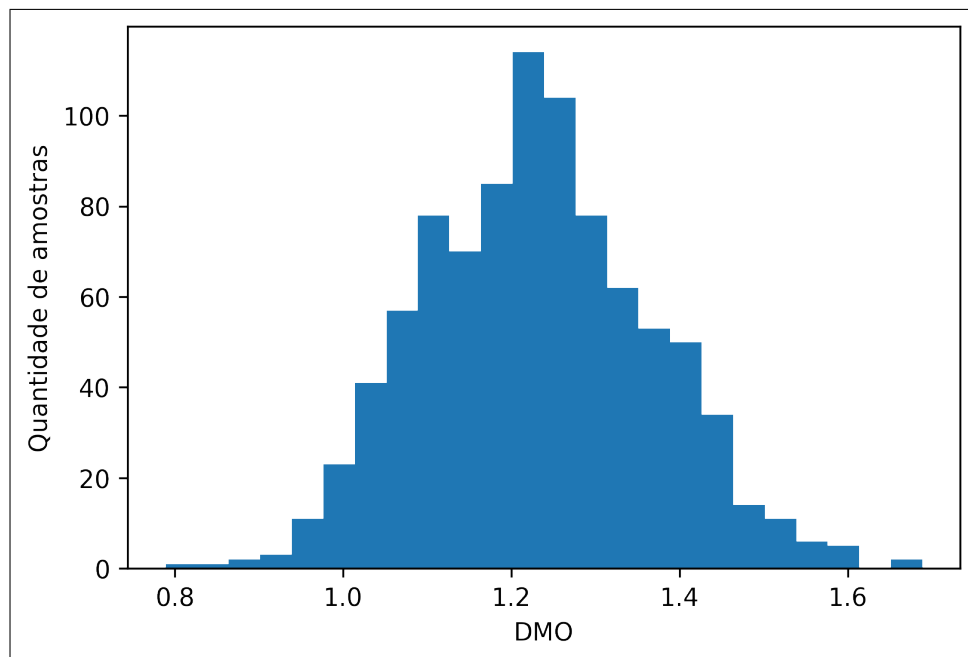
(c) Respostas para “Usa anticoncepcional?”



(d) Respostas para “Tem ciclo irregular?”

Fonte: elaborada pelo autor.

de aprendizado dos algoritmos.

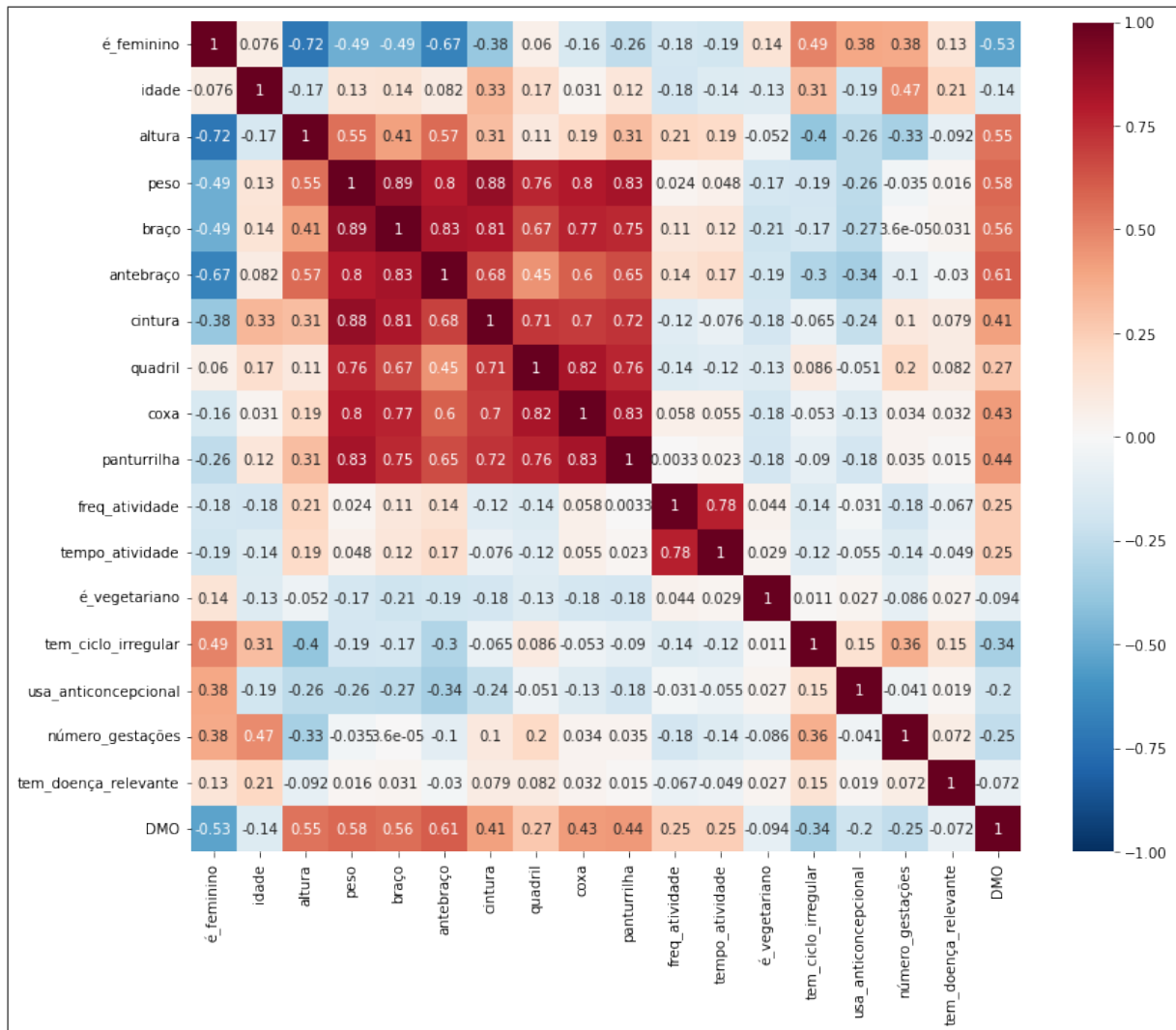
Figura 4 – Histograma dos valores de DMO presentes no *dataset*.

Fonte: elaborada pelo autor.

A Figura 5 apresenta a matriz de correlação de Pearson calculada entre todos

atributos do conjunto de dados, incluindo a DMO. Essa matriz mede a relação linear entre as variáveis. Cada elemento da matriz indica o coeficiente de correlação, que varia de -1 a 1 (MUKAKA, 2012). Se o coeficiente for próximo de 1, há uma forte correlação positiva entre as variáveis, ou seja, à medida que uma variável aumenta, a outra também aumenta. Caso o coeficiente seja perto de -1, há uma forte correlação negativa, ou seja, à medida que uma variável aumenta, a outra diminui. Por fim, se o coeficiente for zero, não há uma correlação linear entre as variáveis. Ressalta-se que a correlação não implica causalidade entre as variáveis, mas apenas identifica uma associação linear entre elas (MUKAKA, 2012). Além disso, enfatiza-se que esse coeficiente analisa a relação linear, ou seja, mesmo que ele seja próximo de 0, podem ser encontrados outros tipos de relação entre as variáveis.

Logo, a partir da Figura 5, percebe-se que há correlações positivas entre todas medidas antropométricas, sendo a maior delas entre peso e circunferência do braço (coeficiente de 0,89) e a menor delas entre a altura e a circunferência de quadril (coeficiente de 0,11). Além disso, verifica-se que há uma correlação negativa entre as medidas antropométricas e a *feature* que identifica amostras do sexo feminino, com exceção da circunferência de quadril. Como já esperado, outro par de atributos com uma alta correlação é a frequência de atividade e o tempo em atividades. Já em relação ao valor de DMO, verifica-se que os atributos de medidas antropométricas e os de atividade física são os únicos com correlação positiva. Dentre esses, a circunferência de antebraço se destaca por ter o maior valor absoluto do coeficiente de correlação: 0,61. Por outro lado, as outras *features* apresentam correlação negativa com DMO, com destaque para o sexo feminino, com coeficiente de -0,53. Portanto, essas correlações podem sugerir que mulheres tendem a ter um menor valor de DMO, pessoas que praticam atividades físicas tendem a ter uma DMO maior, e pessoas com corpos maiores também tendem a ter maiores valores de DMO. Por fim, destaca-se que, na faixa de idades presente no *dataset*, a idade não apresentou uma forte relação linear negativa com a DMO.

Figura 5 – Matriz de correlação de Pearson das variáveis finais do *dataset*.

Fonte: elaborada pelo autor.

4 Metodologia

Este capítulo apresenta a metodologia aplicada para esta pesquisa. Inicialmente, mostra-se uma visão geral das etapas da abordagem proposta. Em seguida, descrevem-se com mais detalhes os componentes do método. Mais especificamente, apresentam-se os experimentos que são conduzidos no trabalho e discute-se sobre as técnicas de pré-processamento aplicadas, o processo de aprendizado dos modelos, os métodos de avaliação e a técnica de interpretação. O repositório com os códigos utilizados nesta metodologia e com instruções de como solicitar acesso ao *dataset* está disponível publicamente na plataforma *GitHub*¹. A requisição de acesso ao *dataset* será analisada pelos pesquisadores que realizaram a coleta, que decidirão sobre a aprovação ou não da solicitação.

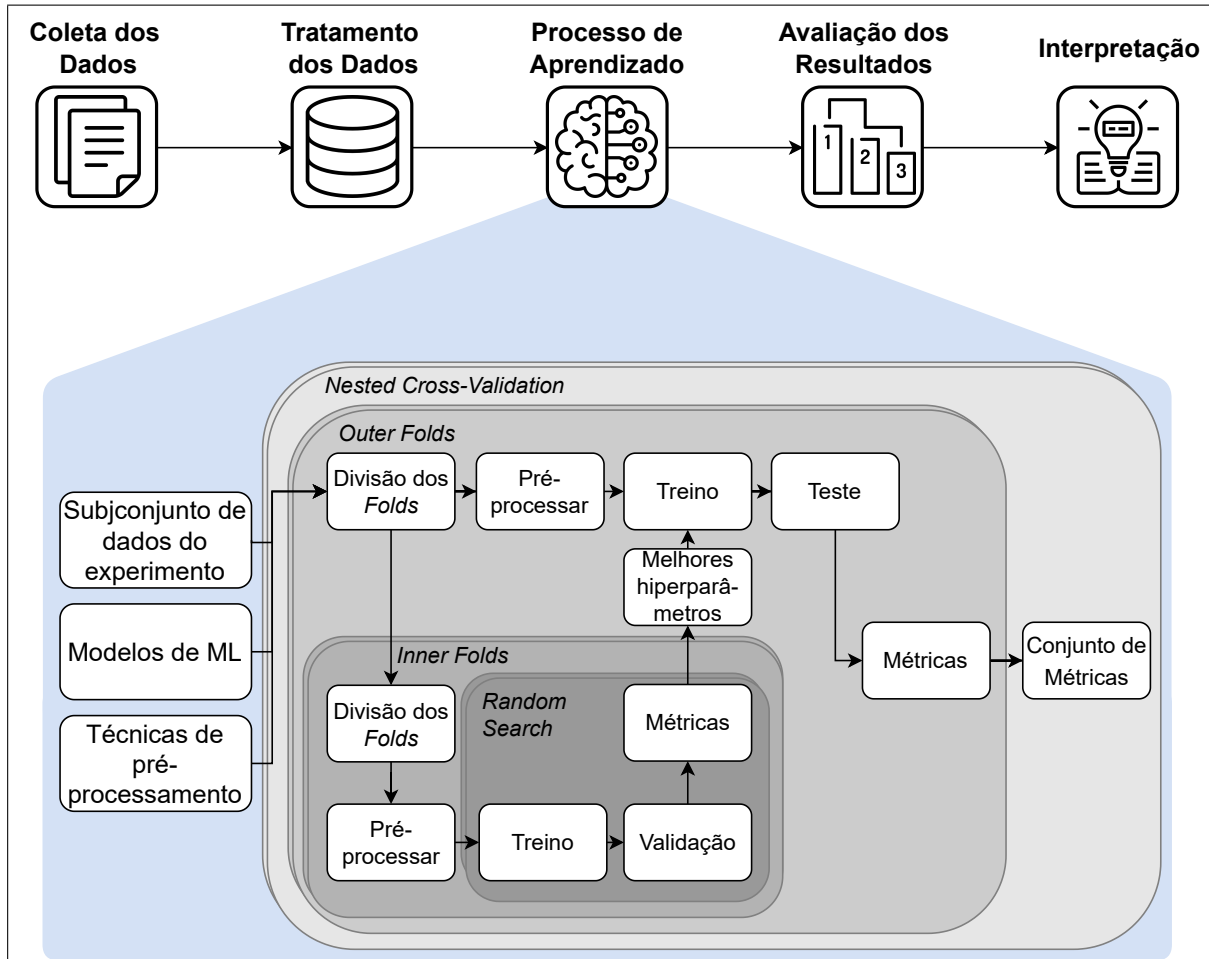
A Figura 6 apresenta um diagrama que ilustra a metodologia desta pesquisa. O fluxo apresentado na região superior do diagrama representa as etapas principais da abordagem. Já o fluxo abaixo, contido na região azul, representa as subetapas da etapa de Processo de Aprendizado. Analisando o fluxo superior, inicialmente são realizadas as etapas de Coleta e Tratamento dos Dados. Essas duas etapas foram detalhadas no Capítulo 3 e foram realizadas apenas uma vez. Em seguida, os dados tratados são utilizados para os experimentos, cujas configurações são especificadas na seção 4.1. Cada experimento inicia pela etapa de Processo de Aprendizado, que é responsável por criar e testar os modelos encarregados de realizar as previsões de DMO. Após a criação e teste dos modelos, os resultados são utilizados na etapa de Avaliação dos Resultados, onde as métricas listadas na subseção 4.4.1 são avaliadas e realizam-se os testes estatísticos descritos na subseção 4.4.2. Assim, essa etapa consegue determinar o modelo mais adequado para cada experimento. Por último, após a definição desse modelo, realiza-se a etapa de Interpretação (especificada na seção 4.5), que é responsável por examinar os impactos dos atributos nas saídas do modelo.

Como mencionado, a Figura 6 também traça as subetapas do Processo de Aprendizado (o fluxo dessa etapa é descrito com maiores detalhes na seção 4.3). Primeiramente, essa etapa é iniciada com os algoritmos de ML listados na seção 2.1, as técnicas de pré-processamento apresentadas na seção 4.2 e o subconjunto de dados do experimento que está sendo investigado. Logo então, esses elementos são utilizados na técnica de validação chamada de *nested cross-validation* (ou validação cruzada aninhada). Essa técnica divide os dados em subconjuntos aninhados, chamados *folds*, e utiliza-se de alguns deles para o treinamento dos algoritmos de ML e outros para testar seus desempenhos. Com isso, é possível escolher os hiperparâmetros mais adequados para cada modelo e, em seguida, gerar métricas para cada modelo. Por fim, esses resultados são passados para a etapa de

¹ Repositório disponível em <https://github.com/gmaiab/Predicao-Densidade-Mineral-Osea>

Avaliação dos Resultados.

Figura 6 – Diagrama da metodologia do trabalho proposto. Na região azul, detalha-se o Processo de Aprendizado.



Fonte: elaborada pelo autor.

4.1 Configuração dos experimentos

Na Tabela 6, são apresentados os experimentos realizados neste trabalho. Designa-se uma sigla para cada um, define-se o grupo de pacientes que faz parte do experimento, e determina-se o subconjunto de atributos utilizados. As siglas são construídas no formato “EX-S”, que significa “Experimento X, com participantes do sexo S”. Cada experimento “X” é realizado em três grupos de amostras. Esses grupos e suas identificações que substituem “S” na sigla são:

- Feminino (F): apenas pessoas com o sexo de nascimento como feminino;
- Masculino (M): apenas pessoas com o sexo de nascimento como masculino;
- Ambos (A): todas pessoas do conjunto de dados.

A divisão dos experimentos nesses três grupos tem como objetivo verificar se

modelos treinados e avaliados em pacientes de um sexo específico apresentam melhores resultados que modelos treinados e avaliados em pacientes de qualquer sexo.

Já a letra “X” na sigla, identifica o tipo do experimento. Cada tipo de experimento utiliza um conjunto específico de atributos do *dataset*. Os atributos selecionados, a identificação dos experimentos e as motivações são listadas abaixo:

- Todos (1): é o experimento base, pois contém todos atributos. É considerado o experimento principal da pesquisa, portanto seus resultados são apresentados com mais detalhes e as análises de interpretabilidade são focadas no melhor modelo desse experimento;
- Sem Medidas de Circunferência (2): ignoram-se os atributos da subcategoria “Medidas de Circunferência”, apresentados na Tabela 4. O objetivo desse experimento é verificar o impacto da retirada dessas medições. Modelos treinados nesse experimento podem ser úteis em casos em que não há alguém capacitado ou instrumentado para realizar essas medidas e em casos que a medição não é possível por outros motivos;
- Sem Anamnese (3): ignoram-se os atributos da categoria “Anamnese”, apresentados na Tabela 4. O objetivo desse experimento é verificar o impacto da retirada dessas perguntas. Modelos treinados nesse experimento podem ser úteis em casos em que não é possível obter esses dados. Por exemplo, o paciente não deseja compartilhar essas informações, ou não sabe/pode responder. Ou até em casos em que um pesquisador possui os atributos de identificação e medidas de um paciente e deseja estimar sua DMO sem ter acesso a dados de anamnese;
- Sem Anamnese e Medidas de Circunferência (4): ignoram-se os atributos da categoria “Anamnese” e da subcategoria “Medidas de Circunferência”, apresentados na Tabela 4. O objetivo desse experimento é verificar o desempenho dos algoritmos utilizando apenas dados básicos (sexo de nascimento, idade, altura e peso). Como são informações fáceis de se obter e estão presentes em grande parte de fichas médicas, um modelo treinado com esses atributos pode ser útil em várias circunstâncias, sem a necessidade de realizar medições de circunferência e as perguntas específicas da categoria de anamnese.

Ressalta-se que, para os experimentos aplicados apenas ao grupo de pacientes do sexo masculino, também são removidos os atributos da subcategoria “Ginecologia” da Tabela 4. Essa remoção é feita pois os dados dessa subcategoria não trazem informações relevantes para o grupo masculino, já que são valores constantes. Por motivos semelhantes, nos experimentos com apenas os pacientes do sexo masculino ou apenas com os do sexo feminino, também é removido o atributo “É feminino?” por não acrescentar informações valiosas ao subconjunto de dados.

Tabela 6 – Siglas dos diferentes experimentos realizados e especificações do conjunto de atributos utilizado e do grupo de pacientes selecionado, em relação ao sexo de nascimento.

Sigla	Sexo de nascimento	Atributos
E1-A	Ambos	Todos
E1-F	Apenas feminino	Todos
E1-M	Apenas masculino	Sem Ginecologia
E2-A	Ambos	Sem Medidas de Circunferência
E2-F	Apenas feminino	Sem Medidas de Circunferência
E2-M	Apenas masculino	Sem Medidas de Circunferência e Ginecologia
E3-A	Ambos	Sem Anamnese
E3-F	Apenas feminino	Sem Anamnese
E3-M	Apenas masculino	Sem Anamnese
E4-A	Ambos	Sem Anamnese e Medidas de Circunferência
E4-F	Apenas feminino	Sem Anamnese e Medidas de Circunferência
E4-M	Apenas masculino	Sem Anamnese e Medidas de Circunferência

Fonte: elaborada pelo autor.

4.2 Pré-processamento

A etapa de pré-processamento refere-se a transformações aplicadas aos dados antes de serem passados para um modelo. A abordagem empregada analisa três alternativas para essa etapa:

- Normalização MinMax;
- Padronização (em inglês, *standardization*) por *z-score*;
- Nenhuma técnica de pré-processamento.

As duas técnicas de pré-processamento listadas acima têm como objetivo minimizar o impacto das escalas de cada atributo. Por exemplo, como apresentado na Tabela 5, a *feature* “altura” varia de 144 a 199, enquanto “número de gestações” varia apenas de 0 a 5. Para alguns algoritmos, em especial aqueles baseados em distância (como o k-NN), essa diferença de escala dá uma maior importância para a variável “altura” comparada ao “número de gestações”. Por isso, o pré-processamento visa trazer os atributos para escalas mais próximas.

A Normalização MinMax transforma os valores de um atributo para uma escala entre um valor mínimo e um valor máximo. Especificamente, o menor valor de um atributo é convertido para 0, o maior é convertido para 1, e os valores intermediários são convertidos proporcionalmente entre 0 e 1. A equação aplicada para essa conversão é apresentada abaixo:

$$x_{imm} = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (4.1)$$

onde x são os valores de um atributo do conjunto de treinamento, x_i é um valor desse atributo que ainda não foi pré-processado e $x_{i_{mm}}$ é o valor correspondente após a normalização MinMax.

Por sua vez, a Padronização centraliza uma *feature* em 0, e escala seus valores ao dividi-los pelo seu desvio-padrão. Dessa forma, ela torna a média e o desvio-padrão do atributo em 0 e 1, respectivamente. A equação aplicada para essa conversão é apresentada abaixo:

$$x_{i_{std}} = \frac{x_i - \text{mean}(x)}{\text{std}(x)}, \quad (4.2)$$

onde x são os valores de um atributo do conjunto de treino, $\text{mean}(x)$ é a média desses valores, $\text{std}(x)$ é o desvio-padrão dos mesmos valores, x_i é um valor desse atributo que ainda não foi pré-processado e $x_{i_{std}}$ é o valor correspondente após a padronização.

Ressalta-se que os pré-processamentos mencionados são aplicados tanto aos valores do conjunto de treino como aos do conjunto de teste. Entretanto, suas equações são definidas com base apenas no conjunto de treino, ou seja, os valores de mínimo, máximo, média e desvio-padrão apresentados nas Equação 4.1 e Equação 4.2 são calculados apenas no conjunto de treino. Essa restrição melhor simula uma situação real (onde os dados de teste não são conhecidos previamente) e evita um vazamento de dados (em inglês, *data leakage*) para o processo de treinamento. Tal vazamento pode gerar resultados superestimados comparados à capacidade real de generalização do modelo.

4.3 Processo de Aprendizado

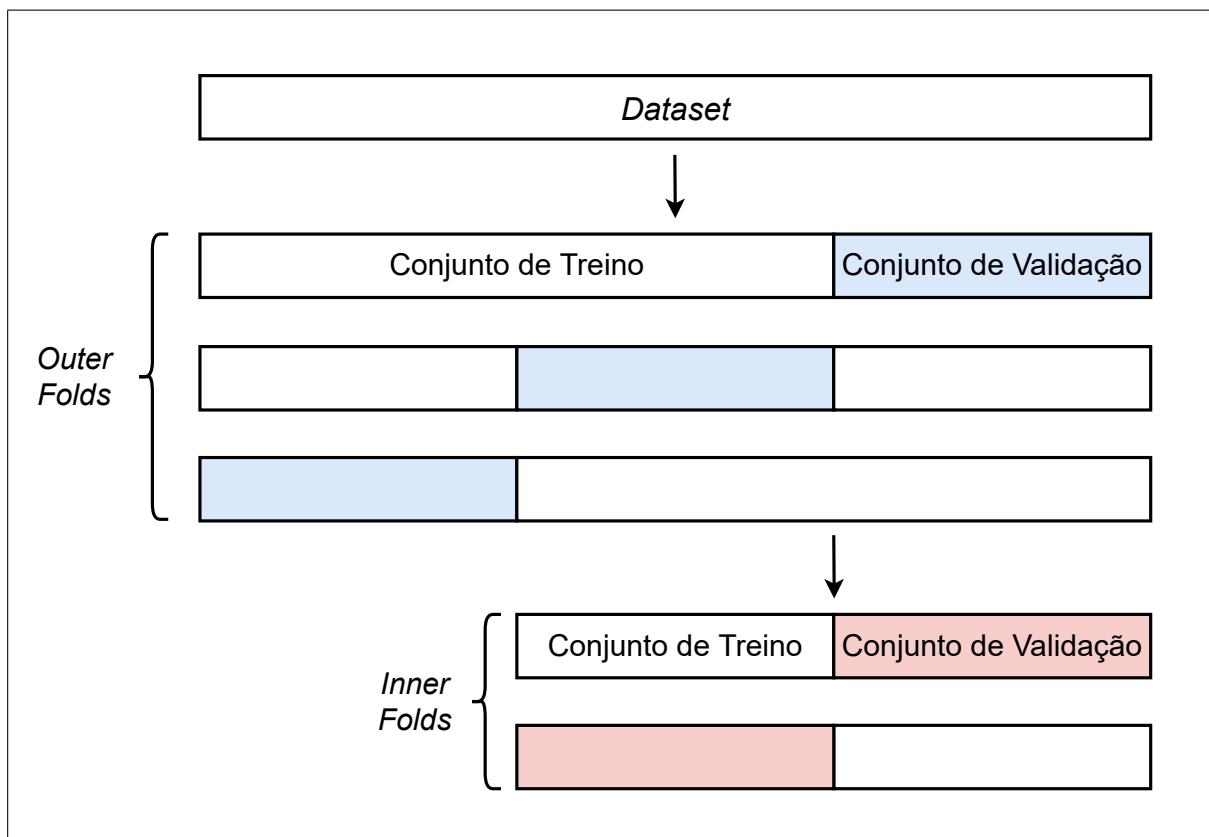
A etapa de Processo de Aprendizado é responsável pela construção dos modelos encarregados de prever DMO e pela validação dos mesmos. Essa etapa pode ser resumida como a validação de cada combinação de algoritmo de ML e técnica de pré-processamento, em cada um dos experimentos listados. Esta seção apresenta a abordagem dessa validação, a estratégia de busca de hiperparâmetros, e o algoritmo que compreende todas as subetapas do Processo de Aprendizado.

A validação cruzada (CV, do inglês *cross-validation*) (STONE, 1974) é uma ferramenta utilizada para avaliar o desempenho de um algoritmo de ML em um *dataset*. Ela permite aferir a capacidade de generalização de um modelo e auxilia na escolha dos hiperparâmetros e algoritmos mais adequados ao problema. Mais especificamente, a CV consiste em dividir o *dataset* em subconjuntos sem interseção (STONE, 1974), utilizando um deles para validação do modelo (etapa de teste, onde o modelo realiza previsões para as amostras e comparam-se esses valores com as anotações das mesmas) e o outro para a construção do modelo (etapa de treinamento, onde o algoritmo “aprende” os seus

parâmetros). Uma configuração popular da CV é a validação cruzada de k -folds (k -fold CV). Nessa configuração, divide-se o *dataset* em k subconjuntos sem interseção, e um dos k subconjuntos é usado para validação, enquanto os outros são utilizados para o treino. Então, realiza-se a etapa treinamento-validação k vezes, sempre utilizando um subconjunto de validação diferente.

Uma forma de validação considerada como uma boa prática é a *nested CV* (CAWLEY; TALBOT, 2010). Essa abordagem aplica k -fold CVs de forma aninhada. Na *nested CV*, há p folds externos (*outer folds*) e q folds internos (*inner folds*). Os *outer folds* são construídos ao dividir o *dataset* em p subconjuntos, gerando assim p conjuntos de treino e de validação diferentes. Já os *inner folds* são gerados dividindo cada um dos p conjuntos de treino dos *outer folds* em q subconjuntos. A Figura 7 representa essas partições para um caso com $p = 3$ e $q = 2$. Por simplificação, a figura ilustra os *inner folds* apenas de um dos *outer folds*, mas essa subdivisão acontece também nos outros *outer folds*.

Figura 7 – Representação visual das divisões de dados realizadas na *nested cross-validation*, com três *outer folds* e dois *inner folds*.



Fonte: elaborada pelo autor.

Essa divisão proposta pela *nested CV* permite que o laço interno de validação seja utilizado para a seleção dos melhores hiperparâmetros, enquanto que o laço externo seja para a seleção do modelo de ML (CAWLEY; TALBOT, 2010). Dessa forma, os conjuntos

de validação dos *outer folds* não são utilizados na otimização dos hiperparâmetros, evitando então a introdução de um viés otimista no desempenho dos modelos (WAINER; CAWLEY, 2021). Portanto, comparando com uma CV tradicional, a *nested CV* fornece um critério mais confiável para a escolha do algoritmo mais adequado. Ainda assim, Wainer e Cawley (2021) compararam o uso de CV tradicional e *nested CV* em diversos *datasets* de classificação binária. Eles concluíram que a CV tradicional geralmente consegue selecionar algoritmos de qualidade semelhante aos selecionados pela *nested CV*, mas o faz de forma mais rápida. Entretanto, como a categoria de problemas investigados por eles é diferente da categoria desta pesquisa, opta-se por utilizar a *nested CV* pelos seus benefícios mencionados, ciente do alto custo que essa abordagem implica. Portanto, o processo de aprendizado deste trabalho faz uso da *nested CV*, utilizando $p = 10$ e $q = 10$, que é uma configuração comumente encontrada na literatura.

Com a definição do método de validação e dos modelos, pré-processamentos e experimentos avaliados, necessita-se definir como os hiperparâmetros são otimizados. Para essa etapa, utiliza-se a estratégia de *Random Search* (busca aleatória), por permitir uma exploração mais diversa de hiperparâmetros e por se mostrar mais eficiente (BERGSTRA; BENGIO, 2012). Nessa abordagem, definem-se os espaços de busca de cada hiperparâmetro e a quantidade de combinações aleatórias de hiperparâmetros que deseja-se formar, com base nos espaços de busca. Nesta pesquisa, a quantidade de combinações a serem avaliadas foi definida como no máximo 30 (dando uma probabilidade de aproximadamente 90% de que pelo menos uma combinação escolhida esteja entre as 10% melhores combinações) e os espaços de busca estão apresentados na Tabela 7. A escolha dos espaços de busca dos hiperparâmetros foi feita com base em outras abordagens encontradas na literatura, em aplicações reais e nas documentações de ferramentas de ML. Pela tabela, percebe-se que apenas a LR não tem hiperparâmetros a serem otimizados.

Reunindo todas as subetapas do Processo de Aprendizado, o Algoritmo 1 apresenta o pseudocódigo para o processo de *nested CV*, pelo qual cada par de modelo e pré-processamento passam em cada experimento. Esse algoritmo também é visualmente representado na região azul da Figura 6. Abaixo estão as descrições de cada parte do algoritmo:

- Linha 1: A função recebe como entrada a sigla do experimento (necessária para definir o conjunto de dados e atributos a serem usados), o nome do modelo e o nome da técnica de pré-processamento.
- Linha 2: Cria o vetor que armazenará os resultados das métricas dos *outer folds*;
- Linhas 3 e 4: Define a quantidade de *outer* e *inner folds*;
- Linha 5: Define a quantidade de conjuntos de hiperparâmetros que serão gerados aleatoriamente;
- Linha 6: Carrega os atributos das amostras para aquele experimento e seus

Tabela 7 – Hiperparâmetros ajustados por *Random Search* e seus respectivos modelos e espaços de busca.

Hiperparâmetro	Espaço de busca	Modelos
alfa ($a + b$)	$[10^{-6}, 10^6]$ (<i>loguniform</i>)	EN
razão L1 ($a/(a + b)$)	$[0, 1]$	EN
k (número de vizinhos)	$[1, 100]$ (apenas inteiros)	k-NN
número de camadas escondidas	$\{1, 2\}$	MLP
neurônios na camada escondida	$\{1, 3, 9, 27\}$	MLP
função de ativação	$\{\text{sigmóide}, \text{tanh}, \text{ReLU}\}$	MLP
C	$[10^{-2}, 10^9]$ (<i>loguniform</i>)	Todos SVM
gamma	$[10^{-6}, 10^2]$ (<i>loguniform</i>)	SVM (RBF, Pol e Sig)
grau	$[2, 7]$ (apenas inteiros)	SVM-Pol
critério	$\{\text{MSE}, \text{MAE}, \text{MSE de Friedman}\}$	DT e RF
mínimo de amostras na folha (%)	$[1, 25]$	DT e RF
profundidade máxima da árvore	$\{1, 2, 4, 6, 8, \text{nenhuma}\}$	DT, RF, e XGBoost
número de árvores	$[10, 500]$ (apenas inteiros)	RF e XGBoost
número máximo de folhas	$[2, 16]$ (apenas inteiros)	XGBoost

Fonte: elaborada pelo autor.

respectivos valores de DMO;

- Linha 7: Define as divisões de treino e teste para os *outer folds*;
- Linhas 8 a 35: Etapas realizadas para cada divisão dos *outer folds*;
- Linha 9: Divide o *dataset* em seus conjuntos de treino e teste e seus respectivos valores de DMO;
- Linha 10: Define as divisões de treino e teste para os *inner folds*, com base no conjunto de treino do *outer fold*;
- Linha 11: Gera as combinações de hiperparâmetros da *random search*, com base no modelo selecionado e na quantidade de combinações desejada;
- Linha 12: Cria o vetor que armazenará cada combinação de hiperparâmetros e as respectivas métricas obtidas nos *inner folds*;
- Linhas 13 a 25: Etapas realizadas para cada divisão dos *inner folds*;
- Linha 14: Divide o conjunto de treino do *outer fold* em seus subconjuntos de treino e teste e seus respectivos valores de DMO;
- Linhas 15 a 17: Cria a etapa de pré-processamento com base no subconjunto de treino e aplica-a nos subconjuntos de treino e teste;
- Linhas 18 a 24: Etapas realizadas para cada conjunto de hiperparâmetros da *random search*;
- Linhas 19 e 20: Cria o modelo de acordo com seus hiperparâmetros e treina-o no subconjunto de treino;
- Linhas 21 a 23: Usa o modelo treinado para prever o DMO das amostras do subconjunto de teste e compara essas previsões com os valores reais dessas amostras, gerando as métricas e salvando-as junto com o conjunto de hiperparâmetros;

- Linha 26: Define o melhor hiperparâmetro desse *outer fold*, com base na melhor média das métricas obtidas nos *inner folds*;
- Linhas 27 a 29: Cria a etapa de pré-processamento com base no conjunto de treino e aplica-a nos conjuntos de treino e teste;
- Linhas 30 e 31: Cria o modelo de acordo com o melhor conjunto de hiperparâmetros e treina-o no conjunto de treino;
- Linhas 32 a 34: Usa o modelo treinado para prever o DMO das amostras do conjunto de teste e compara essas previsões com os valores reais dessas amostras, gerando as métricas e salvando-as no vetor de resultados dos *outer folds*;
- Linha 36: A função retorna como saída o vetor das métricas obtidas em cada *outer fold*, que é então utilizado na etapa de Avaliação dos Resultados.

Por fim, pode-se calcular quantos treinamentos são feitos na metodologia desta pesquisa. Como são avaliados 11 modelos de ML, 3 alternativas de pré-processamento e 12 experimentos, o Algoritmo 1 é aplicado para $11 \times 3 \times 12 = 396$ combinações. Para cada uma dessas combinações, a *nested CV* realiza $p \times q \times \text{combinações_de_hiperparâmetros} = 10 \times 10 \times 30 = 3000$ treinamentos nos laços internos e $p = 10$ treinamentos no laço externo. Portanto, a metodologia realiza até $396 \times (3000 + 10) = 1191960$ treinamentos de modelos. Ressalta-se que algumas combinações de hiperparâmetros podem se repetir, causando treinamentos com configurações iguais.

Algoritmo 1 Pseudocódigo para *Nested Cross-Validation*

```

1: função NESTEDCV(experimento, nome_modelo, nome_preprocesso)
2:   resultados = [ ]
3:   num_outer_folds = 10
4:   num_inner_folds = 10
5:   num_search = 30
6:   X, y = CARREGA_DADOS(experimento)
7:   outer_folds = GERA_FOLDS(X, y, num_outer_folds)
8:   para todo outer_fold em outer_folds faça
9:     X_treino, y_treino, X_teste, y_teste = outer_fold
10:    inner_folds = GERA_FOLDS(X_treino, y_treino, num_inner_folds)
11:    combinações_hiper = GERA_HIPERPARAMETROS(nome_modelo, num_search)
12:    resultados_inner = [ ]
13:    para todo inner_fold em inner_folds faça
14:      X_treino_in, y_treino_in, X_teste_in, y_teste_in = inner_fold
15:      preprocesso = CONSTROI_PREPROCESSO(X_treino_in, nome_preprocesso)
16:      X_treino_in.APLICA_PREPROCESSO(preprocesso)
17:      X_teste_in.APLICA_PREPROCESSO(preprocesso)
18:      para todo hiper em combinações_hiper faça
19:        modelo = CRIA_MODELO(nome_modelo, hiper)
20:        modelo.TREINA(X_treino_in, y_treino_in)
21:        y_pred = modelo.PREDIZ(X_teste_in)
22:        métricas = GERA_MÉTRICAS(y_teste_in, y_pred)
23:        resultados_inner.ADD([hiper, métricas])
24:      fim para
25:    fim para
26:    melhor_hiper = PEGA_MELHOR_HIPER(resultados_inner)
27:    preprocesso = CONSTROI_PREPROCESSO(X_treino, nome_preprocesso)
28:    X_treino.APLICA_PREPROCESSO(preprocesso)
29:    X_teste.APLICA_PREPROCESSO(preprocesso)
30:    modelo = CRIA_MODELO(nome_modelo, melhor_hiper)
31:    modelo.TREINA(X_treino, y_treino)
32:    y_pred = modelo.PREDIZ(X_teste)
33:    métricas = GERA_MÉTRICAS(y_teste, y_pred)
34:    resultados.ADD(métricas)
35:  fim para
36:  devolve resultados
37: fim função

```

4.4 Avaliação dos Resultados

A etapa de Avaliação dos Resultados é encarregada de quantificar o desempenho das abordagens testadas em cada experimento. Além disso, ela é responsável por realizar os testes estatísticos adequados para identificar quais abordagens são superiores. Portanto, primeiramente a subseção 4.4.1 lista e apresenta as métricas utilizadas nesta pesquisa. Em seguida, a subseção 4.4.2 descreve os testes estatísticos aplicados e seus significados.

4.4.1 Métricas

Esta seção apresenta as métricas utilizadas para avaliar o desempenho dos modelos de aprendizagem de máquina utilizados na pesquisa. Elas se dividem em duas categorias: métricas que avaliam a qualidade das predições de DMO e métricas que representam a quantidade de recurso computacional utilizada.

Este trabalho calcula as seguintes métricas para avaliar as predições:

- Erro absoluto médio (MAE, do inglês *Mean absolute error*);
- Raiz do erro quadrático médio (RMSE, do inglês *Root mean square error*);
- Erro percentual absoluto médio (MAPE, do inglês *Mean absolute percentage error*);
- Erro máximo (ME, do inglês *Maximum error*);
- Coeficiente de determinação (R^2).

Para analisar o recurso computacional, este trabalho utiliza as seguintes métricas:

- Tempo de treino (Ttr);
- Tempo de teste (Tte);
- Tamanho do arquivo (TA).

Todas as métricas mencionadas são explicadas nas subseções a seguir. Para facilitar a compreensão das equações, adotam-se as seguintes definições para as variáveis: n representa o tamanho da amostra avaliada, y_i é o valor de DMO verdadeiro para a i -ésima amostra e $\hat{y}_i$ é o valor de DMO predito para a i -ésima amostra. Além disso, a ideia de máximo é representada por $\max$. Por exemplo, o valor de $\max(a, b, c)$ é igual ao maior valor entre a , b e c .

4.4.1.1 MAE

O MAE é calculado como a média dos erros absolutos. Sua equação pode ser descrita como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.3)$$

Essa métrica é comumente utilizada em problemas de regressão por ser intuitiva e fácil de interpretar. O valor do MAE tem a mesma unidade da variável sendo analisada. Ou seja, nesta pesquisa, um MAE de 0,1 indica que o modelo comete, em média, erros de módulo de 0,1 g/cm² em suas predições.

O valor ideal (e o menor valor) para essa métrica é 0, que representa o caso em que todas predições são iguais aos rótulos das amostras, ou seja, não cometeram-se erros.

O MAE não tem limite superior. Portanto, valores altos de MAE são indesejados, já que indicam maiores erros sendo cometidos.

4.4.1.2 RMSE

O RMSE é dado pela raiz quadrada do erro quadrático médio e pode ser definido como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.4)$$

A principal diferença entre o MAE e o RMSE é que o último amplifica erros maiores. Isso ocorre devido à introdução do quadrado do erro em sua fórmula. Portanto, o RMSE é útil quando deseja-se ampliar a penalidade de erros grandes. Além disso, como o RMSE aplica a raiz quadrada após os somatórios, a unidade desse erro também é a mesma unidade da variável analisada. Porém, seu significado não é tão intuitivo quanto o do MAE.

Assim como o MAE, o valor ideal (e o menor valor) para o RMSE é 0. Também não há limite superior no RMSE, e quanto maior o seu valor, pior o desempenho do modelo. Por último, ressalta-se que sempre $RMSE \geq MAE$. O único caso em que as duas métricas são iguais é quando os erros cometidos em todas as amostras têm o mesmo valor absoluto.

4.4.1.3 MAPE

O MAPE computa a média dos erros percentuais absolutos. Sua fórmula é dada por:

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)}, \quad (4.5)$$

onde ϵ é um valor positivo de pequena magnitude. Sua presença é necessária apenas para casos em que o valor real da variável alvo é zero, evitando assim divisões por zero.

O grande diferencial do MAPE é sua sensibilidade a erros relativos. Assim, apenas o valor absoluto do erro não é relevante, mas sim como esse erro se compara ao valor esperado. Além disso, o MAPE é expresso em porcentagem, o que pode facilitar a compreensão do desempenho do modelo para casos em que a escala da variável alvo não é bem conhecida.

Assim como as outras métricas apresentadas, quanto menor o valor do MAPE, melhor o desempenho do modelo. O valor ideal é 0%. Analogamente, quanto maior o MAPE, pior é o desempenho do modelo. Inclusive, esse valor pode ultrapassar 100% quando o valor absoluto do erro é maior que o valor real da amostra.

4.4.1.4 ME

O ME representa o maior erro absoluto encontrado. Ele pode ser descrito como:

$$ME = \max(|y_1 - \hat{y}_1|, \dots, |y_i - \hat{y}_i|, \dots, |y_n - \hat{y}_n|) \quad (4.6)$$

A utilidade do ME é identificar o pior erro cometido pelo modelo. Como as outras métricas apresentadas são médias, elas podem camuflar casos raros em que o modelo comete erros extremos. Já o ME consegue encontrar o pior desses casos.

Assim como o MAE e RMSE, o ME também é expresso pela mesma unidade da variável alvo. Outra semelhança é que quanto menor o ME, melhor o desempenho do modelo. O valor mínimo possível é 0, mas não há limite superior.

4.4.1.5 R^2

O R^2 , conhecido também como coeficiente de determinação, pode ser calculado através da seguinte equação:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.7)$$

onde $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

O R^2 pode ser interpretado como a proporção da variância da variável dependente que é explicada pelas variáveis independentes do modelo (CHICCO; WARRENS; JURMAN, 2021). Nesta pesquisa, a variável dependente é a DMO e as independentes são os atributos das amostras. Dessa forma, o R^2 pode ser usado como um indicador da qualidade do ajuste do modelo.

Através da equação apresentada, percebe-se que o R^2 não possui unidade. Além disso, ele tem o comportamento oposto das outras métricas apresentadas: quanto maior o seu valor, melhor o desempenho do modelo. O valor ideal (e o valor máximo) para o R^2 é 1, porém ele não possui limite inferior. Caso todas as predições de um modelo sejam exatamente a média da variável dependente, o R^2 atinge o valor 0.

4.4.1.6 Tempo de treino

O Ttr é o intervalo que se leva para treinar um modelo no conjunto de treino nos *outer folds* da *nested CV*. Essa métrica representa o quão rápido um modelo consegue aprender seus parâmetros a partir dos dados. O tempo de treino torna-se mais relevante quando há uma necessidade de se retreinar o modelo com frequência, quando há uma grande quantidade de dados, ou caso deseje-se repetir as buscas pelos melhores modelos e hiperparâmetros.

4.4.1.7 Tempo de teste

O Tte é o intervalo que o modelo leva entre receber os atributos de uma amostra e retornar o valor de DMO estimado. Essa métrica torna-se mais significativa quando há uma grande quantidade de chamadas para a função de inferência. Valores altos de Tte podem causar impactos maiores no tempo de resposta quando há um alto volume de requisições para o modelo.

4.4.1.8 Tamanho do arquivo

O TA é o armazenamento ocupado por um modelo treinado. Após a etapa de treino, o modelo pode ser persistido em um arquivo e enviado para produção. Se o ambiente de produção for uma máquina ou sistema embarcado com pouco espaço de armazenamento ou memória RAM, o TA é uma métrica que deve levar maior atenção.

4.4.2 Testes estatísticos

Os testes estatísticos são ferramentas que auxiliam o processo de avaliação e comparação dos resultados. Eles permitem comparar se há diferenças significativas entre o desempenho de duas ou mais abordagens. Para representar o desempenho dos algoritmos, escolheu-se a métrica MAE como variável de interesse, por ela retratar um erro de mais fácil compreensão. Os valores de MAE obtidos em cada partição de teste dos *outer folds* são comparados entre todas combinações de pré-processamento e modelo. Como as partições do *dataset* realizadas pela *nested CV* são as mesmas em um experimento, não há diferenças de dados usados para treino e teste entre os algoritmos, o que torna a comparação mais justa. Para identificar a presença de diferenças significativas entre as abordagens, escolheu-se o teste de Friedman (FRIEDMAN, 1937). Ele foi selecionado por:

- ser um teste não-paramétrico, logo não exige suposições sobre a distribuição da variável;
- conseguir comparar 3 ou mais grupos, logo é possível comparar todas as combinações de pré-processamento e modelo em um experimento;
- conseguir comparar amostras vinculadas, o que ocorre para os valores de MAE, já que os conjuntos de teste são os mesmos para todas abordagens.

Os dados analisados pelo teste de Friedman são estruturados em b blocos, cada um contendo k observações (CONOVER, 1999). Nessa pesquisa, cada bloco é equivalente ao conjunto de teste de cada *outer fold*. Já as observações são representadas por cada combinação de pré-processamento e modelo. Por fim, o valor obtido na i -ésima observação do j -ésimo bloco é dado pelo MAE obtido ao avaliar a i -ésima abordagem nas amostra de teste do j -ésimo *outer fold*. Portanto, cada teste de Friedman analisa $10 \times 33 = 330$ valores de MAE (10 *folds* em 33 abordagens).

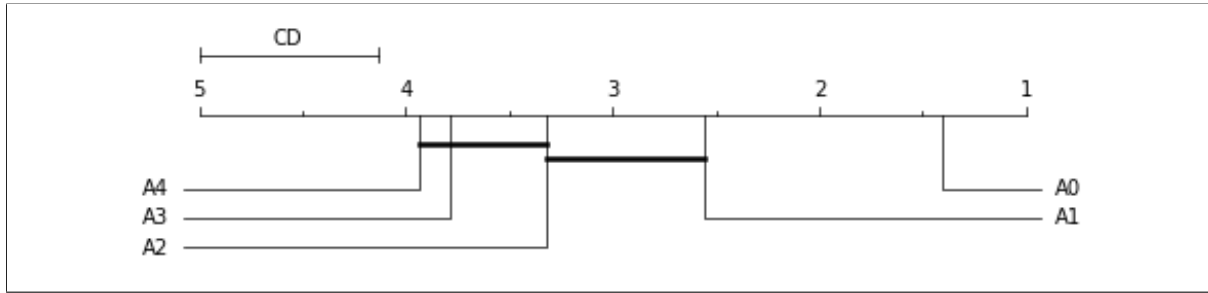
Após a estruturação dos dados, as variáveis de cada bloco são classificadas de acordo com seu valor, onde a menor delas recebe o rank (posto) 1 e a maior delas recebe o rank k (CONOVER, 1999). Esses valores de rank são utilizados para identificar a presença de diferenças significativas nos resultados das abordagens. Mais formalmente, são formuladas as seguintes hipóteses (SHESKIN, 2000):

- Hipótese nula: cada rank é igualmente provável de ser atribuído a cada observação dentro de um bloco. Ou seja, as abordagens têm desempenhos equivalentes e as diferenças em seus resultados podem ser atribuídas a uma chance aleatória.
- Hipótese alternativa: pelo menos uma observação tende a obter ranks diferentes de pelo menos uma outra observação. Ou seja, pelo menos duas abordagens têm desempenhos diferentes entre si.

Para que a hipótese nula seja rejeitada, é necessário que o valor-p do teste de Friedman seja abaixo do nível de significância (α). O valor adotado para α nesse estudo é de 0,05, que é comumente encontrado na literatura.

Após o teste de Friedman, se a hipótese nula houver sido rejeitada, é necessário aplicar um teste *post-hoc* para descobrir quais abordagens tiveram desempenhos significativamente diferentes. O teste *post-hoc* correspondente ao de Friedman é o de Nemenyi (NEMENYI, 1963). No teste de Nemenyi, são feitas múltiplas comparações em pares, com o intuito de identificar aqueles que são estatisticamente equivalentes. Esse teste calcula um valor chamado Diferença Crítica (CD, do inglês *Critical Difference*). Duas abordagens são significativamente diferentes se a diferença entre seus ranks médios for maior ou igual ao CD.

Para melhor visualizar os resultados do teste de Nemenyi, é construído o Diagrama de CD (DEMŠAR, 2006). Um exemplo genérico desse diagrama é apresentado na Figura 8. O eixo horizontal indica o rank médio de cada abordagem, onde os ranks mais baixos (abordagens melhores) estão mais à direita e os ranks mais altos (abordagens piores) mais à esquerda. No topo do diagrama, apresenta-se a magnitude do CD. Além disso, os grupos de abordagem que não são significativamente diferentes são conectados entre si por uma barra horizontal. Por exemplo, na Figura 8, as abordagens A1 e A2 não têm desempenhos significativamente diferentes entre si e o mesmo é válido para as abordagens A2, A3 e A4. Por outro lado, a abordagem A0 é significativamente diferente de todas as outras.

Figura 8 – Exemplo genérico do Diagrama de Diferença Crítica. ($CD = 0,9$)

Fonte: elaborada pelo autor.

4.5 Interpretabilidade

A etapa de interpretabilidade é responsável por mensurar o impacto de cada atributo nas predições. Como apresentado na seção 2.2, as interpretações geram importantes *insights* sobre o comportamento do modelo e o torna mais aceito por especialistas. Portanto, após a etapa de Avaliação dos Resultados definir o modelo mais adequado para o experimento, realiza-se o processo de interpretação. Caso esse seja um modelo linear, o impacto das *features* é dado pelos próprios coeficientes da reta ajustada. Já nos outros casos, utiliza-se o método SHAP descrito na seção 2.2.

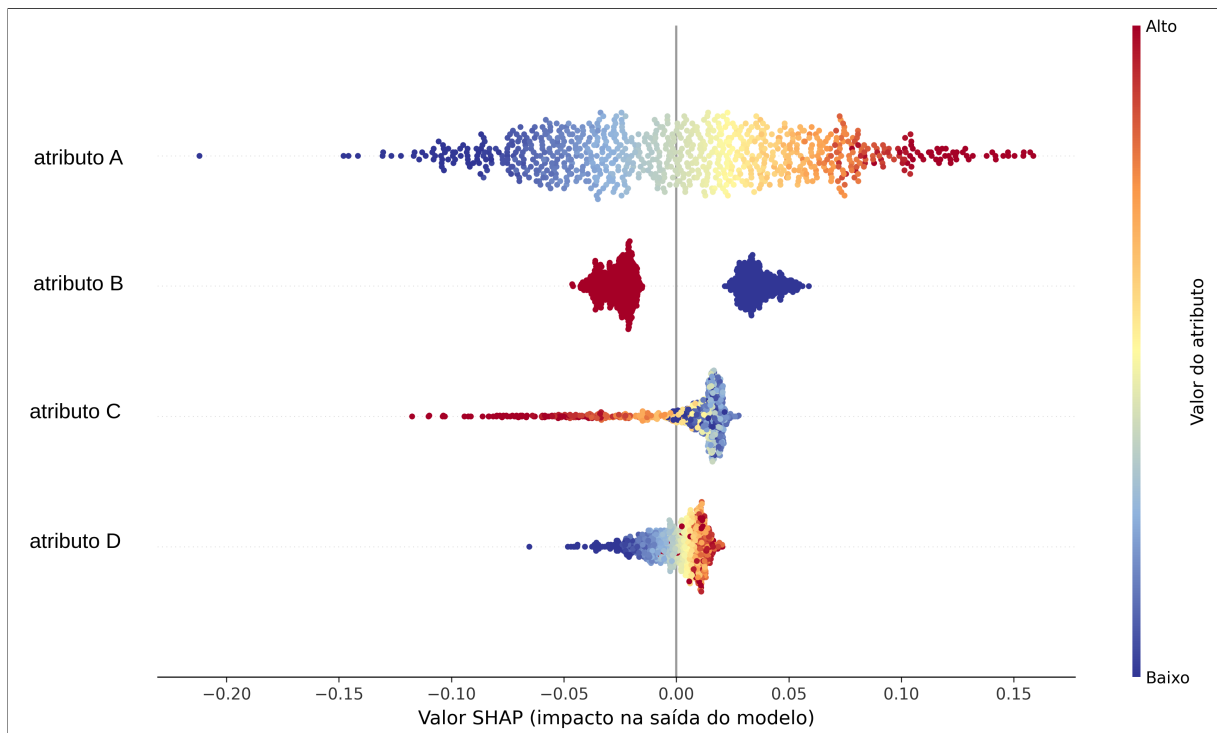
No caso do SHAP ser a técnica de interpretação, seguem-se as instruções indicadas na documentação dessa ferramenta (LUNDBERG, 2020). Após a definição do melhor par de pré-processamento e modelo para um experimento específico, realiza-se uma nova *10-fold CV* para otimizar os hiperparâmetros, com as mesmas configurações de busca apresentadas na seção 4.3. Com o melhor conjunto de hiperparâmetros encontrado, utiliza-se esse conjunto para treinar o modelo no *dataset* completo. Em seguida, o modelo treinado, o pré-processamento e todas as amostras são passadas para a ferramenta do SHAP. Por último, o SHAP calcula o impacto de todos atributos em cada uma das amostras do *dataset*. Esses valores SHAP são então utilizados para construir avaliações gráficas que retratam as influências das *features* nas predições.

Em particular, nesta pesquisa são realizados dois tipos de análise gráfica. A primeira é o gráfico *beeswarm* (enxame de abelha), que permite uma visualização geral do comportamento dos valores SHAP conforme os valores dos atributos. A Figura 9 apresenta um gráfico *beeswarm* para um conjunto de dados genérico com quatro *features*. Os atributos são ordenados verticalmente de acordo com o valor absoluto médio dos impactos causados na saída de cada amostra, alocando a *feature* com maior impacto no topo do gráfico e a com menor impacto na base. Na linha de cada atributo, os pontos representam uma amostra do *dataset*. A posição no eixo *x* desses pontos indica o impacto que essa *feature* tem na predição dessa amostra (LUNDBERG et al., 2020). Quando vários pontos caem na mesma

posição x , eles se amontoam para mostrar densidade. Além disso, os pontos são codificados por cores que indicam o valor do atributo: vermelho, amarelo e azul representam valores altos, intermediários e baixos, respectivamente.

Na Figura 9, os atributos que têm o maior impacto absoluto médio são o A, B, C e D, nessa ordem. Percebe-se que valores altos do atributo A ou do atributo D aumentam o valor predito pelo modelo. Já para os atributos B e C, valores altos causam redução do valor predito. Além disso, pela figura percebe-se que o atributo B é uma *feature* binária. Também verifica-se que valores intermediários do atributo C podem ter impactos parecidos com valores baixos do mesmo atributo. Por fim, ao analisar o atributo A, percebe-se que o valor desse atributo em algumas amostras é responsável por uma redução de cerca de 0,20 na saída do modelo ou até um acréscimo de aproximadamente 0,15 na saída.

Figura 9 – Gráfico *beeswarm* genérico representando o comportamento dos valores SHAP de acordo com os valores de quatro atributos.



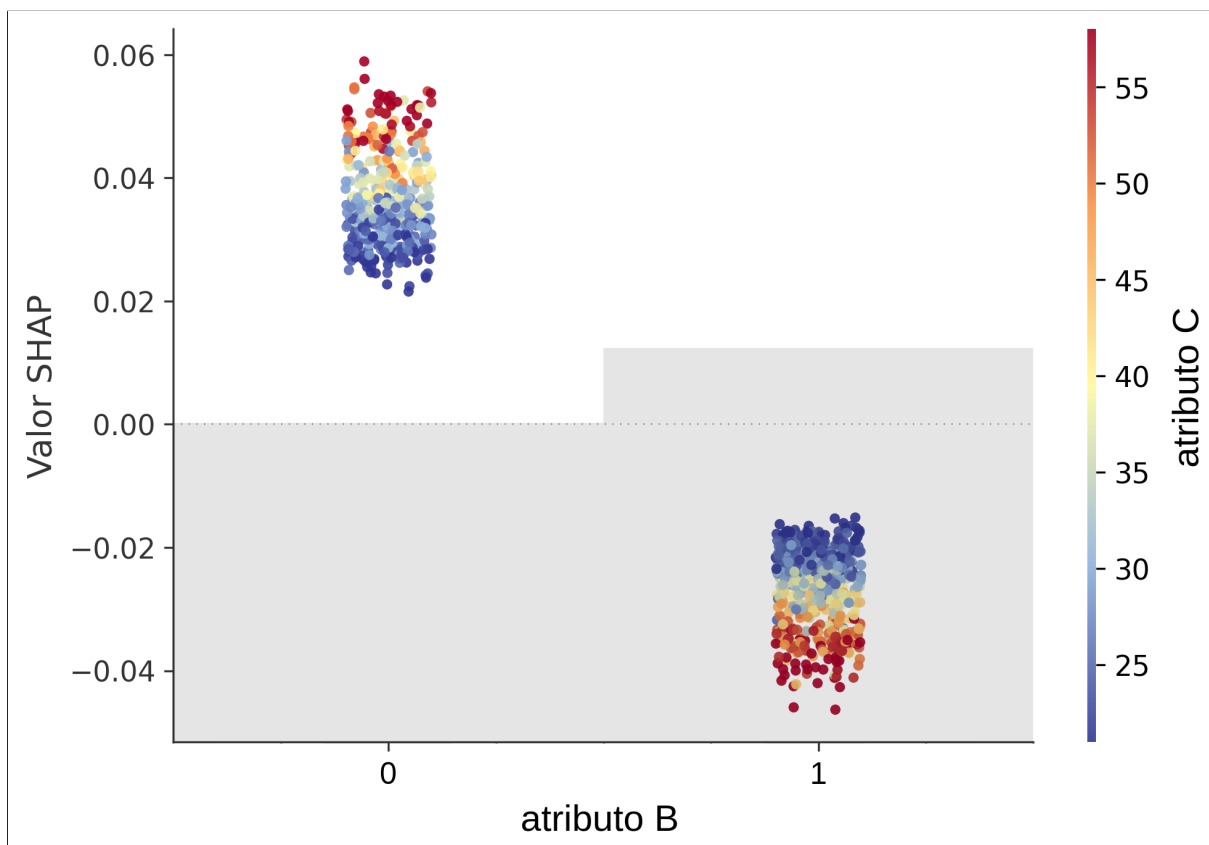
Fonte: elaborada pelo autor.

O segundo tipo de gráfico analisado é o de interação entre dois atributos. Esse gráfico explora como os valores de um atributo influenciam no valor SHAP de um outro atributo. Por exemplo, a Figura 10 apresenta a interação do atributo C com os valores SHAP do atributo B. No gráfico, o eixo x apresenta os valores da *feature* primária analisada. Já o eixo y indica o valor SHAP dessa *feature*. Assim como no gráfico *beeswarm*, os pontos indicam amostras do dataset. Além disso, as cores desses pontos recebem a mesma codificação de cor, mas essa cor representa o valor do atributo secundário, o qual deseja-se investigar a interação com o valor SHAP do atributo primário. Adicionalmente, o gráfico apresenta um

histograma cinza ao fundo, indicando a frequência de amostras com os respectivos valores do atributo primário.

Na Figura 10, percebe-se que amostras com valor 0 no atributo B aumentam a predição do modelo, enquanto que aquelas com valor 1 reduzem a predição. Além disso, pode-se analisar as cores das amostras para investigar as interações dos valores do atributo B com os do atributo C. Para as amostras que possuem 0 na *feature* B, um valor mais alto da *feature* C torna o impacto na saída ainda maior. Entretanto, o oposto ocorre para aquelas amostras que possuem 1 no atributo B: um valor mais alto da *feature* C torna o impacto na saída ainda menor. Ou seja, ter um valor alto no atributo C pode causar diferentes impactos no valor SHAP, dependendo do atributo B da amostra avaliada. Por último, o histograma ao fundo indica que nesse *dataset* genérico há um pouco mais de amostras com valor 1 no atributo B do que valor 0.

Figura 10 – Gráfico genérico representando o comportamento dos valores SHAP do atributo B de acordo com sua interação com o atributo C.



Fonte: elaborada pelo autor.

Por fim, ressalta-se que o foco da etapa de Interpretabilidade é dado para os experimentos E1-A, E1-F e E1-M, mas principalmente ao E1-A. Acredita-se que a interpretação nesses experimentos é mais relevante, pois eles utilizam todas as *features* do *dataset*, o que permite ter uma visão geral das diferenças de importância entre elas. Adicionalmente,

além do E1-A permitir investigar todos atributos, esse experimento também permite a análise de todos pacientes do conjunto de dados.

5 Resultados e Discussão

Este capítulo destina-se à apresentação dos resultados e à discussão dos mesmos. Inicialmente, na seção 5.1, são investigados em maior detalhe os resultados dos experimentos E1. Em seguida, na seção 5.2, apresentam-se brevemente os resultados dos outros experimentos e comparam-se os resultados de todos experimentos. Por fim, a seção 5.3 apresenta as análises de interpretação feitas nos experimentos E1.

Para tornar mais justa a comparação dos resultados de métricas de recurso computacional, todos os experimentos apresentados nesta pesquisa foram executados em um mesmo computador. Para também facilitar comparações com resultados de experimentos futuros, detalham-se as especificações da máquina utilizada: computador com sistema operacional Linux (Ubuntu 20.04), com 16 GB de memória RAM e processador Intel® Core™ i7-9750H. Os resultados apresentados nesta seção não são comparados com trabalhos relacionados, pois os experimentos encontrados na literatura utilizam uma variável alvo diferente ou possuem *datasets* com grupos demográficos diferentes.

5.1 Experimentos Principais

Esta seção apresenta os resultados e discussões dos experimentos do tipo E1, considerados os experimentos principais desta pesquisa. Os resultados do experimento E1-A, tido como o mais importante deste trabalho, é discutido na subseção 5.1.1. Já a subseção 5.1.2 e a subseção 5.1.3 discutem os resultados dos experimentos E1-F e E1-M, respectivamente.

5.1.1 E1-A

Esta subseção apresenta os resultados do experimento E1-A, considerado o experimento mais relevante da pesquisa, pois contém todas as amostras e todos atributos do *dataset*. Primeiramente, apresentam-se e discutem-se sobre as métricas de avaliação e métricas de recurso. Em seguida, analisam-se os testes estatísticos aplicados. Então, comenta-se sobre a abordagem que obteve destaque. Por fim, realiza-se uma investigação sobre o desempenho do modelo em amostras mais extremas.

A Tabela 8 mostra as métricas de avaliação para todas combinações de técnica de pré-processamento e algoritmo de ML obtidas no experimento E1-A. Percebe-se que o modelo SVM com o *kernel* RBF, combinado com a normalização MinMax, obteve as melhores médias das métricas, com exceção do ME. O MAE médio de 0,0774 g/cm² e o MAPE médio de 6,39% sugerem que esse par consegue prever o valor de DMO próximo

ao valor real. Já o modelo XGBoost, especialmente combinado com o pré-processamento MinMax, obteve as menores médias de ME. Esse comportamento mostra que os maiores erros do XGBoost tendem a não ser tão altos quanto os dos outros algoritmos.

Tabela 8 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0786 (0,0077)	0,0988 (0,0102)	6,49 (0,66)	0,2863 (0,0489)	0,4645 (0,0719)
Nenhum	EN	0,0784 (0,0075)	0,0986 (0,0099)	6,47 (0,65)	0,2856 (0,0478)	0,4665 (0,0698)
Nenhum	DT	0,0849 (0,0113)	0,1054 (0,0130)	7,00 (0,97)	0,2983 (0,0566)	0,3879 (0,1133)
Nenhum	MLP	0,0790 (0,0071)	0,0995 (0,0097)	6,53 (0,62)	0,2885 (0,0419)	0,4571 (0,0652)
Nenhum	k-NN	0,0790 (0,0074)	0,0986 (0,0097)	6,55 (0,63)	0,2736 (0,0458)	0,4659 (0,0712)
Nenhum	SVM_Lin	0,0934 (0,0212)	0,1186 (0,0287)	7,66 (1,77)	0,3477 (0,1283)	0,2101 (0,3328)
Nenhum	SVM_RBF	0,0776 (0,0068)	0,0990 (0,0117)	6,39 (0,60)	0,3071 (0,1033)	0,4609 (0,1004)
Nenhum	SVM_Pol	0,0805 (0,0071)	0,1018 (0,0091)	6,66 (0,62)	0,3000 (0,0453)	0,4270 (0,1031)
Nenhum	SVM_Sig	0,0821 (0,0112)	0,1037 (0,0130)	6,78 (0,93)	0,3028 (0,0478)	0,4033 (0,1383)
Nenhum	RF	0,0794 (0,0090)	0,0988 (0,0104)	6,55 (0,77)	0,2808 (0,0374)	0,4618 (0,0895)
Nenhum	XGBoost	0,0789 (0,0090)	0,0978 (0,0107)	6,50 (0,75)	0,2641 (0,0434)	0,4739 (0,0799)
MinMax	LR	0,0786 (0,0077)	0,0988 (0,0102)	6,49 (0,66)	0,2863 (0,0489)	0,4645 (0,0719)
MinMax	EN	0,0789 (0,0080)	0,0991 (0,0104)	6,51 (0,68)	0,2867 (0,0402)	0,4613 (0,0756)
MinMax	DT	0,0854 (0,0116)	0,1061 (0,0131)	7,04 (0,99)	0,2757 (0,0375)	0,3778 (0,1240)
MinMax	MLP	0,0793 (0,0080)	0,0997 (0,0107)	6,53 (0,67)	0,3001 (0,0650)	0,4545 (0,0808)
MinMax	k-NN	0,0819 (0,0078)	0,1022 (0,0090)	6,78 (0,67)	0,2880 (0,0382)	0,4284 (0,0526)
MinMax	SVM_Lin	0,0787 (0,0078)	0,0987 (0,0100)	6,49 (0,66)	0,2772 (0,0357)	0,4656 (0,0686)
MinMax	SVM_RBF	0,0774 (0,0077)	0,0974 (0,0095)	6,39 (0,66)	0,2842 (0,0475)	0,4794 (0,0643)
MinMax	SVM_Pol	0,0793 (0,0082)	0,1007 (0,0120)	6,54 (0,76)	0,3068 (0,1125)	0,4391 (0,1201)
MinMax	SVM_Sig	0,0788 (0,0076)	0,0988 (0,0098)	6,49 (0,65)	0,2766 (0,0373)	0,4644 (0,0664)
MinMax	RF	0,0784 (0,0090)	0,0981 (0,0110)	6,47 (0,77)	0,2704 (0,0466)	0,4713 (0,0865)
MinMax	XGBoost	0,0785 (0,0081)	0,0976 (0,0104)	6,47 (0,67)	0,2634 (0,0425)	0,4767 (0,0761)
Padronização	LR	0,0786 (0,0077)	0,0988 (0,0102)	6,49 (0,66)	0,2863 (0,0489)	0,4645 (0,0719)
Padronização	EN	0,0783 (0,0078)	0,0983 (0,0099)	6,46 (0,67)	0,2806 (0,0382)	0,4697 (0,0677)
Padronização	DT	0,0845 (0,0113)	0,1054 (0,0131)	6,98 (0,97)	0,2970 (0,0591)	0,3871 (0,1199)
Padronização	MLP	0,0802 (0,0082)	0,1000 (0,0105)	6,61 (0,70)	0,2904 (0,0503)	0,4510 (0,0752)
Padronização	k-NN	0,0794 (0,0078)	0,0992 (0,0095)	6,58 (0,67)	0,2834 (0,0414)	0,4615 (0,0610)
Padronização	SVM_Lin	0,0786 (0,0077)	0,0987 (0,0102)	6,48 (0,66)	0,2790 (0,0482)	0,4665 (0,0685)
Padronização	SVM_RBF	0,0786 (0,0077)	0,0984 (0,0095)	6,49 (0,65)	0,2799 (0,0415)	0,4688 (0,0672)
Padronização	SVM_Pol	0,0912 (0,0072)	0,1161 (0,0105)	7,52 (0,69)	0,3392 (0,0666)	0,2567 (0,1205)
Padronização	SVM_Sig	0,0781 (0,0075)	0,0981 (0,0097)	6,44 (0,64)	0,2750 (0,0405)	0,4720 (0,0647)
Padronização	RF	0,0782 (0,0079)	0,0976 (0,0098)	6,45 (0,69)	0,2794 (0,0362)	0,4753 (0,0814)
Padronização	XGBoost	0,0789 (0,0084)	0,0983 (0,0105)	6,51 (0,71)	0,2669 (0,0418)	0,4688 (0,0769)

Fonte: elaborada pelo autor.

A Tabela 9 complementa os resultados do experimento E1-A, apresentando as médias e desvios-padrão das métricas de recurso. Verifica-se que o k-NN sem pré-processamento teve o Ttr médio mais baixo, apenas 1 ms. Esse resultado é esperado, já que a etapa de “treino” do k-NN consiste em apenas carregar os dados. Em contrapartida, o k-NN apresenta uns dos valores de Tte mais altos, já que há a necessidade de calcular a distância entre a amostra testada e todas as outras carregadas no treino. Em relação ao armazenamento, os modelos LR e EN, sem pré-processamento, apresentaram o menor valor de TA, com a média de 1 kB. Esses algoritmos ocupam pouco espaço pois precisam apenas armazenar os coeficientes da reta ajustada. Já a RF apresenta os maiores valores de Ttr, Tte e TA. Esses tempos longos e esse grande tamanho devem-se ao alto número de árvores utilizadas para a construção da RF. Por fim, os resultados encontrados na Tabela 9 demonstram que a maioria dos modelos poderia ser colocada em produção em sistemas com poucos recursos

Tabela 9 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	2 (0,7)	0,8 (0,15)	1 (0,0)
Nenhum	EN	4 (2,1)	0,8 (0,04)	1 (0,0)
Nenhum	DT	15 (14,4)	0,7 (0,05)	3 (0,5)
Nenhum	MLP	739 (1137,9)	0,8 (0,11)	7 (1,6)
Nenhum	k-NN	1 (0,2)	2,3 (0,34)	116 (0,1)
Nenhum	SVM_Lin	64 (5,6)	1,7 (0,34)	38 (3,6)
Nenhum	SVM_RBF	16 (5,6)	1,2 (0,47)	39 (1,4)
Nenhum	SVM_Pol	30 (18,8)	0,9 (0,18)	40 (1,6)
Nenhum	SVM_Sig	19 (5,8)	0,8 (0,04)	44 (5,0)
Nenhum	RF	1545 (1588,1)	14,9 (6,51)	961 (1197,3)
Nenhum	XGBoost	78 (63,3)	1,7 (0,14)	137 (88,0)
MinMax	LR	4 (1,4)	0,9 (0,08)	2 (0,0)
MinMax	EN	5 (1,0)	0,8 (0,04)	2 (0,0)
MinMax	DT	19 (15,6)	0,8 (0,04)	5 (0,7)
MinMax	MLP	77 (37,7)	0,9 (0,03)	7 (1,0)
MinMax	k-NN	3 (0,4)	2,7 (0,65)	117 (0,1)
MinMax	SVM_Lin	13 (4,5)	2,0 (0,07)	40 (1,2)
MinMax	SVM_RBF	20 (8,5)	1,1 (0,43)	39 (1,3)
MinMax	SVM_Pol	21 (10,0)	0,9 (0,10)	40 (1,3)
MinMax	SVM_Sig	20 (7,1)	0,9 (0,07)	40 (1,2)
MinMax	RF	1386 (1410,8)	13,9 (6,63)	752 (462,9)
MinMax	XGBoost	66 (119,9)	1,9 (0,04)	82 (56,6)
Padronização	LR	4 (1,5)	0,9 (0,11)	2 (0,0)
Padronização	EN	4 (0,5)	0,8 (0,07)	2 (0,0)
Padronização	DT	8 (6,5)	0,8 (0,04)	4 (0,3)
Padronização	MLP	82 (50,8)	0,9 (0,04)	9 (3,2)
Padronização	k-NN	3 (0,3)	3,1 (1,60)	117 (0,1)
Padronização	SVM_Lin	20 (1,7)	2,1 (0,03)	40 (0,9)
Padronização	SVM_RBF	17 (6,1)	1,0 (0,38)	39 (1,2)
Padronização	SVM_Pol	20 (9,5)	0,8 (0,05)	48 (3,5)
Padronização	SVM_Sig	16 (3,1)	0,9 (0,08)	40 (1,1)
Padronização	RF	2562 (2950,0)	14,9 (7,90)	668 (530,8)
Padronização	XGBoost	60 (73,6)	2,0 (0,23)	79 (51,1)

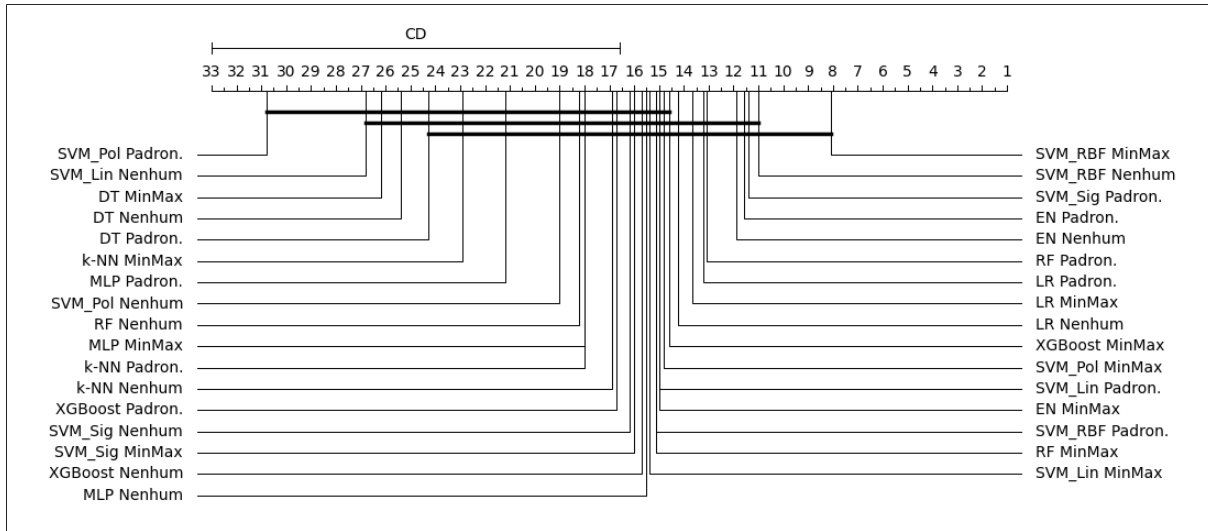
Fonte: elaborada pelo autor.

computacionais, inclusive em sistemas embarcados para serem usados por especialistas.

Para avaliar se há alguma diferença significativa entre os resultados de MAE obtidos no experimento E1-A, aplicou-se inicialmente o teste de Friedman. Nesse teste, obteve-se um valor-p de 2×10^{-7} (valor-p < 0.05). Assim, rejeita-se a hipótese nula e assume-se que há pelo menos um par de combinações que geraram resultados significativamente diferentes. Em seguida, realizou-se o teste de Nemenyi para identificar quais pares possuem essa diferença significativa. Na Figura 11 está apresentado o diagrama de diferença crítica obtido pelo teste de Nemenyi. Pela figura, percebe-se que a combinação SMV_RBF com MinMax teve o melhor rank médio. Esse par apresentou apenas diferenças significativas com as combinações SVM_Pol com Padronização, SVM_Lin sem pré-processamento, DT com MinMax e DT sem pré-processamento.

Em conclusão, a combinação SMV_RBF com MinMax apresentou as melhores

Figura 11 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-A. (CD = 16,4)



Fonte: elaborada pelo autor.

médias de 4 das 5 métricas de avaliação. Seu MAE não é significativamente diferente do valor do MAE de outras 28 combinações. Entretanto, esse par obteve a posição mais alta no diagrama de diferença crítica. Portanto, considera-se o mais adequado para prever DMO ao considerar todas amostras e *features* do *dataset*.

Não é surpreendente que o SVM_RBF tenha obtido os melhores resultados, já que é um algoritmo que é comumente utilizado em problemas de regressão por apresentar uma boa capacidade de generalização. Além disso, o MinMax leva todos os atributos para o mesmo intervalo de valores. Como as SVMs trabalham com distâncias, esse pré-processamento pode ter ajudado o modelo, já que a normalização diminui o impacto das escalas originais das *features*, como discutido na seção 4.2.

Além das avaliações já discutidas, realiza-se uma análise adicional do par SVM_RBF com MinMax no experimento E1-A. Pressupõe-se que amostras com valores de DMO mais extremos (muito baixos ou muito altos) são passíveis de sofrer maiores erros nas predições. Portanto, efetuou-se uma investigação extra para analisar o desempenho que a abordagem teve nesses casos. Primeiramente, agregaram-se todas as amostras de teste dos *outer folds* e as respectivas predições realizadas pelo par escolhido. Em seguida, esse conjunto foi ordenado crescentemente, com base nos valores reais de DMO das amostras. Então, dividiu-se o conjunto em quatro subconjuntos, cada um contendo 25% das amostras. A Tabela 10 apresenta esses subconjuntos e os respectivos intervalos de DMO encontrados em suas amostras. Por fim, as predições obtidas da etapa de *nested cross-validation* foram respectivamente comparadas com os valores de DMO das amostras, e calcularam-se as métricas para cada subconjunto separadamente. Os resultados encontrados são apresentados na Tabela 10.

Tabela 10 – Métricas obtidas pela abordagem selecionada no E1-A ao avaliar as amostras de acordo com intervalos de DMO. As amostras de teste são ordenadas pelo valor de DMO e então divididas em quatro subconjuntos de mesmo tamanho.

Subconjunto	Intervalo de DMO	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)
S_{0-25}	[0,789, 1,131]	0,0952	0,1144	9,2	0,3027
S_{25-50}	[1,132, 1,229]	0,0555	0,0722	4,69	0,3462
S_{50-75}	[1,230, 1,320]	0,0619	0,0785	4,87	0,3688
S_{75-100}	[1,320, 1,689]	0,0968	0,1176	6,77	0,3325

Fonte: elaborada pelo autor.

Pelos resultados apresentados na Tabela 10, verifica-se que o subconjunto S_{25-50} obteve os menores MAE, RMSE e MAPE. Isso indica que, no geral, a abordagem comete os menores erros médios e percentuais em pacientes que têm um valor de DMO médio-baixo. Já em relação ao ME, o subconjunto S_{0-25} foi o que atingiu o menor valor. Esse resultado sugere que pacientes com baixo valor de DMO sofrem os menores erros máximos. Analisando a tabela por completo, percebe-se que as amostras com valores extremos de DMO sofrem os maiores erros médios e percentuais, que pode ter acontecido por esses grupos terem uma maior faixa de DMO e, portanto, serem mais difíceis de modelar. Ainda assim, o maior MAPE foi de apenas 9,2% e o maior MAE foi de 0,0968, o que sugere que o modelo consegue manter bons resultados mesmo em casos mais extremos. Por fim, ressalta-se que a métrica R^2 foi omitida nessa avaliação, pois a ordenação dos subconjuntos interfere diretamente na variância dos mesmos, tornando essa métrica pouco relevante.

5.1.2 E1-F

Esta subseção apresenta os resultados do experimento E1-F, que contém todos atributos do *dataset* e apenas amostras do sexo feminino. De início, apresentam-se e discutem-se sobre as métricas de avaliação e métricas de recurso. Logo então, avaliam-se os testes estatísticos aplicados. Por fim, discute-se sobre a abordagem mais adequada para este experimento.

A Tabela 11 apresenta as métricas de avaliação para todos pares de técnica de pré-processamento e modelo obtidos no experimento E1-F. Verifica-se que o algoritmo RF, antecedido pela técnica de Padronização, obteve as melhores médias das métricas, com exceção do ME. Essa combinação obteve MAE médio de 0,0718 g/cm², MAPE médio de 6,24% e RMSE médio de apenas 0,0899 g/cm², o que indica que essa abordagem costuma cometer erros pequenos ao predizer a DMO. Já em relação ao ME médio, o modelo k-NN sem pré-processamento obteve a menor média: apenas 0,2216 g/cm².

Algumas combinações tiveram desempenhos inferiores ou instáveis e podem não ser adequadas para resolver a atividade de estimar DMO de pacientes do sexo feminino. Por exemplo, o SVM com *kernel* linear, sem pré-processamento, teve um ME de 0,3021 (0,1717) g/cm² e R^2 de 0,0126 (0,5603). Esse alto valor do desvio-padrão indica que este

par teve um ótimo desempenho em alguns *folds*, mas péssimo em outros, demonstrando um comportamento instável. Logo, não se mostra como uma abordagem confiável para ser aplicada em um ambiente real.

Tabela 11 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0750 (0,0102)	0,0938 (0,0114)	6,51 (0,89)	0,2485 (0,0648)	0,2108 (0,1062)
Nenhum	EN	0,0744 (0,0100)	0,0930 (0,0110)	6,48 (0,89)	0,2439 (0,0567)	0,2249 (0,0822)
Nenhum	DT	0,0758 (0,0095)	0,0951 (0,0113)	6,59 (0,89)	0,2497 (0,0450)	0,1875 (0,1049)
Nenhum	MLP	0,0742 (0,0102)	0,0930 (0,0108)	6,44 (0,83)	0,2391 (0,0625)	0,2220 (0,1037)
Nenhum	k-NN	0,0737 (0,0095)	0,0914 (0,0105)	6,43 (0,85)	0,2216 (0,0284)	0,2514 (0,0738)
Nenhum	SVM_Lin	0,0795 (0,0144)	0,1020 (0,0227)	6,90 (1,21)	0,3021 (0,1717)	0,0126 (0,5603)
Nenhum	SVM_RBF	0,0733 (0,0087)	0,0907 (0,0097)	6,36 (0,76)	0,2334 (0,0428)	0,2595 (0,0871)
Nenhum	SVM_Pol	0,0750 (0,0104)	0,0942 (0,0116)	6,52 (0,90)	0,2671 (0,0717)	0,2014 (0,1227)
Nenhum	SVM_Sig	0,0789 (0,0116)	0,0978 (0,0139)	6,86 (1,04)	0,2470 (0,0338)	0,1440 (0,1284)
Nenhum	RF	0,0727 (0,0094)	0,0910 (0,0104)	6,32 (0,83)	0,2316 (0,0400)	0,2568 (0,0796)
Nenhum	XGBoost	0,0760 (0,0093)	0,0952 (0,0098)	6,60 (0,79)	0,2566 (0,0625)	0,1782 (0,1370)
MinMax	LR	0,0750 (0,0102)	0,0938 (0,0114)	6,51 (0,89)	0,2485 (0,0648)	0,2108 (0,1062)
MinMax	EN	0,0741 (0,0102)	0,0922 (0,0108)	6,44 (0,87)	0,2379 (0,0487)	0,2383 (0,0748)
MinMax	DT	0,0775 (0,0087)	0,0967 (0,0101)	6,74 (0,83)	0,2422 (0,0519)	0,1598 (0,0939)
MinMax	MLP	0,0736 (0,0094)	0,0912 (0,0098)	6,39 (0,80)	0,2223 (0,0271)	0,2541 (0,0550)
MinMax	k-NN	0,0781 (0,0091)	0,0972 (0,0094)	6,79 (0,83)	0,2423 (0,0393)	0,1503 (0,0845)
MinMax	SVM_Lin	0,0739 (0,0103)	0,0917 (0,0107)	6,42 (0,90)	0,2337 (0,0373)	0,2473 (0,0687)
MinMax	SVM_RBF	0,0729 (0,0096)	0,0903 (0,0104)	6,33 (0,81)	0,2241 (0,0336)	0,2704 (0,0689)
MinMax	SVM_Pol	0,0745 (0,0094)	0,0922 (0,0096)	6,46 (0,78)	0,2314 (0,0322)	0,2351 (0,0864)
MinMax	SVM_Sig	0,0737 (0,0105)	0,0917 (0,0111)	6,41 (0,93)	0,2373 (0,0389)	0,2481 (0,0645)
MinMax	RF	0,0722 (0,0093)	0,0905 (0,0102)	6,28 (0,83)	0,2288 (0,0439)	0,2662 (0,0762)
MinMax	XGBoost	0,0754 (0,0110)	0,0942 (0,0113)	6,53 (0,92)	0,2394 (0,0543)	0,1997 (0,1271)
Padronização	LR	0,0750 (0,0102)	0,0938 (0,0114)	6,51 (0,89)	0,2485 (0,0648)	0,2108 (0,1062)
Padronização	EN	0,0742 (0,0101)	0,0924 (0,0107)	6,46 (0,86)	0,2409 (0,0568)	0,2350 (0,0832)
Padronização	DT	0,0774 (0,0099)	0,0964 (0,0115)	6,74 (0,92)	0,2418 (0,0500)	0,1675 (0,0865)
Padronização	MLP	0,0745 (0,0102)	0,0929 (0,0105)	6,48 (0,85)	0,2321 (0,0401)	0,2257 (0,0858)
Padronização	k-NN	0,0749 (0,0088)	0,0929 (0,0090)	6,53 (0,77)	0,2300 (0,0333)	0,2261 (0,0446)
Padronização	SVM_Lin	0,0740 (0,0101)	0,0922 (0,0109)	6,43 (0,88)	0,2374 (0,0493)	0,2392 (0,0804)
Padronização	SVM_RBF	0,0746 (0,0111)	0,0928 (0,0116)	6,50 (0,97)	0,2338 (0,0309)	0,2297 (0,0847)
Padronização	SVM_Pol	0,0851 (0,0087)	0,1044 (0,0097)	7,37 (0,77)	0,2737 (0,0315)	0,0203 (0,0766)
Padronização	SVM_Sig	0,0744 (0,0101)	0,0923 (0,0107)	6,47 (0,88)	0,2328 (0,0407)	0,2367 (0,0723)
Padronização	RF	0,0718 (0,0095)	0,0899 (0,0102)	6,24 (0,81)	0,2287 (0,0368)	0,2762 (0,0720)
Padronização	XGBoost	0,0744 (0,0107)	0,0927 (0,0112)	6,46 (0,90)	0,2383 (0,0487)	0,2240 (0,1376)

Fonte: elaborada pelo autor.

A Tabela 12 acrescenta aos resultados do experimento E1-F, indicando as médias e desvios-padrão das métricas de recurso. Percebe-se que, para essas métricas, a relação entre as abordagens se assemelha à relação encontrada na Tabela 9, mas os valores do E1-F são menores na maioria dos casos. Nenhuma das abordagens levou mais que um segundo em média para treinar o modelo. Além disso, com exceção do RF, todos modelos levam menos de 10 ms para predizer a DMO e ocupam menos de 100 kB de armazenamento. Por sua vez, a combinação RF com Padronização ocupa em média aproximadamente 2 MB, o que é consideravelmente maior que o TA dos outros modelos, mas ainda é um tamanho pequeno para grande parte das máquinas.

Para verificar se há alguma diferença significativa entre os resultados de MAE obtidos no experimento E1-F, também aplicou-se o teste de Friedman. Nesse teste, obteve-se um valor-p de 5×10^{-7} (valor-p < 0.05). Com isso, rejeita-se a hipótese nula e assume-se que

Tabela 12 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

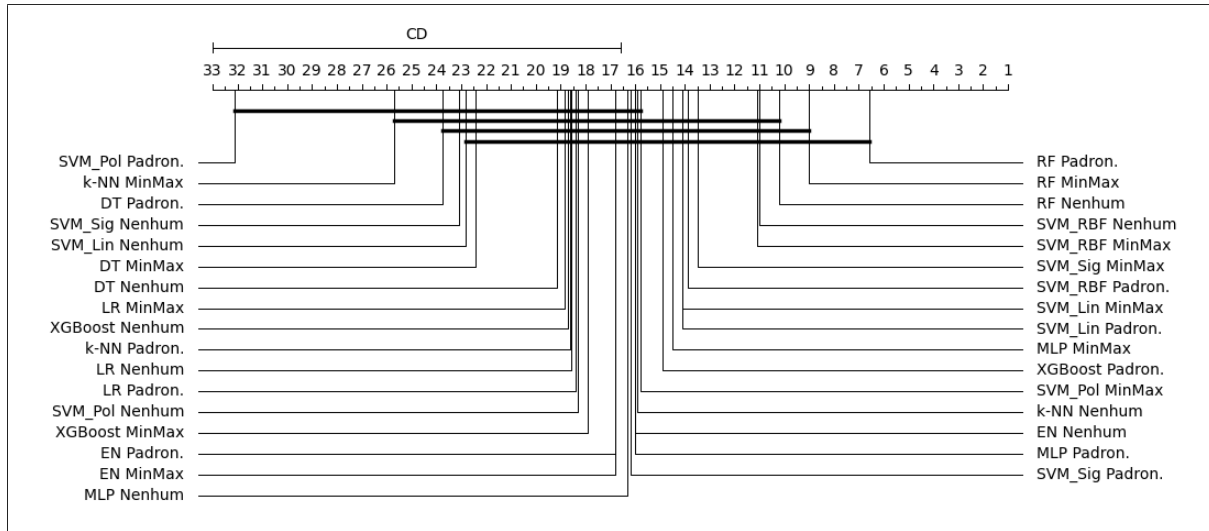
Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,1)	0,7 (0,07)	1 (0,0)
Nenhum	EN	2 (0,3)	0,7 (0,04)	1 (0,0)
Nenhum	DT	4 (4,3)	0,7 (0,03)	3 (0,5)
Nenhum	MLP	199 (270,7)	0,7 (0,02)	7 (1,8)
Nenhum	k-NN	1 (0,1)	1,0 (0,03)	61 (0,0)
Nenhum	SVM_Lin	33 (5,7)	0,7 (0,03)	21 (1,2)
Nenhum	SVM_RBF	6 (2,2)	0,8 (0,03)	20 (0,8)
Nenhum	SVM_Pol	8 (4,1)	0,7 (0,02)	21 (1,2)
Nenhum	SVM_Sig	8 (4,8)	0,7 (0,00)	21 (1,3)
Nenhum	RF	930 (1004,3)	18,8 (5,02)	882 (715,5)
Nenhum	XGBoost	135 (206,8)	1,8 (0,24)	45 (14,3)
MinMax	LR	3 (0,2)	0,8 (0,09)	2 (0,0)
MinMax	EN	3 (0,1)	0,8 (0,04)	2 (0,0)
MinMax	DT	4 (0,6)	0,8 (0,02)	4 (0,3)
MinMax	MLP	39 (22,3)	0,8 (0,03)	9 (2,3)
MinMax	k-NN	3 (1,8)	1,1 (0,03)	62 (0,0)
MinMax	SVM_Lin	9 (4,0)	0,9 (0,03)	21 (0,7)
MinMax	SVM_RBF	14 (7,9)	0,9 (0,02)	22 (1,1)
MinMax	SVM_Pol	8 (2,8)	0,9 (0,01)	22 (0,8)
MinMax	SVM_Sig	9 (2,4)	0,9 (0,02)	21 (0,7)
MinMax	RF	411 (194,7)	14,8 (7,43)	703 (675,6)
MinMax	XGBoost	93 (216,7)	2,5 (1,81)	60 (41,6)
Padronização	LR	3 (0,2)	0,9 (0,11)	2 (0,0)
Padronização	EN	3 (0,4)	0,8 (0,04)	2 (0,0)
Padronização	DT	4 (0,8)	0,9 (0,08)	3 (0,4)
Padronização	MLP	25 (15,2)	0,8 (0,03)	9 (2,5)
Padronização	k-NN	3 (0,6)	1,1 (0,11)	62 (0,0)
Padronização	SVM_Lin	13 (3,2)	0,9 (0,11)	21 (0,8)
Padronização	SVM_RBF	8 (3,1)	0,9 (0,01)	21 (0,8)
Padronização	SVM_Pol	8 (4,5)	0,9 (0,11)	24 (1,3)
Padronização	SVM_Sig	8 (2,2)	0,9 (0,09)	21 (0,6)
Padronização	RF	581 (461,9)	12,7 (6,66)	1752 (3495,2)
Padronização	XGBoost	62 (92,8)	2,0 (0,26)	63 (55,7)

Fonte: elaborada pelo autor.

há pelo menos um par de combinações que geraram resultados significativamente diferentes. Com a rejeição da hipótese nula, realizou-se o teste de Nemenyi para identificar quais combinações possuem diferença significativa. Na Figura 12 apresenta-se o diagrama de CD obtido no teste de Nemenyi. Analisando o diagrama, verifica-se que os três pares no topo do rank incluem RF como o modelo, sendo a combinação com Padronização a que alcançou o melhor rank médio. Considerando uma diferença crítica de 16,4, o par RF com Padronização mostrou-se significativamente diferente das combinações SVM_Pol com Padronização, k-NN com MinMax, DT com Padronização e SVM_Sig sem pré-processamento.

Portanto, a combinação RF com Padronização apresentou as melhores médias de 4 das 5 métricas de avaliação. O MAE desse par não é significativamente diferente do valor do MAE de outras 28 combinações. Todavia, esse par também obteve a posição mais alta no diagrama de CD. Portanto, considera-se o mais adequado para prever DMO ao

Figura 12 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-F. (CD = 16,4)



Fonte: elaborada pelo autor.

considerar todas as *features* do subconjunto de pacientes do sexo feminino.

Como a floresta aleatória é um método *ensemble* que usa o poder preditivo de um conjunto de árvores de decisão, ela possui uma alta capacidade de aprendizado ao mesmo tempo que evita o *overfitting*. Por serem construídas com DT, a diferença de escala das *features* não causa impacto na modelagem. Ao analisar os resultados do experimento E1-F, percebe-se que as três abordagens que utilizam RF tiveram resultados semelhantes. As pequenas diferenças encontradas devem-se à aleatoriedade presente no processo de aprendizado.

5.1.3 E1-M

Esta subseção apresenta os resultados do experimento E1-M, que contém todos atributos do *dataset* e apenas amostras do sexo masculino. Primeiramente, apresentam-se e discutem-se sobre as métricas de avaliação e métricas de recurso. Depois, examinam-se os resultados dos testes estatísticos. Por fim, analisa-se a combinação de modelo e pré-processamento mais adequada para este experimento.

A Tabela 13 mostra as métricas de avaliação para as combinações de pré-processamento e algoritmo de ML obtidas no experimento E1-M. Nota-se que o algoritmo SVM com o *kernel* RBF, sem uso de pré-processamento, obteve as melhores médias em todas métricas, com exceção do ME. Esse par obteve MAE médio de 0,0814 g/cm² e MAPE médio de 6,30%, indicando que essa abordagem costuma prever valores de DMO próximos ao real. A combinação que obteve o ME médio mais baixo foi a MLP com o MinMax, com apenas 0,2554 g/cm².

Assim como no E1-F, algumas abordagens no E1-M também apresentaram resultados inconstantes ou inferiores. Por exemplo, MLP sem pré-processamento teve um MAE médio próximo do mais baixo, mas apresentou um ME de 0,3894 (0,3202) g/cm², indicando que em alguns casos esse par cometeu drásticos erros. Outro exemplo é o par SMV_Pol com Padronização, o qual teve a pior ou uma das piores médias para cada uma das métricas de avaliação.

Tabela 13 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E1-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0836 (0,0108)	0,1047 (0,0143)	6,50 (0,91)	0,2608 (0,0551)	0,1977 (0,2161)
Nenhum	EN	0,0829 (0,0114)	0,1039 (0,0142)	6,44 (0,95)	0,2573 (0,0526)	0,2144 (0,1979)
Nenhum	DT	0,0925 (0,0115)	0,1149 (0,0152)	7,18 (0,99)	0,2776 (0,0564)	0,0555 (0,1437)
Nenhum	MLP	0,0905 (0,0186)	0,1214 (0,0440)	7,02 (1,55)	0,3894 (0,3202)	-0,0697 (0,5681)
Nenhum	k-NN	0,0856 (0,0115)	0,1076 (0,0143)	6,68 (0,98)	0,2713 (0,0496)	0,1666 (0,1677)
Nenhum	SVM_Lin	0,0901 (0,0138)	0,1133 (0,0187)	6,92 (1,02)	0,2843 (0,0620)	-0,0629 (0,8020)
Nenhum	SVM_RBF	0,0814 (0,0100)	0,1014 (0,0131)	6,30 (0,85)	0,2574 (0,0526)	0,2496 (0,1857)
Nenhum	SVM_Pol	0,0842 (0,0145)	0,1057 (0,0184)	6,55 (1,25)	0,2674 (0,0527)	0,1896 (0,2131)
Nenhum	SVM_Sig	0,0907 (0,0153)	0,1134 (0,0185)	7,03 (1,22)	0,2742 (0,0495)	0,0871 (0,1476)
Nenhum	RF	0,0844 (0,0103)	0,1063 (0,0142)	6,56 (0,87)	0,2656 (0,0626)	0,1857 (0,1585)
Nenhum	XGBoost	0,0826 (0,0115)	0,1048 (0,0152)	6,41 (0,98)	0,2671 (0,0537)	0,2060 (0,1801)
MinMax	LR	0,0836 (0,0108)	0,1047 (0,0143)	6,50 (0,91)	0,2608 (0,0551)	0,1977 (0,2161)
MinMax	EN	0,0834 (0,0114)	0,1047 (0,0146)	6,49 (0,97)	0,2605 (0,0511)	0,2079 (0,1767)
MinMax	DT	0,0931 (0,0114)	0,1164 (0,0153)	7,23 (0,98)	0,2890 (0,0577)	0,0272 (0,1762)
MinMax	MLP	0,0845 (0,0106)	0,1050 (0,0136)	6,57 (0,90)	0,2554 (0,0541)	0,1959 (0,1955)
MinMax	k-NN	0,0855 (0,0122)	0,1072 (0,0161)	6,68 (1,05)	0,2617 (0,0506)	0,1783 (0,1650)
MinMax	SVM_Lin	0,0831 (0,0114)	0,1041 (0,0145)	6,43 (0,95)	0,2573 (0,0553)	0,2140 (0,1831)
MinMax	SVM_RBF	0,0830 (0,0107)	0,1034 (0,0137)	6,43 (0,90)	0,2634 (0,0436)	0,2223 (0,1810)
MinMax	SVM_Pol	0,0852 (0,0136)	0,1073 (0,0178)	6,60 (1,15)	0,2623 (0,0516)	0,1724 (0,1808)
MinMax	SVM_Sig	0,0830 (0,0114)	0,1045 (0,0143)	6,43 (0,94)	0,2606 (0,0535)	0,2082 (0,1819)
MinMax	RF	0,0841 (0,0116)	0,1062 (0,0155)	6,55 (0,98)	0,2676 (0,0670)	0,1836 (0,1870)
MinMax	XGBoost	0,0839 (0,0104)	0,1068 (0,0140)	6,51 (0,90)	0,2792 (0,0495)	0,1690 (0,2031)
Padronização	LR	0,0836 (0,0108)	0,1047 (0,0143)	6,50 (0,91)	0,2608 (0,0551)	0,1977 (0,2161)
Padronização	EN	0,0827 (0,0118)	0,1040 (0,0150)	6,44 (0,99)	0,2578 (0,0573)	0,2173 (0,1817)
Padronização	DT	0,0905 (0,0119)	0,1152 (0,0172)	7,03 (1,02)	0,2874 (0,0798)	0,0472 (0,1931)
Padronização	MLP	0,0843 (0,0110)	0,1051 (0,0137)	6,55 (0,92)	0,2607 (0,0538)	0,1939 (0,2047)
Padronização	k-NN	0,0852 (0,0108)	0,1061 (0,0141)	6,65 (0,94)	0,2615 (0,0437)	0,1895 (0,1645)
Padronização	SVM_Lin	0,0843 (0,0110)	0,1054 (0,0145)	6,52 (0,90)	0,2649 (0,0567)	0,1842 (0,2317)
Padronização	SVM_RBF	0,0832 (0,0111)	0,1039 (0,0149)	6,45 (0,93)	0,2589 (0,0596)	0,2131 (0,2044)
Padronização	SVM_Pol	0,0945 (0,0122)	0,1203 (0,0171)	7,39 (1,08)	0,3231 (0,0876)	-0,0277 (0,1362)
Padronização	SVM_Sig	0,0829 (0,0113)	0,1042 (0,0147)	6,42 (0,93)	0,2571 (0,0545)	0,2103 (0,1941)
Padronização	RF	0,0846 (0,0116)	0,1068 (0,0155)	6,59 (0,98)	0,2676 (0,0663)	0,1737 (0,1887)
Padronização	XGBoost	0,0837 (0,0119)	0,1056 (0,0156)	6,50 (0,98)	0,2703 (0,0487)	0,1889 (0,2134)

Fonte: elaborada pelo autor.

As médias e desvios-padrão das métricas de recurso do experimento E1-M são apresentadas na Tabela 14. Verifica-se que os valores encontrados para essas métricas são muito semelhantes aos da Tabela 12. A combinação anteriormente mencionada, SVM_RBF sem pré-processamento, apresentou o Tte médio mais baixo, somente 0,7 ms. Além disso, requer apenas 15 kB em média de espaço de armazenamento.

Assim como nos outros experimentos, aplicou-se o teste de Friedman para avaliar se há alguma diferença significativa entre os resultados de MAE obtidos no experimento E1-M. No teste, obteve-se um valor-p de 6×10^{-9} (valor-p < 0.05). Dessa forma, também rejeita-se a hipótese nula e assume-se que há pelo menos um par de combinações que geraram resultados

Tabela 14 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E1-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida para a etapa de interpretação.

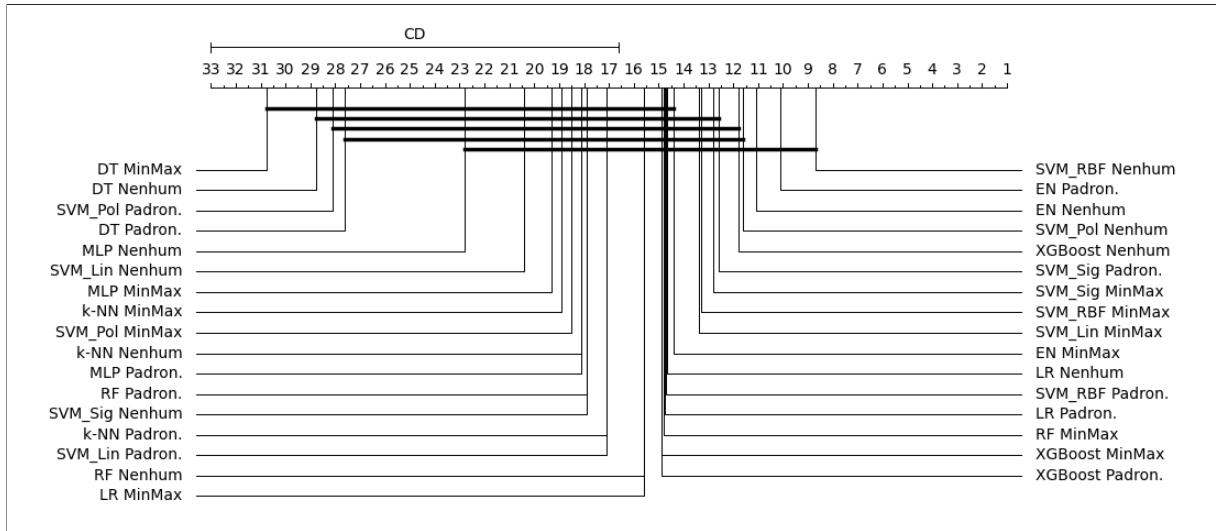
Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,0)	0,7 (0,02)	1 (0,0)
Nenhum	EN	2 (0,1)	0,8 (0,11)	1 (0,0)
Nenhum	DT	4 (2,1)	0,7 (0,01)	2 (0,2)
Nenhum	MLP	82 (101,0)	0,7 (0,06)	6 (1,4)
Nenhum	k-NN	1 (0,2)	0,9 (0,03)	41 (0,1)
Nenhum	SVM_Lin	24 (9,2)	0,7 (0,02)	17 (0,9)
Nenhum	SVM_RBF	7 (2,3)	0,7 (0,02)	15 (0,4)
Nenhum	SVM_Pol	9 (3,9)	0,7 (0,02)	16 (0,8)
Nenhum	SVM_Sig	5 (1,8)	0,7 (0,01)	17 (1,6)
Nenhum	RF	807 (514,7)	14,2 (7,35)	351 (210,1)
Nenhum	XGBoost	259 (469,9)	2,5 (2,53)	44 (23,7)
MinMax	LR	2 (0,0)	0,8 (0,09)	2 (0,0)
MinMax	EN	4 (0,8)	0,8 (0,06)	2 (0,0)
MinMax	DT	5 (2,4)	0,8 (0,04)	3 (0,3)
MinMax	MLP	46 (50,6)	0,8 (0,02)	7 (2,6)
MinMax	k-NN	3 (0,4)	1,0 (0,02)	42 (0,1)
MinMax	SVM_Lin	6 (2,2)	0,8 (0,03)	17 (0,3)
MinMax	SVM_RBF	6 (1,6)	0,8 (0,02)	16 (0,5)
MinMax	SVM_Pol	4 (0,6)	0,8 (0,02)	17 (0,5)
MinMax	SVM_Sig	7 (4,0)	0,8 (0,02)	17 (0,4)
MinMax	RF	698 (501,3)	15,0 (6,53)	351 (213,7)
MinMax	XGBoost	134 (309,6)	1,9 (0,04)	90 (70,0)
Padronização	LR	3 (0,1)	0,8 (0,10)	2 (0,0)
Padronização	EN	3 (0,1)	0,8 (0,02)	2 (0,0)
Padronização	DT	5 (1,4)	0,8 (0,03)	3 (0,3)
Padronização	MLP	30 (15,8)	0,8 (0,04)	10 (2,9)
Padronização	k-NN	3 (0,5)	1,0 (0,02)	42 (0,1)
Padronização	SVM_Lin	10 (5,0)	0,8 (0,02)	16 (0,6)
Padronização	SVM_RBF	5 (1,5)	0,8 (0,02)	16 (0,3)
Padronização	SVM_Pol	5 (0,5)	0,8 (0,04)	19 (1,2)
Padronização	SVM_Sig	8 (4,8)	0,8 (0,02)	16 (0,5)
Padronização	RF	709 (447,0)	15,2 (5,21)	418 (256,5)
Padronização	XGBoost	56 (87,5)	3,5 (4,67)	56 (30,2)

Fonte: elaborada pelo autor.

significativamente diferentes. Logo então, aplicou-se o teste de Nemenyi para identificar quais pares possuem diferença significativa entre si. Na Figura 27 está apresentado o diagrama de CD obtido nesse teste. Pela figura, percebe-se que a combinação SMV_RBF sem pré-processamento teve o melhor rank médio. Essa combinação apresentou diferenças significativas apenas com as combinações DT com MinMax, DT sem pré-processamento, SMV_Pol com Padronização e DT com padronização.

Portanto, a combinação SMV_RBF sem pré-processamento obteve as melhores médias de 4 das 5 métricas de avaliação. Pelo teste de Nemenyi, o MAE desse par não é significativamente diferente do valor do MAE de outras 28 combinações. Entretanto, esse par obteve o melhor rank no diagrama de CD. Assim, considera-se o mais adequado para predição de DMO ao considerar todas as *features* do subconjunto de pacientes do sexo masculino.

Figura 13 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E1-M. (CD = 16,4)



Fonte: elaborada pelo autor.

Apesar do algoritmo SVM trabalhar com distâncias e por isso sofrer impacto com as diferenças de escala dos atributos, o par sem pré-processamento foi o que obteve rank e médias melhores. Uma possível causa dessa distinção é que algumas *features* de escalas maiores (como peso ou altura) podem ter maior importância que outras de escalas menores (como os atributos binários). Assim, o impacto da escala pode ter tido uma contribuição positiva para o desempenho do modelo.

5.2 Experimentos Adicionais

Esta seção foca na comparação entre as melhores abordagens encontradas em cada experimento. Inicialmente, comparam-se os resultados dos experimentos EX-A. Em seguida, avaliam-se os resultados dos experimentos EX-F, os quais são comparados com as melhores abordagens dos experimentos EX-A, mas testadas apenas nas amostras femininas. Logo então, analisam-se os resultados dos experimentos EX-M, os quais também são comparados com as melhores abordagens dos experimentos EX-A, mas testadas apenas nas amostras masculinas. Por fim, são feitas algumas conclusões sobre os experimentos em geral.

A melhor abordagem em cada experimento foi escolhida em relação ao valor de MAE. Mais especificamente, foi selecionada aquela que alcançou o melhor rank no teste de Nemenyi. Para facilitar a leitura, as tabelas contendo as métricas completas dos experimentos E2, E3 e E4 e os diagramas de CD desses mesmos experimentos não são apresentadas nesta seção. Todas essas tabelas e figuras encontram-se no Apêndice A.

Na Tabela 15 são apresentadas as métricas de avaliação para a melhor combinação

Tabela 15 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.

Experimento	Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
E1-A	MinMax	SVM_RBF	0,0774 (0,0077)	0,0974 (0,0095)	6,39 (0,66)	0,2842 (0,0475)	0,4794 (0,0643)
E2-A	Padronização	SVM_RBF	0,0780 (0,0074)	0,0979 (0,0090)	6,43 (0,62)	0,2787 (0,0335)	0,4743 (0,0639)
E3-A	MinMax	MLP	0,0771 (0,0070)	0,0975 (0,0104)	6,34 (0,61)	0,2913 (0,0628)	0,4787 (0,0742)
E4-A	Padronização	MLP	0,0778 (0,0065)	0,0981 (0,0083)	6,40 (0,56)	0,2923 (0,0550)	0,4716 (0,0603)

Fonte: elaborada pelo autor.

de cada um dos quatro experimentos que utilizam todas as amostras do *dataset* (EX-A). As melhores abordagens dos experimentos E1-A, E2-A e E3-A alcançaram o melhor resultado em pelo menos uma das métricas. O par SMV_RBF com MinMax (E1-A) obteve as melhores médias de RMSE e R². Já o SVM_RBF com Padronização (E2-A) teve o menor ME médio. Por fim, a MLP com MinMax (E3-A) atingiu as menores médias de MAE e MAPE. Ao aplicar o teste de Friedman com os resultados de MAE das quatro abordagens, obteve-se um valor-p de 0,92 (valor-p > 0.05). Dessa forma, falha-se em rejeitar a hipótese nula. Portanto, não encontram-se diferenças significativas entre os melhores modelos de cada um dos tipos de experimento.

Na Tabela 16 mostram-se as métricas de avaliação para a melhor abordagem de cada um dos quatro experimentos que utilizam apenas as amostras do sexo feminino (EX-F). Nessa mesma tabela, também são apresentados os resultados dos melhores modelos treinados nos experimentos EX-A, mas testando apenas nos pacientes equivalentes dos experimentos EX-F (sexo feminino). As melhores combinações dos experimentos E1-F e E4-F atingiram o melhor resultado em duas das métricas. O par RF com Padronização (E1-F) teve as menores médias de MAE e MAPE. Já a MLP com MinMax (E4-F) obteve as melhores médias de RMSE e R². O par SVM_RBF com Padronização (treinado no experimento E2-A, mas testado nas amostras do sexo feminino), obteve a média de ME mais baixa. Entretanto, ao aplicar o teste de Friedman nos resultados de MAE das oito combinações, encontrou-se um valor-p de 0,51 (valor-p > 0.05). Assim, falha-se em rejeitar a hipótese nula. Logo, não encontram-se diferenças significativas entre os melhores modelos de cada um dos tipos de experimento.

Por fim, a Tabela 17 apresenta as métricas de avaliação para a melhor combinação de cada um dos quatro experimentos que utilizam apenas as amostras do sexo masculino (EX-M). Nessa mesma tabela, também são apresentados os resultados dos melhores modelos treinados nos experimentos EX-A, mas testando apenas nos pacientes equivalentes dos experimentos EX-M (sexo masculino). A melhor abordagem do experimento E1-M alcançou as melhores médias em todas as métricas, exceto ME. A menor média de ME foi obtida pelo SVM_RBF com MinMax treinado no E1-A, mas testado apenas nas amostras do sexo masculino. Porém, ao aplicar o teste de Friedman com os resultados de MAE das oito

Tabela 16 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-F e de cada experimento EX-A, analisando apenas amostras do sexo feminino. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.

Experimento	Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
E1-F	Padronização	RF	0,0718 (0,0095)	0,0899 (0,0102)	6,24 (0,81)	0,2287 (0,0368)	0,2762 (0,0720)
E2-F	Padronização	XGBoost	0,0726 (0,0096)	0,0904 (0,0105)	6,30 (0,81)	0,2294 (0,0467)	0,2599 (0,1324)
E3-F	MinMax	RF	0,0727 (0,0110)	0,0909 (0,0120)	6,32 (0,97)	0,2356 (0,0458)	0,2611 (0,0833)
E4-F	MinMax	MLP	0,0720 (0,0093)	0,0896 (0,0116)	6,26 (0,84)	0,2279 (0,0490)	0,2801 (0,0917)
E1-A (F)	MinMax	SVM_RBF	0,0738 (0,0096)	0,0921 (0,0121)	6,42 (0,84)	0,2426 (0,0605)	0,2361 (0,1297)
E2-A (F)	Padronização	SVM_RBF	0,0739 (0,0092)	0,0914 (0,0099)	6,41 (0,83)	0,2232 (0,0320)	0,2513 (0,0480)
E3-A (F)	MinMax	MLP	0,0723 (0,0105)	0,0914 (0,0141)	6,27 (0,93)	0,2538 (0,0823)	0,2495 (0,1441)
E4-A (F)	Padronização	MLP	0,0723 (0,0094)	0,0902 (0,0121)	6,28 (0,88)	0,2292 (0,0574)	0,2709 (0,0955)

Fonte: elaborada pelo autor.

abordagens, obteve-se um valor-p de 0,33 (valor-p > 0.05). Portanto, falha-se em rejeitar a hipótese nula. Assim, não encontram-se diferenças significativas entre os melhores modelos de cada um dos tipos de experimento.

Tabela 17 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as melhores combinações de modelo e pré-processamento de cada experimento EX-M e de cada experimento EX-A, analisando apenas amostras do sexo masculino. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal.

Experimento	Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
E1-M	Nenhum	SVM_RBF	0,0814 (0,0100)	0,1014 (0,0131)	6,30 (0,85)	0,2574 (0,0526)	0,2496 (0,1857)
E2-M	Nenhum	SVM_Pol	0,0828 (0,0096)	0,1043 (0,0131)	6,43 (0,82)	0,2667 (0,0472)	0,2127 (0,1663)
E3-M	Padronização	EN	0,0831 (0,0110)	0,1045 (0,0144)	6,46 (0,94)	0,2657 (0,0552)	0,2054 (0,2000)
E4-M	Nenhum	SVM_RBF	0,0853 (0,0102)	0,1073 (0,0130)	6,62 (0,89)	0,2754 (0,0451)	0,1687 (0,1652)
E1-A (M)	MinMax	SVM_RBF	0,0819 (0,0112)	0,1026 (0,0152)	6,35 (0,95)	0,2522 (0,0501)	0,2376 (0,1823)
E2-A (M)	Padronização	SVM_RBF	0,0831 (0,0115)	0,1044 (0,0157)	6,45 (0,97)	0,2609 (0,0422)	0,2084 (0,2039)
E3-A (M)	MinMax	MLP	0,0830 (0,0090)	0,1037 (0,0126)	6,44 (0,79)	0,2678 (0,0505)	0,2065 (0,2199)
E4-A (M)	Padronização	MLP	0,0845 (0,0113)	0,1059 (0,0147)	6,55 (0,95)	0,2702 (0,0497)	0,1879 (0,1956)

Fonte: elaborada pelo autor.

Pelas tabelas 15, 16 e 17, percebe-se que em todos casos os experimentos do tipo E1 apresentam as melhores médias em pelo menos duas métricas. Mas ao mesmo tempo, os testes de Friedman indicaram que utilizar um subconjunto menor das *features* não causou diferenças significativas nos resultados de MAE. Isso sugere que os atributos básicos já detêm alto poder preditivo.

Analisando os resultados das tabelas 16 e 17, nota-se que os modelos especializados, aqueles treinados em apenas amostras de um sexo específico (EX-F e EX-M), tiveram mais médias de métricas próximas ao ideal do que os modelos treinados em ambos sexos, mas testados apenas em um (EX-A (F) e EX-A (M)). Mesmo assim, os testes de Friedman não encontraram diferenças significativas entre as possíveis abordagens. Isso indica que modelos treinados em todo o conjunto podem se sair tão bem quanto os modelos especializados. Ao treinar com ambos os sexos, os modelos possuem a vantagem de possuir mais amostras disponíveis para o aprendizado. Em contrapartida, os modelos especializados podem ter vantagens ao estarem modelando funções menos complexas.

Ao comparar os resultados obtidos utilizando apenas o grupo de amostras do sexo feminino com as do sexo masculino, verifica-se que todas as métricas dos experimentos EX-F alcançaram médias melhores que dos experimentos EX-M. Alguns fatores podem ter contribuído para essa diferença:

- O conjunto do sexo feminino possui mais amostras que o do masculino (500 contra 405, respectivamente). Essa discrepância pode ter facilitado o treinamento dos modelos nos experimentos EX-F comparado aos do EX-M, já que se pode aprender padrões de mais exemplos.
- Os pacientes do sexo masculino tendem a ter valores de DMO mais altos e têm uma faixa de valores de DMO maior. Logo, alvos mais altos tendem a gerar erros maiores e uma maior variância do alvo pode ser mais difícil de modelar.
- Os atributos utilizados podem ter maior poder preditivo de DMO para pessoas do sexo feminino. Assim, as *features* podem explicar a variação de DMO melhor em pessoas do sexo feminino do que masculino.

5.3 Interpretação

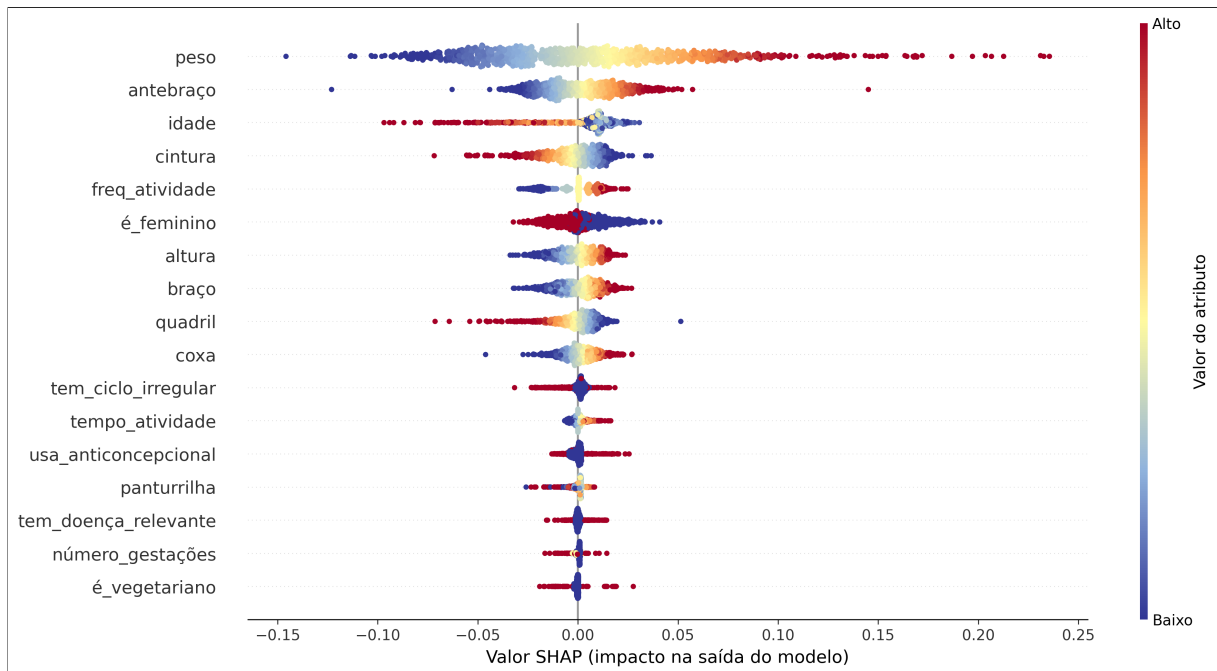
Esta seção apresenta as análises de interpretação feitas para as melhores abordagens dos experimentos E1. Inicialmente, discute-se sobre os impactos gerais das *features* do experimento E1-A. Em seguida, são avaliados alguns casos relevantes de interação de *features* nesse mesmo experimento. Logo então, apresentam-se os comportamentos dos valores SHAP tanto para o experimento E1-F como para o E1-M. Por fim, realizam-se algumas conclusões sobre as interpretações feitas.

A Figura 14 apresenta a influência dos valores dos atributos no comportamento geral dos valores SHAP, referentes à melhor combinação de pré-processamento e modelo obtida no experimento E1-A. Percebe-se que a *feature* peso possui o maior impacto na saída do SVM_RBF, seguida por circunferência de antebraço, idade, circunferência de cintura, frequência de atividade e sexo designado no nascimento. Os atributos com menor impacto na predição da DMO foram vegetarianismo, número de gestações e presença de doença relevante.

Ainda pela Figura 14, verifica-se que há um impacto positivo na DMO predita quando os pacientes são do sexo masculino ou quando apresentam valores altos de peso, circunferência de antebraço, frequência de atividade, altura, circunferência de braço, circunferência de coxa e tempo de atividade. Por outro lado, ser do sexo feminino ou apresentar valores altos de idade, circunferência de cintura e circunferência de quadril impacta negativamente na saída do modelo. No geral, os outros atributos tiveram impactos divergentes.

A associação positiva entre peso e DMO é esperada e está de acordo com o

Figura 14 – Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-A).



Fonte: elaborada pelo autor.

encontrado em outros estudos (MEYBODI et al., 2011). Já a circunferência de antebraço pode estar relacionada com a DMO por ser um indicador de massa corporal magra. Pesquisadores identificaram uma alta correlação entre circunferência de antebraço e massa magra (CHIEN et al., 2020). A massa magra por sua vez, tem uma correlação positiva com DMO (HO-PHAM; NGUYEN; NGUYEN, 2014).

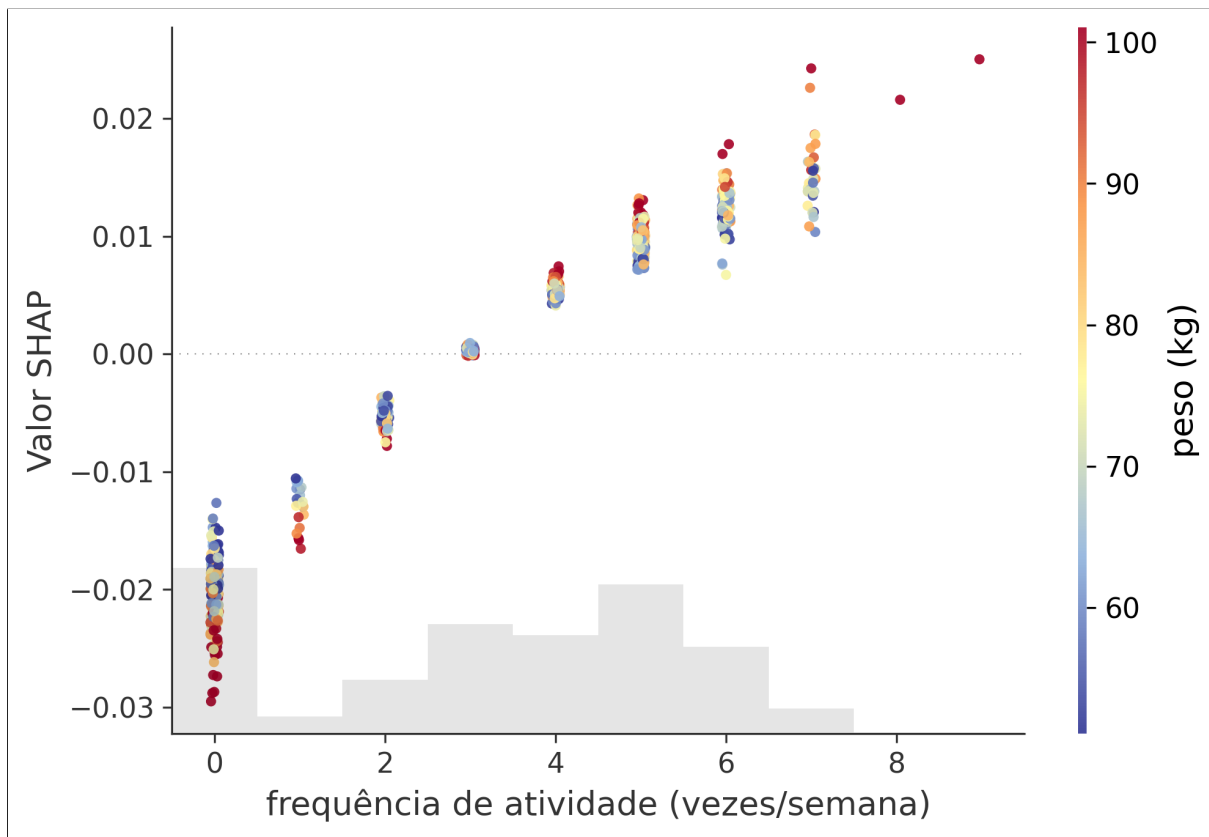
Ao contrário do antebraço, os valores altos de circunferência de cintura e de quadril impactam negativamente no valor de DMO. Esse comportamento pode ser causado pelo fato que as circunferências de cintura e quadril possuem uma correlação positiva com o percentual de gordura corporal (FREEDMAN et al., 2012). Além disso, estudos anteriores encontraram uma associação negativa entre percentual de gordura corporal e DMO (LIU et al., 2013), o que poderia explicar o impacto negativo da cintura e quadril na saída do modelo.

A frequência e duração de atividades físicas apresentaram um impacto positivo na DMO. Esse relacionamento é coerente com pesquisas que encontraram uma associação positiva entre DMO e exercícios físicos (RONDANELLI et al., 2021), e outras que apresentam uma correlação negativa entre atividade física e gordura corporal (BRADBURY et al., 2017).

A Figura 15 explora melhor as associações mencionadas. Essa figura mostra o impacto na DMO de acordo com a relação entre frequência de atividade e peso do paciente.

No grupo de pessoas que não praticam atividade física, valores altos de peso diminuem a saída do modelo mais do que valores baixos de peso. Quando a frequência de atividade é maior que três vezes por semana, o comportamento se inverte: pesos maiores aumentam a saída do modelo mais do que pesos menores. Essa interação é coerente com as associações apresentadas anteriormente entre massa corporal e DMO. A razão entre gordura corporal e massa magra provavelmente é maior em pessoas com peso alto que não praticam atividades do que em pessoas com o peso alto que praticam atividades com maior frequência.

Figura 15 – Comportamento dos valores SHAP de acordo com a relação entre frequência de atividade e peso (referentes à abordagem selecionada no experimento E1-A).

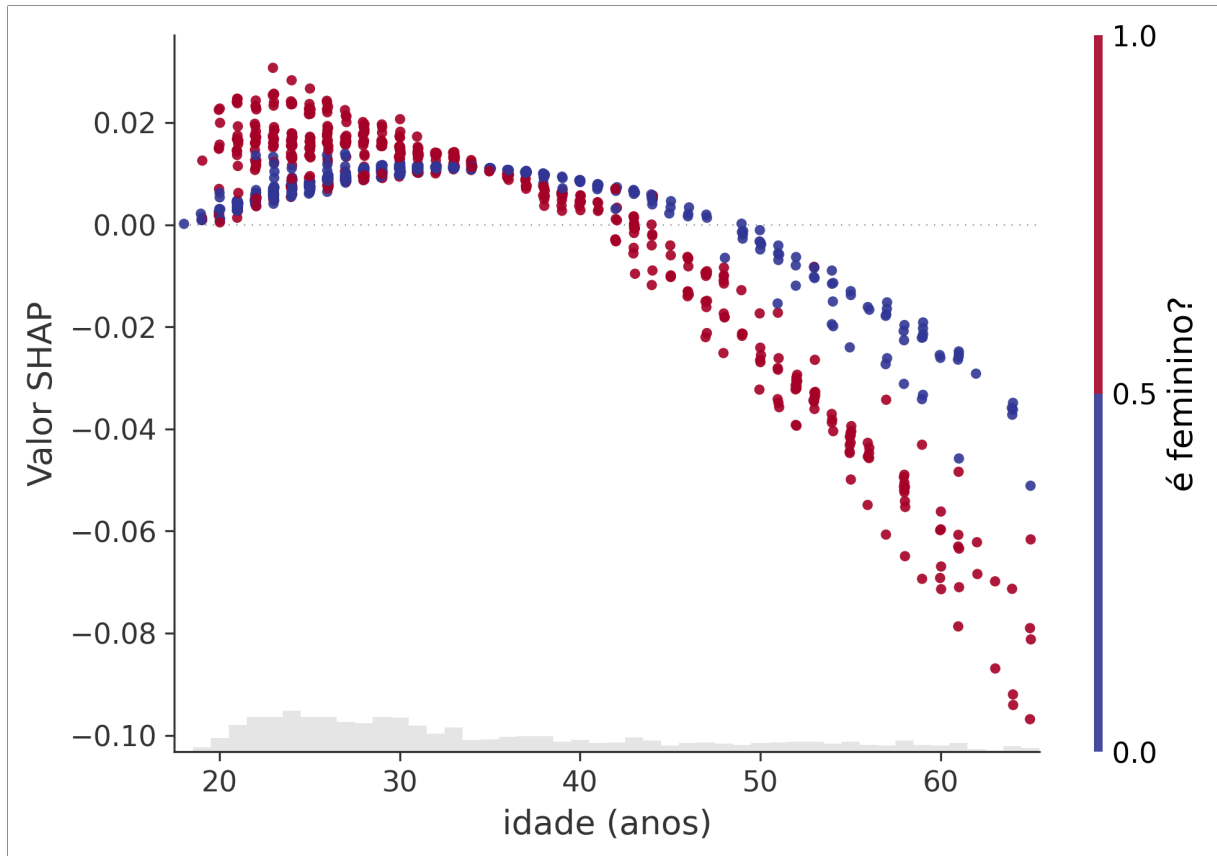


Fonte: elaborada pelo autor.

Outras *features* que apresentaram alto impacto na saída do modelo são a idade e o sexo designado no nascimento. Na Figura 16, explora-se os efeitos da interação entre esses dois atributos. Percebe-se que a idade tem um impacto diferente na saída dependendo do sexo. Ao comparar os pacientes mais jovens, verifica-se que os do sexo feminino (pontos vermelhos) têm um impacto positivo na DMO maior que os do sexo masculino (pontos azuis). O pico desse valor SHAP acontece mais cedo para os pacientes femininos. Entretanto, a partir de aproximadamente 40 anos, o avanço da idade faz com que os impactos decresçam e que a idade seja um fator que impacta a saída mais negativamente nas mulheres do que nos homens. Esse comportamento é consistente com estudos que mostram que homens atingem o pico de DMO mais tarde do que as mulheres (AVDAGIĆ

et al., 2009; National Institutes of Health, 2018b) e com estudos que indicam que a DMO da mulher cai consideravelmente após a menopausa (JI; YU, 2015).

Figura 16 – Comportamento dos valores SHAP de acordo com a relação entre idade e sexo de nascimento (referentes à abordagem selecionada no experimento E1-A).



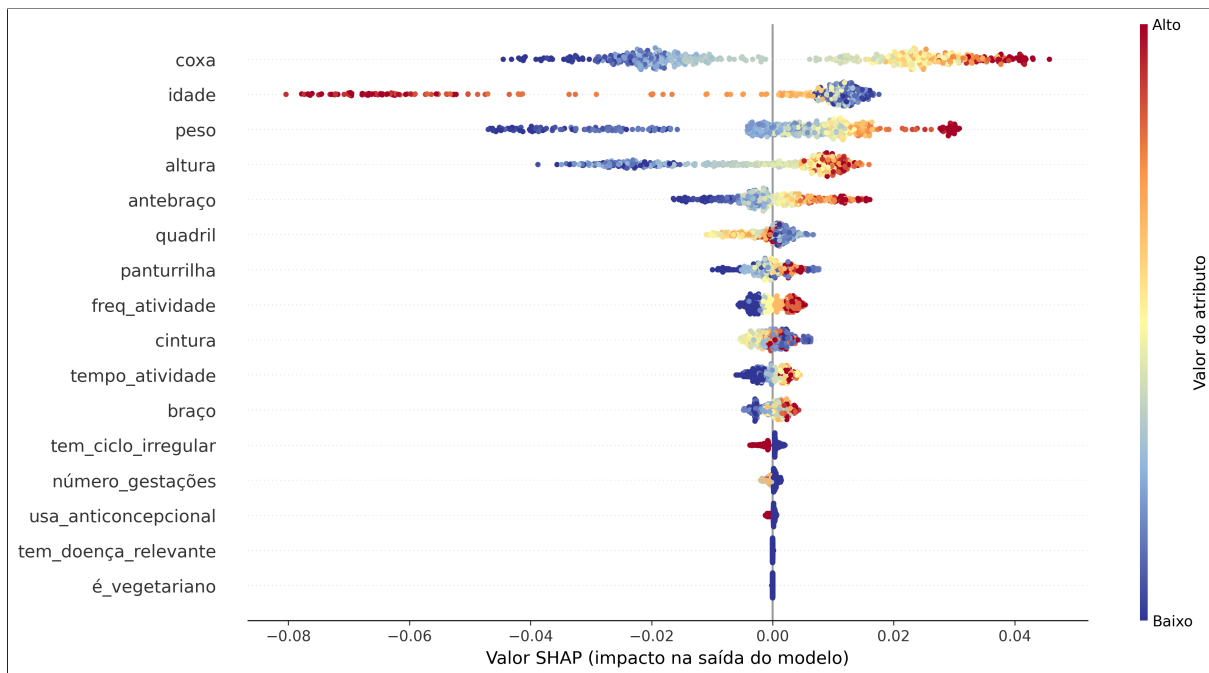
Fonte: elaborada pelo autor.

Além das análises feitas no experimento E1-A, realizam-se análises interpretativas mais breves dos experimentos E1-F e E1-M. As figuras 17 e 18 apresentam a influência dos valores dos atributos no comportamento geral dos valores SHAP, referentes à melhor combinação de pré-processamento e modelo obtida no experimento E1-F e no experimento E1-M, respectivamente. Percebe-se que, em ambos os gráficos, o sentido do impacto da maioria dos atributos é semelhante ao do experimento E1-A. O que diferem-se mais são a magnitude dos impactos e a ordem das *features* em relação ao valor SHAP. Verifica-se que o peso e a circunferência de antebraço continuam sendo algumas das variáveis mais relevantes. Entretanto, a circunferência de coxa, que no E1-A não esteve entre os atributos mais influentes, aparece como o mais importante no E1-F e o quarto mais importante no E1-M.

Uma diferença clara entre os gráficos das figuras 17 e 18 é o comportamento dos valores de SHAP referentes à variação da idade. Esse atributo no E1-F tem a segunda maior importância, provavelmente ligado ao fato da idade ser crucial para definir o início

da menopausa. Já para o E1-M, a idade está entre as *features* de menos importância e tem um comportamento diferente. Para o modelo treinado apenas com pacientes masculinos, a idade mais avançada causa um pequeno aumento na DMO predita. Isso pode estar ligado ao fato de homens atingirem o pico de massa óssea mais tarde, como mencionado anteriormente, e pela pesquisa se limitar a pacientes com até 65 anos, onde os efeitos da idade na DMO dos homens não são tão perceptíveis quanto nas mulheres.

Figura 17 – Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-F).



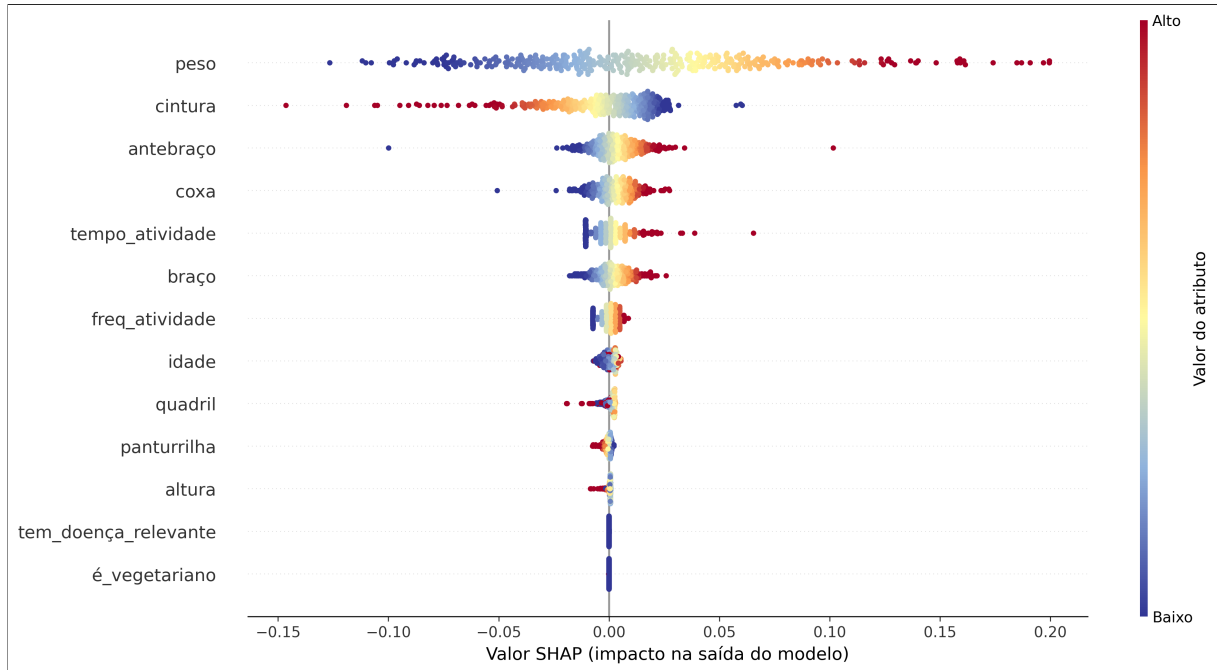
Fonte: elaborada pelo autor.

Portanto, a técnica SHAP permitiu quantificar a importância dos atributos para a predição de DMO e identificar como a variação de seus valores e suas interações impactam a saída dos modelos. Essas interpretações sugerem que os modelos conseguiram aprender padrões nos dados que condizem com associações encontradas na literatura, tais como:

- Pessoas do sexo feminino tendem a ter uma DMO inferior àquelas do sexo masculino;
- Após a menopausa, a DMO decresce com maior velocidade;
- Pessoas mais leves tendem a ter menores valores de DMO;
- Exercícios físicos ajudam a ter uma DMO maior.

Além disso, a interpretação pode destacar comportamentos menos conhecidos e ajudar no entendimento científico. Por exemplo, é comumente encontrado na literatura que o peso é associado positivamente com DMO. Porém, é menos comentado e estudado o fato que, na composição desse peso, a massa magra contribui positivamente para a DMO, enquanto

Figura 18 – Comportamento dos valores SHAP de acordo com os valores dos atributos (referentes à abordagem selecionada no experimento E1-M).



Fonte: elaborada pelo autor.

a gordura corporal tende a ter um efeito contrário. Esse comportamento mencionado foi sugerido pela interpretação do modelo. Por último, alguns atributos não tiveram grande impacto na DMO predita, tais como o vegetarianismo e a presença de doenças relevantes. Essa pouca influência na saída do modelo pode ser causada pela falta de relação dessas *features* com a DMO, logo elas não carregariam informação tão relevante para a densidade óssea como as outras *features* carregam. Outra causa plausível é a pequena presença de casos positivos para esses atributos. Como visto na Figura 3, apenas cerca de 6% das pessoas no *dataset* são vegetarianos e cerca de 10% apresentam pelo menos uma das doenças consideradas relevantes. Essa amostra pequena desses casos pode ter dificultado o reconhecimento de padrões nesses atributos, tornando seus impactos irrisórios.

6 Conclusão

A DMO é uma medida importante para a identificação e prevenção da osteoporose e da osteopenia. Porém, seu valor é obtido através do exame DXA, que pode ser inacessível a muitas pessoas, devido ao seu custo ou à sua escassez no sistema de saúde. Em vista disso, este trabalho propôs uma abordagem de ML para a predição de DMO, utilizando atributos de fácil obtenção: dados de identificação, medidas antropométricas e dados de anamnese.

Adicionalmente, esta pesquisa avaliou diferentes algoritmos de regressão e técnicas de pré-processamento. Percebeu-se que os métodos SVM_RBF, MLP e RF se destacaram, alcançando as melhores métricas na maioria dos experimentos. Além disso, verificou-se que técnicas de pré-processamento também contribuíram para esse bom desempenho, na maioria dos casos. Por exemplo, no experimento que utiliza o *dataset* completo, o método SVM_RBF com a técnica MinMax obteve MAE de 0,0774 g/cm², R² de 0,4794 e MAPE de apenas 6,39%. Esses resultados indicam que a abordagem proposta se mostra capaz de prever DMO com pequenos erros. Ainda assim, os testes estatísticos apontaram que muitos outros métodos não tiveram resultados significativamente diferentes, então também poderiam ser utilizados no problema apresentado.

Não apenas foram avaliadas diferentes técnicas de ML, como também foram investigados os impactos de se utilizar apenas parte dos atributos e de realizar treinamentos com todas amostras ou apenas as de um sexo específico. Nesse primeiro ponto, verificou-se que o uso de todos atributos resultou em métricas médias mais próximas do ideal. Porém, esses resultados não foram significativamente diferentes daqueles dos experimentos que utilizaram apenas a idade, sexo, altura e peso do paciente. Assim, mesmo que o profissional da saúde não tenha acesso aos dados de anamnese do paciente e não consiga fazer as medições de circunferência, os outros dados mais básicos são suficientes para obter predições de qualidade equivalente. Já em relação ao segundo ponto, as abordagens treinadas especificamente em pessoas de um sexo apresentaram desempenho semelhante às que foram treinadas com ambos sexos. Dessa forma, não há necessidade do uso de modelos especializados em apenas um sexo.

Além das análises de desempenho das abordagens, também foram realizadas interpretações dos impactos das *features* na saída dos modelos. Essas explicações permitiram constatar que grande parte do comportamento do modelo condiz com relações já detectadas por outros pesquisadores. Exemplos são a associação positiva entre peso e DMO e o impacto negativo na DMO com o avanço da idade nas mulheres. Tais interpretações permitiram averiguar que o comportamento do modelo é coerente e, portanto, traz maior

confiança para os profissionais de saúde. Além disso, a interpretabilidade revela comportamentos menos explorados na literatura, que podem servir de *insight* para investigação de pesquisadores da área.

Outra contribuição da pesquisa é a apresentação de um novo *dataset*, contendo indivíduos de 18 a 65 anos, de ambos sexos. O *dataset* possui 905 pacientes, com seus respectivos dados de sexo, idade, medidas antropométricas, anamnese e valor de DMO do corpo inteiro. A introdução desse conjunto de dados ataca problemas encontrados na literatura, como uso de atributos difíceis de se obter, restrição rigorosa dos participantes do *dataset* e indisponibilidade desses dados. Portanto, este trabalho fornece uma abordagem mais acessível e abrangente. Além disso, como as técnicas propostas têm um baixo custo computacional na predição e não necessitam de um grande espaço de armazenamento, as mesmas poderiam ser embarcadas em aplicações móveis. Isso facilitaria o uso dessa ferramenta em comunidades afastadas de centros urbanos, que podem não ter acesso à internet, computadores, ou até à energia elétrica.

Ressalta-se que a abordagem proposta ainda não foi aplicada e avaliada em um ambiente clínico. Portanto, ela não substitui os exames clínicos recomendados. Entretanto, esse método pode se tornar uma ferramenta para complementar o exame médico e auxiliar na decisão do profissional. Uma predição baixa de DMO pode sugerir que o paciente necessita de um exame oficial para confirmar o valor. Além disso, uma predição baixa poderia motivar o profissional de saúde a sugerir estilos de vida para a prevenção da osteoporose, tais como mudanças na dieta e exercício físico. Finalmente, uma predição alta de DMO, combinada com a análise de outros fatores feita pelo especialista, poderia indicar que o paciente não precisa realizar um DXA naquele momento, o que reduziria os custos e evitaria a exposição à radiação.

6.1 Limitações

Apesar do conjunto de dados usado ter uma abrangência demográfica maior que a maioria dos estudos encontrados, ainda há limitações quanto aos critérios de inclusão. O *dataset* não contém crianças e nem idosos, portanto os métodos utilizados podem apresentar resultados diferentes para essas faixas etárias. Além disso, pessoas com doenças debilitantes ou que não podiam chegar ao local de coleta não foram incluídas na pesquisa. Essa restrição retira pessoas acamadas, que poderiam se beneficiar da técnica. Por último, pessoas que sofreram amputação não participaram da coleta, o que exclui uma parcela da população e deixa em aberto como deve ser feito o processamento dos atributos de circunferência em casos de membros amputados.

Outra limitação do *dataset* encontra-se na variável alvo. Como o conjunto de dados desta pesquisa contém apenas a informação de DMO do corpo inteiro, não consegue-

se utilizar essa informação para identificar a presença de osteoporose em uma região específica, como a coluna lombar, a bacia e o fêmur. Tais regiões costumam ser o foco das diretrizes para avaliação de osteoporose (BORRUD et al., 2012). Apesar disso, a DMO do corpo inteiro tem alta associação com DMO de regiões específicas (BAGUR; VEGA; MAUTALEN, 1995; SAKAI et al., 2010). Logo, o valor de DMO utilizado também é uma informação valiosa sobre a saúde óssea do paciente.

Já uma limitação presente na abordagem está na interpretabilidade. Ressalta-se que o método SHAP analisa o impacto dos atributos na saída do modelo. Portanto, se o modelo não generalizasse bem os dados, os impactos encontrados com o SHAP não refletiriam bem a realidade do problema. Além disso, como muitos atributos do *dataset* são altamente correlacionados, o modelo pode concentrar a importância dessas *features* em apenas uma delas, ou pode até distribuir o impacto delas. Esse comportamento afeta diretamente nos valores SHAP, podendo fazer com que atributos tenham uma relevância diferente do esperado. Por fim, salienta-se que as associações encontradas com o SHAP não implicam causalidade. Por exemplo, por mais que as interpretações tenham identificado uma relação positiva entre peso e DMO, não se pode afirmar que o valor de DMO de uma pessoa irá aumentar se ela aumentar seu peso.

6.2 Trabalhos Futuros

Planeja-se que haja continuidade para esta pesquisa e que ela resulte em novos trabalhos no futuro. Um dos planos é aplicar a mesma metodologia para a predição de DMO de regiões específicas (lombar, bacia e fêmur), que costumam ser as regiões de interesse para o diagnóstico de osteoporose. Além disso, pretende-se afrouxar os critérios de seleção e aumentar a quantidade de pacientes no *dataset*. Também deseja-se incluir novos atributos que possam ajudar a estimar a DMO, tais como cor, histórico de fraturas, histórico familiar de fraturas e osteoporose, uso de medicamentos, consumo de álcool e uso de tabaco (National Institutes of Health, 2019). Adicionalmente, pretende-se investigar estratégias para melhorar a capacidade de generalização do modelo para pacientes com valores de DMO mais extremos, especialmente aqueles com valores mais baixos, por serem casos mais críticos.

Outro objetivo futuro é aplicar o método em um ambiente clínico, como uma ferramenta de auxílio ao profissional da saúde. Dessa forma, poderia-se avaliar o desempenho do método em uma aplicação real, analisar a aceitação por parte dos usuários e receber retorno sobre possíveis melhorias para a abordagem. Por último, planeja-se avaliar o uso de fotos dos pacientes para a predição de DMO. Há pesquisas recentes que estimam as circunferências corporais através de imagens e usam esses valores para predizer o percentual de gordura corporal (LOUREIRO et al., 2022). Uma abordagem semelhante poderia ser

testada ou poderia-se até utilizar outras técnicas de aprendizado profundo para imagens.

Referências

ABRAHAMSEN, B. et al. Ten-year prediction of osteoporosis from baseline bone mineral density: development of prognostic thresholds in healthy postmenopausal women. the danish osteoporosis prevention study. *Osteoporosis international*, Springer, v. 17, n. 2, p. 245–251, 2006. Citado na página 25.

ARRIETA, A. B. et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, Elsevier, v. 58, p. 82–115, 2020. Citado na página 41.

AVDAGIĆ, S. et al. Differences in peak bone density between male and female students. *Archives of Industrial Hygiene and Toxicology*, Walter de Gruyter GmbH, v. 60, n. 1, p. 79–86, mar. 2009. Disponível em: <<https://doi.org/10.2478/10004-1254-60-2009-1886>>. Citado na página 89.

AWAD, M. et al. Support vector regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, Springer, p. 67–80, 2015. Citado na página 37.

AZIZIYEH, R. et al. A scorecard for osteoporosis in four latin american countries: Brazil, mexico, colombia, and argentina. *Archives of Osteoporosis*, Springer Science and Business Media LLC, v. 14, n. 1, p. 1–10, jun. 2019. Disponível em: <<https://doi.org/10.1007/s11657-019-0622-1>>. Citado na página 25.

BAGUR, A.; VEGA, E.; MAUTALEN, C. Discrimination of total body bone mineral density measured by dxa in vertebral osteoporosis. *Calcified tissue international*, Springer, v. 56, p. 263–267, 1995. Citado na página 95.

BENKE, K.; BENKE, G. Artificial intelligence and big data in public health. *International Journal of Environmental Research and Public Health*, MDPI AG, v. 15, n. 12, p. 2796, dez. 2018. Disponível em: <<https://doi.org/10.3390/ijerph15122796>>. Citado na página 26.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of machine learning research*, v. 13, n. 2, 2012. Citado na página 59.

BISHOP, C. M.; NASRABADI, N. M. *Pattern recognition and machine learning*. [S.l.]: Springer, 2006. v. 4. Citado 3 vezes nas páginas 33, 35 e 37.

BONACCORSO, G. *Machine learning algorithms*. [S.l.]: Packt Publishing Ltd, 2017. Citado na página 34.

BORRUD, L. G. et al. Osteoporosis or low bone mass at the femur neck or lumbar spine in older adults, united states, 2005-2008. National Center for Health Statistics, 2012. Citado na página 95.

BRADBURY, K. E. et al. Association between physical activity and body fat percentage, with adjustment for bmi: a large cross-sectional analysis of uk biobank. *BMJ open*, British Medical Journal Publishing Group, v. 7, n. 3, p. e011843, 2017. Citado na página 87.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 39.

BREIMAN, L. et al. *Classification and regression trees*. [S.l.]: Routledge, 2017. Citado na página 38.

CABITZA, F.; BANFI, G. Machine learning in laboratory medicine: waiting for the flood? *Clinical Chemistry and Laboratory Medicine (CCLM)*, De Gruyter, v. 56, n. 4, p. 516–524, 2018. Citado na página 26.

CAWLEY, G. C.; TALBOT, N. L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, JMLR. org, v. 11, p. 2079–2107, 2010. Citado na página 58.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794. Citado na página 40.

CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, PeerJ Inc., v. 7, p. e623, 2021. Citado na página 65.

CHIEN, K.-Y. et al. A community-based approach to lean body mass and appendicular skeletal muscle mass prediction using body circumferences in community-dwelling elderly in taiwan. *Asia Pacific Journal of Clinical Nutrition*, v. 29, n. 1, p. 94–100, 2020. Citado na página 87.

CHUANG, C.-C. et al. Robust support vector regression networks for function approximation with outliers. *IEEE Transactions on Neural Networks*, IEEE, v. 13, n. 6, p. 1322–1330, 2002. Citado na página 38.

CNES, M. da Saúde Cadastro Nacional dos Estabelecimentos de Saúde do B. *TabNet Win32 3.0: CNES - Recursos Físicos - Equipamentos - Brasil*. 2022. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?cnes/cnv/equipobr.def>>. Citado na página 26.

CONOVER, W. J. *Practical nonparametric statistics*. [S.l.]: john wiley & sons, 1999. v. 350. Citado 2 vezes nas páginas 66 e 67.

CUNNINGHAM, P.; DELANY, S. J. K-nearest neighbour classifiers-a tutorial. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 54, n. 6, p. 1–25, 2021. Citado na página 36.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, JMLR. org, v. 7, p. 1–30, 2006. Citado na página 67.

DEO, R. C. Machine learning in medicine. *Circulation*, Ovid Technologies (Wolters Kluwer Health), v. 132, n. 20, p. 1920–1930, nov. 2015. Disponível em: <<https://doi.org/10.1161/circulationaha.115.001593>>. Citado na página 26.

FREEDMAN, D. S. et al. The body adiposity index (hip circumference ÷ heights^{sup1.5}/sup) is not a more accurate measure of adiposity than is BMI, waist circumference, or hip circumference. *Obesity*, Wiley, v. 20, n. 12, p. 2438–2444, dez. 2012. Disponível em: <<https://doi.org/10.1038/oby.2012.81>>. Citado na página 87.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, NIH Public Access, v. 33, n. 1, p. 1, 2010. Citado na página 35.

FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, Taylor & Francis, v. 32, n. 200, p. 675–701, 1937. Citado na página 66.

GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. [S.l.]: "O'Reilly Media, Inc.", 2022. Citado 4 vezes nas páginas 37, 38, 39 e 40.

HASANI-RANJBAR, S. et al. Association of osteoporosis with anthropometric measures in a representative sample of iranian adults: The iranian multicenter osteoporosis study. *International Journal of Preventive Medicine*, Wolters Kluwer–Medknow Publications, v. 10, 2019. Citado na página 27.

HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado 3 vezes nas páginas 34, 36 e 39.

HAYKIN, S. *Redes neurais: princípios e prática*. 2. ed. Porto Alegre: Bookman, 2001. Citado 2 vezes nas páginas 36 e 37.

Healthdirect. *Osteoporosis*. 2020. Disponível em: <<https://www.healthdirect.gov.au/osteoporosis>>. Citado na página 46.

HIEN, V. T. T. et al. Determining the prevalence of osteoporosis and related factors using quantitative ultrasound in vietnamese adult women. *American Journal of Epidemiology*, Oxford University Press (OUP), v. 161, n. 9, p. 824–830, maio 2005. Disponível em: <<https://doi.org/10.1093/aje/kwi105>>. Citado na página 26.

HO, C.-S. et al. Application of deep learning neural network in predicting bone mineral density from plain x-ray radiography. *Archives of Osteoporosis*, Springer, v. 16, n. 1, p. 1–12, 2021. Citado 2 vezes nas páginas 27 e 28.

HO-PHAM, L. T.; NGUYEN, U. D. T.; NGUYEN, T. V. Association between lean mass, fat mass, and bone mineral density: A meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, The Endocrine Society, v. 99, n. 1, p. 30–38, jan. 2014. Disponível em: <<https://doi.org/10.1210/jc.2013-3190>>. Citado na página 87.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Taylor & Francis, v. 12, n. 1, p. 55–67, 1970. Citado na página 35.

HOLZINGER, A. et al. Explainable ai methods-a brief overview. In: SPRINGER. *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. [S.l.], 2022. p. 13–38. Citado na página 42.

ILIOU, T.; ANAGNOSTOPOULOS, C.-N.; ANASTASSOPOULOS, G. Osteoporosis detection using machine learning techniques and feature selection. *International Journal on Artificial Intelligence Tools*, World Scientific Pub Co Pte Lt, v. 23, n. 05, p. 1450014, out. 2014. Disponível em: <<https://doi.org/10.1142/s0218213014500146>>. Citado 2 vezes nas páginas 27 e 28.

- JI, M.-X.; YU, Q. Primary osteoporosis in postmenopausal women. *Chronic Diseases and Translational Medicine*, Wiley, v. 1, n. 1, p. 9–13, mar. 2015. Disponível em: <<https://doi.org/10.1016/j.cdtm.2015.02.006>>. Citado na página 89.
- KANIS, J. A. et al. Assessment of osteoporosis at the primary health-care level. *WHO Collaborating Centre for Metabolic Bone Diseases, University of Sheffield*, 2007. Citado na página 25.
- KARAGUZEL, G.; HOLICK, M. F. Diagnosis and treatment of osteopenia. *Reviews in endocrine and metabolic disorders*, Springer, v. 11, n. 4, p. 237–251, 2010. Citado na página 25.
- KILIC, N.; HOSGORMEZ, E. Automatic estimation of osteoporotic fracture cases by using ensemble learning approaches. *Journal of Medical Systems*, Springer Science and Business Media LLC, v. 40, n. 3, dez. 2015. Disponível em: <<https://doi.org/10.1007/s10916-015-0413-1>>. Citado 2 vezes nas páginas 27 e 29.
- KIM, S. K. et al. Osteoporosis risk prediction using machine learning and conventional methods. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013. Disponível em: <<https://doi.org/10.1109/embc.2013.6609469>>. Citado 2 vezes nas páginas 27 e 28.
- KLING, J. M.; CLARKE, B. L.; SANDHU, N. P. Osteoporosis prevention, screening, and treatment: a review. *Journal of women's health*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 23, n. 7, p. 563–572, 2014. Citado na página 25.
- LAURENT, M. R. et al. Age-related bone loss and sarcopenia in men. *Maturitas*, Elsevier, v. 122, p. 51–56, 2019. Citado na página 27.
- LIU, Y. hua et al. Association of weight-adjusted body fat and fat distribution with bone mineral density in middle-aged chinese adults: A cross-sectional study. *PLoS ONE*, Public Library of Science (PLOS), v. 8, n. 5, p. e63339, maio 2013. Disponível em: <<https://doi.org/10.1371/journal.pone.0063339>>. Citado na página 87.
- LOUREIRO, L. L. et al. Development of a method for evaluating body composition and perimeters using cell phone photos. *Medicine & Science in Sports & Exercise*, LWW, v. 54, n. 9S, p. 368–369, 2022. Citado na página 95.
- LUNDBERG, S. *Welcome to the shap documentation*. 2020. Disponível em: <<https://shap.readthedocs.io/en/latest/>>. Citado na página 68.
- LUNDBERG, S. M. et al. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, Nature Publishing Group UK London, v. 2, n. 1, p. 56–67, 2020. Citado na página 68.
- LUNDBERG, S. M.; ERION, G. G.; LEE, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018. Citado na página 42.
- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, v. 30, 2017. Citado 2 vezes nas páginas 41 e 42.

- MARKUS, A. F.; KORS, J. A.; RIJNBEEK, P. R. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, Elsevier, v. 113, p. 103655, jan. 2021. Disponível em: <<https://doi.org/10.1016/j.jbi.2020.103655>>. Citado na página 41.
- MEYBODI, H. A. et al. Association between anthropometric measures and bone mineral density: population-based study. *Iranian Journal of Public Health*, Tehran University of Medical Sciences, v. 40, n. 2, p. 18, 2011. Citado na página 87.
- MILLER, P. D. Underdiagnoses and undertreatment of osteoporosis: the battle to be won. *The Journal of Clinical Endocrinology & Metabolism*, Oxford University Press, v. 101, n. 3, p. 852–859, 2016. Citado na página 25.
- MOLNAR, C. *Interpretable Machine Learning: A guide for making black box models explainable*. 2. ed. [s.n.], 2022. Disponível em: <<https://christophm.github.io/interpretable-ml-book>>. Citado na página 41.
- MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, v. 24, n. 3, p. 69–71, 2012. Citado na página 50.
- MYSZCZYNSKA, M. A. et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, Nature Publishing Group, v. 16, n. 8, p. 440–456, 2020. Citado na página 26.
- National Institutes of Health. *Bone mass measurement: What the numbers mean*. 2018. Disponível em: <<https://www.bones.nih.gov/health-info/bone/bone-health/bone-mass-measure>>. Citado na página 25.
- National Institutes of Health. *Osteoporosis: Peak Bone Mass in Women*. 2018. Disponível em: <<https://www.bones.nih.gov/health-info/bone/osteoporosis/bone-mass>>. Citado na página 89.
- National Institutes of Health. *What people with celiac disease need to know about osteoporosis*. 2018. Disponível em: <<https://www.bones.nih.gov/health-info/bone/osteoporosis/conditions-behaviors/celiac>>. Citado na página 46.
- National Institutes of Health. *Osteoporosis overview*. 2019. Disponível em: <<https://www.bones.nih.gov/health-info/bone/osteoporosis/overview>>. Citado 3 vezes nas páginas 27, 46 e 95.
- NAYAK, S.; ROBERTS, M. S.; GREENSPAN, S. L. Cost-effectiveness of different screening strategies for osteoporosis in postmenopausal women. *Annals of Internal Medicine*, American College of Physicians, v. 155, n. 11, p. 751, dez. 2011. Disponível em: <<https://doi.org/10.7326/0003-4819-155-11-201112060-00007>>. Citado na página 26.
- NEMENYI, P. B. *Distribution-free multiple comparisons*. [S.l.]: Princeton University, 1963. Citado na página 67.
- Organização Mundial da Saúde. *Prevention and management of osteoporosis: report of a WHO scientific group*. [S.l.]: Organização Mundial da Saúde, 2003. Citado 2 vezes nas páginas 25 e 26.

RADOMINSKI, S. C. et al. Brazilian guidelines for the diagnosis and treatment of postmenopausal osteoporosis. *Revista brasileira de reumatologia*, SciELO Brasil, v. 57, p. s452–s466, 2017. Citado na página 25.

RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. "why should i trust you?"explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 1135–1144. Citado na página 41.

RONDANELLI, M. et al. Nutrition, physical activity, and dietary supplementation to prevent bone mineral density loss: A food pyramid. *Nutrients*, MDPI, v. 14, n. 1, p. 74, 2021. Citado na página 87.

RUDIN, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, Nature Publishing Group, v. 1, n. 5, p. 206–215, 2019. Citado 2 vezes nas páginas 37 e 38.

SAKAI, Y. et al. Is whole body bone mineral density measured by the dual energy x-ray absorptiometry applied to evaluate risk of osteoporosis among japanese adult females? *Fukuoka Igaku Zasshi= Hukuoka Acta Medica*, v. 101, n. 4, p. 69–74, 2010. Citado na página 95.

SHESKIN, D. J. Parametric and nonparametric statistical procedures. *Boca Raton: CRC*, 2000. Citado na página 67.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and computing*, Springer, v. 14, n. 3, p. 199–222, 2004. Citado na página 37.

SONG, Y. et al. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, Elsevier, v. 251, p. 26–34, 2017. Citado na página 36.

SÖZEN, T.; ÖZİŞİK, L.; BAŞARAN, N. Ç. An overview and management of osteoporosis. *European journal of rheumatology*, AVES, v. 4, n. 1, p. 46, 2017. Citado na página 25.

SRIDHARAN, K. et al. Utility of anthropometric indicators in predicting osteoporosis in ambulant community dwelling rural postmenopausal women from southern india. *Tropical Doctor*, SAGE Publications Sage UK: London, England, v. 50, n. 3, p. 228–232, 2020. Citado na página 27.

STONE, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, Wiley Online Library, v. 36, n. 2, p. 111–133, 1974. Citado na página 57.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996. Citado na página 35.

WAINER, J.; CAWLEY, G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, Elsevier, v. 182, p. 115222, 2021. Citado na página 59.

YANG, W.-Y. O. et al. Development of machine learning models for prediction of osteoporosis from clinical health examination data. *International Journal of Environmental Research and Public Health*, MDPI AG, v. 18, n. 14, p. 7635, jul. 2021. Disponível em: <<https://doi.org/10.3390/ijerph18147635>>. Citado 2 vezes nas páginas 27 e 28.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005. Citado 2 vezes nas páginas 34 e 35.

Apêndices

APÊNDICE A – Resultados completos dos experimentos adicionais

Tabela 18 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0784 (0,0074)	0,0990 (0,0093)	6,47 (0,62)	0,2824 (0,0302)	0,4619 (0,0672)
Nenhum	EN	0,0782 (0,0074)	0,0988 (0,0091)	6,45 (0,62)	0,2790 (0,0310)	0,4641 (0,0666)
Nenhum	DT	0,0815 (0,0081)	0,1032 (0,0096)	6,71 (0,69)	0,2888 (0,0420)	0,4132 (0,0902)
Nenhum	MLP	0,0786 (0,0063)	0,0989 (0,0077)	6,47 (0,54)	0,2876 (0,0577)	0,4639 (0,0492)
Nenhum	k-NN	0,0790 (0,0067)	0,0994 (0,0085)	6,53 (0,59)	0,2711 (0,0493)	0,4581 (0,0589)
Nenhum	SVM_Lin	0,0808 (0,0069)	0,1023 (0,0083)	6,63 (0,63)	0,2877 (0,0364)	0,4243 (0,0737)
Nenhum	SVM_RBF	0,0777 (0,0072)	0,0972 (0,0086)	6,39 (0,59)	0,2673 (0,0468)	0,4817 (0,0573)
Nenhum	SVM_Pol	0,0779 (0,0067)	0,0982 (0,0084)	6,40 (0,55)	0,2767 (0,0391)	0,4719 (0,0509)
Nenhum	SVM_Sig	0,0850 (0,0111)	0,1081 (0,0147)	7,01 (0,87)	0,3065 (0,0627)	0,3479 (0,1799)
Nenhum	RF	0,0783 (0,0072)	0,0982 (0,0084)	6,46 (0,62)	0,2863 (0,0375)	0,4698 (0,0676)
Nenhum	XGBoost	0,0781 (0,0075)	0,0979 (0,0091)	6,43 (0,63)	0,2712 (0,0358)	0,4743 (0,0610)
MinMax	LR	0,0784 (0,0074)	0,0990 (0,0093)	6,47 (0,62)	0,2824 (0,0302)	0,4619 (0,0672)
MinMax	EN	0,0783 (0,0073)	0,0989 (0,0093)	6,47 (0,62)	0,2838 (0,0286)	0,4633 (0,0667)
MinMax	DT	0,0839 (0,0083)	0,1049 (0,0109)	6,91 (0,69)	0,2885 (0,0507)	0,3953 (0,0866)
MinMax	MLP	0,0793 (0,0077)	0,0996 (0,0085)	6,53 (0,62)	0,2997 (0,0490)	0,4549 (0,0646)
MinMax	k-NN	0,0830 (0,0067)	0,1046 (0,0086)	6,88 (0,58)	0,2970 (0,0365)	0,4008 (0,0481)
MinMax	SVM_Lin	0,0782 (0,0072)	0,0989 (0,0093)	6,44 (0,60)	0,2855 (0,0328)	0,4624 (0,0679)
MinMax	SVM_RBF	0,0779 (0,0077)	0,0984 (0,0096)	6,41 (0,64)	0,2856 (0,0314)	0,4688 (0,0671)
MinMax	SVM_Pol	0,0801 (0,0085)	0,1013 (0,0109)	6,61 (0,74)	0,2781 (0,0408)	0,4372 (0,0761)
MinMax	SVM_Sig	0,0784 (0,0072)	0,0992 (0,0092)	6,46 (0,59)	0,2865 (0,0330)	0,4597 (0,0690)
MinMax	RF	0,0793 (0,0079)	0,0997 (0,0094)	6,55 (0,67)	0,2865 (0,0405)	0,4544 (0,0709)
MinMax	XGBoost	0,0785 (0,0075)	0,0986 (0,0091)	6,48 (0,64)	0,2812 (0,0357)	0,4664 (0,0639)
Padronização	LR	0,0784 (0,0074)	0,0990 (0,0093)	6,47 (0,62)	0,2824 (0,0302)	0,4619 (0,0672)
Padronização	EN	0,0782 (0,0072)	0,0988 (0,0091)	6,46 (0,61)	0,2832 (0,0282)	0,4636 (0,0657)
Padronização	DT	0,0819 (0,0067)	0,1034 (0,0083)	6,75 (0,58)	0,2904 (0,0385)	0,4128 (0,0693)
Padronização	MLP	0,0785 (0,0076)	0,0989 (0,0093)	6,48 (0,63)	0,2844 (0,0304)	0,4629 (0,0668)
Padronização	k-NN	0,0813 (0,0072)	0,1024 (0,0088)	6,74 (0,63)	0,2925 (0,0333)	0,4255 (0,0542)
Padronização	SVM_Lin	0,0782 (0,0072)	0,0989 (0,0092)	6,45 (0,60)	0,2830 (0,0328)	0,4629 (0,0670)
Padronização	SVM_RBF	0,0780 (0,0074)	0,0979 (0,0090)	6,43 (0,62)	0,2787 (0,0335)	0,4743 (0,0639)
Padronização	SVM_Pol	0,0903 (0,0104)	0,1279 (0,0480)	7,42 (0,87)	0,5198 (0,6145)	-0,0430 (1,0819)
Padronização	SVM_Sig	0,0783 (0,0073)	0,0990 (0,0093)	6,46 (0,60)	0,2856 (0,0317)	0,4621 (0,0671)
Padronização	RF	0,0784 (0,0080)	0,0988 (0,0089)	6,46 (0,66)	0,2855 (0,0416)	0,4626 (0,0753)
Padronização	XGBoost	0,0785 (0,0076)	0,0985 (0,0091)	6,47 (0,64)	0,2803 (0,0404)	0,4667 (0,0694)

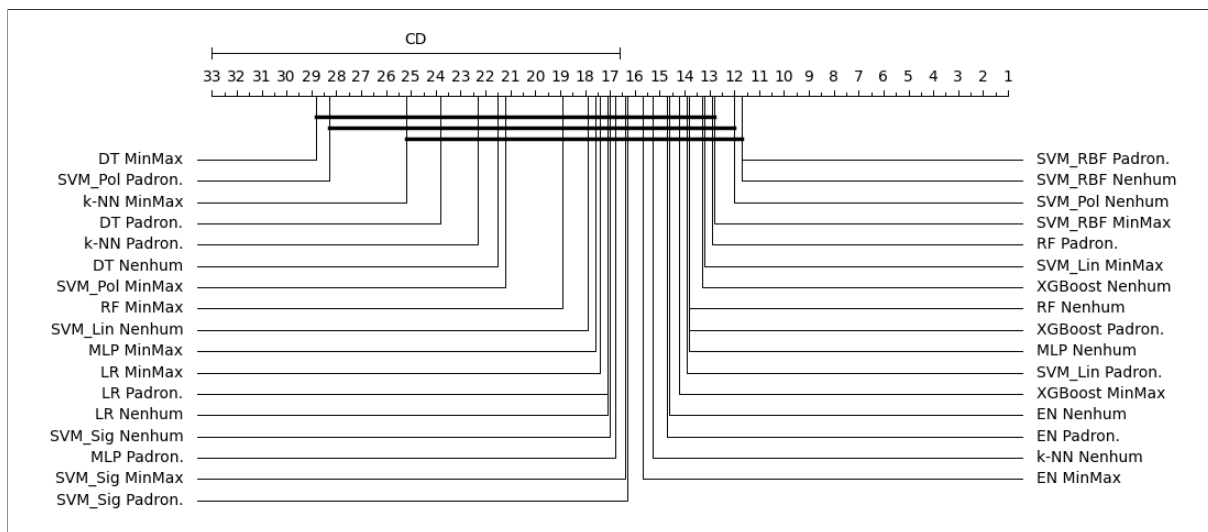
Fonte: elaborada pelo autor.

Tabela 19 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,1)	0,7 (0,04)	1 (0,0)
Nenhum	EN	2 (0,4)	0,7 (0,04)	1 (0,0)
Nenhum	DT	7 (5,4)	0,7 (0,08)	3 (0,7)
Nenhum	MLP	1694 (1769,9)	0,7 (0,06)	9 (2,3)
Nenhum	k-NN	1 (0,0)	1,0 (0,02)	78 (0,0)
Nenhum	SVM_Lin	47 (7,9)	0,7 (0,04)	28 (0,8)
Nenhum	SVM_RBF	23 (12,4)	0,7 (0,01)	27 (0,8)
Nenhum	SVM_Pol	30 (16,1)	0,7 (0,04)	27 (1,2)
Nenhum	SVM_Sig	20 (6,2)	0,8 (0,07)	32 (5,3)
Nenhum	RF	1811 (1381,8)	16,2 (7,20)	759 (595,3)
Nenhum	XGBoost	122 (284,7)	1,9 (0,45)	70 (42,5)
MinMax	LR	2 (0,3)	0,8 (0,11)	2 (0,0)
MinMax	EN	3 (0,5)	0,8 (0,05)	2 (0,0)
MinMax	DT	6 (4,1)	0,8 (0,05)	4 (0,8)
MinMax	MLP	67 (92,4)	0,8 (0,06)	6 (0,4)
MinMax	k-NN	3 (1,8)	1,0 (0,01)	78 (0,0)
MinMax	SVM_Lin	27 (6,8)	0,8 (0,03)	28 (0,8)
MinMax	SVM_RBF	26 (5,6)	0,8 (0,02)	28 (0,7)
MinMax	SVM_Pol	20 (7,0)	0,9 (0,06)	29 (0,8)
MinMax	SVM_Sig	27 (8,4)	0,9 (0,05)	28 (0,8)
MinMax	RF	633 (650,9)	18,9 (5,34)	804 (549,1)
MinMax	XGBoost	60 (107,2)	2,9 (3,06)	98 (70,8)
Padronização	LR	3 (0,0)	0,9 (0,17)	2 (0,0)
Padronização	EN	3 (0,3)	0,8 (0,08)	2 (0,0)
Padronização	DT	10 (6,0)	0,8 (0,09)	4 (0,7)
Padronização	MLP	20 (10,5)	0,8 (0,03)	7 (0,6)
Padronização	k-NN	3 (0,6)	1,1 (0,05)	78 (0,0)
Padronização	SVM_Lin	27 (8,4)	0,8 (0,03)	28 (0,7)
Padronização	SVM_RBF	22 (7,1)	0,9 (0,10)	27 (0,7)
Padronização	SVM_Pol	21 (17,3)	0,8 (0,00)	34 (1,8)
Padronização	SVM_Sig	23 (9,5)	0,9 (0,07)	28 (0,8)
Padronização	RF	1532 (1377,7)	15,1 (6,05)	870 (715,0)
Padronização	XGBoost	42 (49,5)	2,3 (1,26)	93 (79,6)

Fonte: elaborada pelo autor.

Figura 19 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-A. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 20 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0790 (0,0069)	0,0996 (0,0097)	6,52 (0,61)	0,2964 (0,0646)	0,4555 (0,0711)
Nenhum	EN	0,0790 (0,0068)	0,0996 (0,0095)	6,52 (0,61)	0,2938 (0,0611)	0,4548 (0,0693)
Nenhum	DT	0,0838 (0,0089)	0,1046 (0,0117)	6,92 (0,77)	0,2926 (0,0472)	0,3978 (0,1020)
Nenhum	MLP	0,0781 (0,0068)	0,0986 (0,0094)	6,43 (0,60)	0,2947 (0,0553)	0,4665 (0,0669)
Nenhum	k-NN	0,0793 (0,0072)	0,0993 (0,0097)	6,57 (0,62)	0,2779 (0,0441)	0,4590 (0,0703)
Nenhum	SVM_Lin	0,0881 (0,0131)	0,1113 (0,0170)	7,22 (1,11)	0,3213 (0,0859)	0,3065 (0,2102)
Nenhum	SVM_RBF	0,0782 (0,0068)	0,1006 (0,0133)	6,45 (0,60)	0,3246 (0,1187)	0,4415 (0,1273)
Nenhum	SVM_Pol	0,0810 (0,0081)	0,1023 (0,0112)	6,69 (0,68)	0,3091 (0,0617)	0,4176 (0,1330)
Nenhum	SVM_Sig	0,0846 (0,0136)	0,1069 (0,0169)	6,99 (1,16)	0,3123 (0,0712)	0,3619 (0,1918)
Nenhum	RF	0,0791 (0,0091)	0,0993 (0,0117)	6,53 (0,77)	0,2822 (0,0428)	0,4565 (0,0966)
Nenhum	XGBoost	0,0792 (0,0089)	0,0991 (0,0114)	6,54 (0,75)	0,2749 (0,0535)	0,4599 (0,0873)
MinMax	LR	0,0790 (0,0069)	0,0996 (0,0097)	6,52 (0,61)	0,2964 (0,0646)	0,4555 (0,0711)
MinMax	EN	0,0790 (0,0070)	0,0994 (0,0097)	6,52 (0,62)	0,2934 (0,0619)	0,4572 (0,0714)
MinMax	DT	0,0854 (0,0094)	0,1062 (0,0113)	7,04 (0,76)	0,2770 (0,0388)	0,3790 (0,1017)
MinMax	MLP	0,0771 (0,0070)	0,0975 (0,0104)	6,34 (0,61)	0,2913 (0,0628)	0,4787 (0,0742)
MinMax	k-NN	0,0787 (0,0067)	0,0986 (0,0091)	6,51 (0,58)	0,2816 (0,0440)	0,4669 (0,0630)
MinMax	SVM_Lin	0,0793 (0,0073)	0,1001 (0,0102)	6,54 (0,64)	0,2934 (0,0498)	0,4496 (0,0747)
MinMax	SVM_RBF	0,0775 (0,0075)	0,0982 (0,0113)	6,38 (0,65)	0,2927 (0,0758)	0,4706 (0,0843)
MinMax	SVM_Pol	0,0787 (0,0078)	0,0997 (0,0108)	6,49 (0,68)	0,3003 (0,0753)	0,4546 (0,0784)
MinMax	SVM_Sig	0,0787 (0,0069)	0,0991 (0,0098)	6,49 (0,61)	0,2828 (0,0458)	0,4612 (0,0680)
MinMax	RF	0,0794 (0,0090)	0,0991 (0,0110)	6,56 (0,77)	0,2858 (0,0397)	0,4581 (0,0958)
MinMax	XGBoost	0,0798 (0,0081)	0,0994 (0,0108)	6,57 (0,68)	0,2679 (0,0520)	0,4573 (0,0827)
Padronização	LR	0,0790 (0,0069)	0,0996 (0,0097)	6,52 (0,61)	0,2964 (0,0646)	0,4555 (0,0711)
Padronização	EN	0,0791 (0,0070)	0,0996 (0,0097)	6,53 (0,62)	0,2936 (0,0604)	0,4552 (0,0710)
Padronização	DT	0,0850 (0,0114)	0,1060 (0,0131)	7,01 (0,99)	0,2841 (0,0329)	0,3770 (0,1328)
Padronização	MLP	0,0796 (0,0079)	0,1002 (0,0110)	6,56 (0,68)	0,3120 (0,0805)	0,4484 (0,0838)
Padronização	k-NN	0,0786 (0,0070)	0,0987 (0,0098)	6,52 (0,60)	0,2871 (0,0485)	0,4656 (0,0671)
Padronização	SVM_Lin	0,0791 (0,0073)	0,1000 (0,0102)	6,53 (0,64)	0,2955 (0,0559)	0,4508 (0,0744)
Padronização	SVM_RBF	0,0785 (0,0069)	0,0986 (0,0096)	6,47 (0,61)	0,2970 (0,0612)	0,4656 (0,0752)
Padronização	SVM_Pol	0,0912 (0,0089)	0,1168 (0,0116)	7,54 (0,79)	0,3495 (0,0841)	0,2546 (0,0745)
Padronização	SVM_Sig	0,0790 (0,0069)	0,0995 (0,0097)	6,52 (0,61)	0,2885 (0,0518)	0,4562 (0,0689)
Padronização	RF	0,0792 (0,0092)	0,0993 (0,0112)	6,54 (0,79)	0,2893 (0,0408)	0,4561 (0,0957)
Padronização	XGBoost	0,0797 (0,0082)	0,0992 (0,0107)	6,57 (0,70)	0,2682 (0,0489)	0,4590 (0,0857)

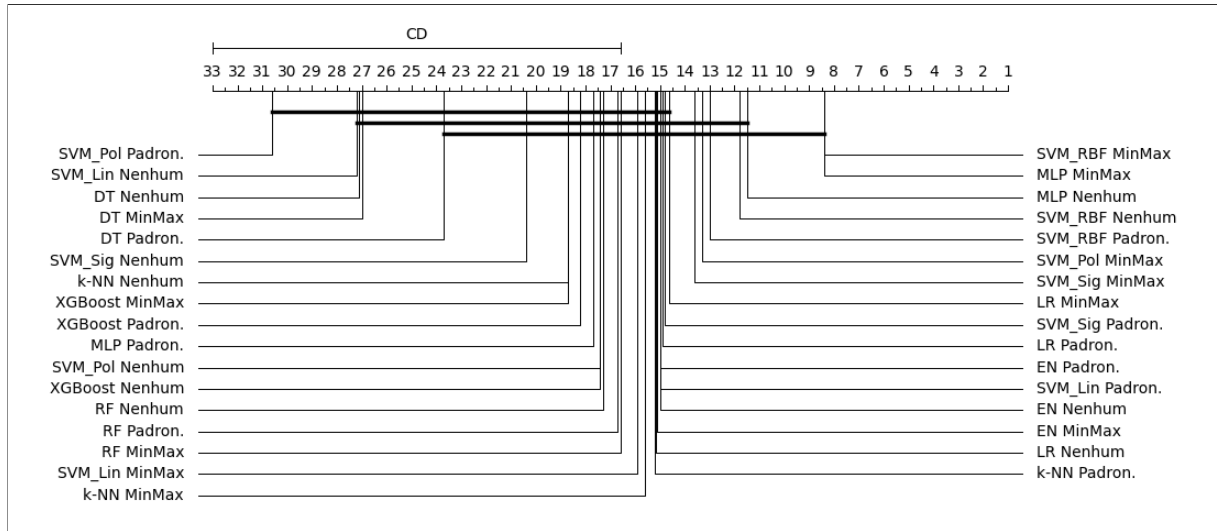
Fonte: elaborada pelo autor.

Tabela 21 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,2)	0,7 (0,07)	1 (0,0)
Nenhum	EN	2 (0,4)	0,8 (0,17)	1 (0,0)
Nenhum	DT	16 (8,8)	0,7 (0,07)	3 (0,6)
Nenhum	MLP	1355 (1617,8)	0,7 (0,07)	9 (2,7)
Nenhum	k-NN	2 (0,9)	1,0 (0,03)	71 (0,0)
Nenhum	SVM_Lin	72 (13,4)	0,8 (0,08)	26 (1,2)
Nenhum	SVM_RBF	22 (13,6)	0,8 (0,04)	26 (0,5)
Nenhum	SVM_Pol	26 (10,1)	0,7 (0,04)	27 (3,0)
Nenhum	SVM_Sig	17 (4,8)	0,8 (0,07)	30 (4,8)
Nenhum	RF	1210 (1725,7)	10,8 (5,58)	547 (504,3)
Nenhum	XGBoost	112 (213,7)	1,6 (0,15)	129 (90,7)
MinMax	LR	2 (0,1)	0,8 (0,09)	2 (0,0)
MinMax	EN	3 (0,1)	0,8 (0,02)	2 (0,0)
MinMax	DT	15 (9,0)	0,8 (0,03)	4 (0,4)
MinMax	MLP	160 (104,2)	0,8 (0,02)	7 (0,9)
MinMax	k-NN	3 (0,5)	1,0 (0,03)	72 (0,0)
MinMax	SVM_Lin	19 (9,8)	0,8 (0,02)	27 (0,4)
MinMax	SVM_RBF	25 (14,0)	0,8 (0,02)	26 (0,7)
MinMax	SVM_Pol	22 (8,0)	0,8 (0,02)	27 (0,6)
MinMax	SVM_Sig	32 (13,2)	0,8 (0,03)	26 (0,6)
MinMax	RF	773 (653,1)	14,9 (6,62)	528 (397,5)
MinMax	XGBoost	22 (13,3)	2,7 (2,34)	64 (45,9)
Padronização	LR	3 (0,1)	0,8 (0,06)	2 (0,0)
Padronização	EN	4 (0,9)	0,8 (0,03)	2 (0,0)
Padronização	DT	10 (9,7)	0,8 (0,02)	4 (0,8)
Padronização	MLP	35 (19,8)	0,8 (0,02)	5 (0,3)
Padronização	k-NN	4 (1,9)	1,1 (0,04)	72 (0,0)
Padronização	SVM_Lin	28 (10,7)	0,8 (0,02)	26 (0,6)
Padronização	SVM_RBF	23 (16,8)	0,8 (0,04)	26 (0,7)
Padronização	SVM_Pol	25 (11,1)	0,8 (0,03)	32 (3,3)
Padronização	SVM_Sig	23 (7,8)	0,8 (0,01)	26 (0,6)
Padronização	RF	1303 (1764,9)	11,8 (6,16)	635 (602,3)
Padronização	XGBoost	88 (202,3)	2,3 (1,12)	61 (31,1)

Fonte: elaborada pelo autor.

Figura 20 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-A. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 22 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0797 (0,0060)	0,1008 (0,0090)	6,58 (0,52)	0,2864 (0,0374)	0,4415 (0,0750)
Nenhum	EN	0,0797 (0,0060)	0,1008 (0,0090)	6,58 (0,52)	0,2865 (0,0376)	0,4413 (0,0751)
Nenhum	DT	0,0831 (0,0073)	0,1040 (0,0088)	6,84 (0,63)	0,2910 (0,0446)	0,4014 (0,1017)
Nenhum	MLP	0,0791 (0,0065)	0,0995 (0,0086)	6,51 (0,55)	0,2925 (0,0387)	0,4575 (0,0586)
Nenhum	k-NN	0,0799 (0,0064)	0,1008 (0,0085)	6,59 (0,56)	0,2772 (0,0492)	0,4427 (0,0603)
Nenhum	SVM_Lin	0,1043 (0,0546)	0,1321 (0,0675)	8,53 (4,45)	0,3741 (0,1708)	-0,2286 (1,6553)
Nenhum	SVM_RBF	0,0783 (0,0052)	0,0996 (0,0077)	6,45 (0,45)	0,2974 (0,0600)	0,4556 (0,0533)
Nenhum	SVM_Pol	0,0813 (0,0089)	0,1032 (0,0131)	6,70 (0,83)	0,2942 (0,0643)	0,4103 (0,1453)
Nenhum	SVM_Sig	0,0897 (0,0095)	0,1132 (0,0129)	7,38 (0,82)	0,3056 (0,0659)	0,2820 (0,1878)
Nenhum	RF	0,0802 (0,0071)	0,1009 (0,0088)	6,61 (0,61)	0,2970 (0,0432)	0,4394 (0,0787)
Nenhum	XGBoost	0,0793 (0,0073)	0,0995 (0,0097)	6,52 (0,62)	0,2820 (0,0431)	0,4556 (0,0762)
MinMax	LR	0,0797 (0,0060)	0,1008 (0,0090)	6,58 (0,52)	0,2864 (0,0374)	0,4415 (0,0750)
MinMax	EN	0,0798 (0,0060)	0,1008 (0,0090)	6,58 (0,52)	0,2874 (0,0373)	0,4412 (0,0748)
MinMax	DT	0,0823 (0,0086)	0,1035 (0,0102)	6,78 (0,74)	0,2903 (0,0443)	0,4094 (0,0938)
MinMax	MLP	0,0788 (0,0064)	0,0990 (0,0086)	6,48 (0,54)	0,2990 (0,0547)	0,4618 (0,0645)
MinMax	k-NN	0,0800 (0,0054)	0,1007 (0,0076)	6,60 (0,47)	0,2911 (0,0484)	0,4428 (0,0573)
MinMax	SVM_Lin	0,0800 (0,0057)	0,1010 (0,0088)	6,61 (0,49)	0,2897 (0,0368)	0,4388 (0,0728)
MinMax	SVM_RBF	0,0780 (0,0059)	0,0982 (0,0082)	6,42 (0,51)	0,2887 (0,0383)	0,4703 (0,0586)
MinMax	SVM_Pol	0,0804 (0,0058)	0,1013 (0,0084)	6,63 (0,54)	0,2950 (0,0395)	0,4359 (0,0653)
MinMax	SVM_Sig	0,0804 (0,0066)	0,1012 (0,0094)	6,63 (0,55)	0,2902 (0,0362)	0,4367 (0,0801)
MinMax	RF	0,0800 (0,0069)	0,1004 (0,0092)	6,59 (0,60)	0,2881 (0,0476)	0,4459 (0,0755)
MinMax	XGBoost	0,0796 (0,0075)	0,1005 (0,0101)	6,55 (0,62)	0,2928 (0,0487)	0,4442 (0,0835)
Padronização	LR	0,0797 (0,0060)	0,1008 (0,0090)	6,58 (0,52)	0,2864 (0,0374)	0,4415 (0,0750)
Padronização	EN	0,0797 (0,0059)	0,1007 (0,0090)	6,58 (0,52)	0,2864 (0,0379)	0,4419 (0,0744)
Padronização	DT	0,0824 (0,0068)	0,1034 (0,0089)	6,79 (0,59)	0,2965 (0,0468)	0,4119 (0,0814)
Padronização	MLP	0,0778 (0,0065)	0,0981 (0,0083)	6,40 (0,56)	0,2923 (0,0550)	0,4716 (0,0603)
Padronização	k-NN	0,0791 (0,0058)	0,0998 (0,0084)	6,53 (0,51)	0,2892 (0,0489)	0,4527 (0,0652)
Padronização	SVM_Lin	0,0798 (0,0062)	0,1008 (0,0092)	6,59 (0,53)	0,2880 (0,0365)	0,4409 (0,0752)
Padronização	SVM_RBF	0,0778 (0,0062)	0,0981 (0,0084)	6,40 (0,53)	0,2962 (0,0440)	0,4726 (0,0567)
Padronização	SVM_Pol	0,0828 (0,0062)	0,1058 (0,0089)	6,82 (0,58)	0,3000 (0,0430)	0,3823 (0,1025)
Padronização	SVM_Sig	0,0800 (0,0060)	0,1010 (0,0091)	6,61 (0,52)	0,2900 (0,0351)	0,4392 (0,0743)
Padronização	RF	0,0802 (0,0070)	0,1009 (0,0088)	6,61 (0,60)	0,2965 (0,0458)	0,4401 (0,0748)
Padronização	XGBoost	0,0803 (0,0061)	0,1009 (0,0085)	6,61 (0,50)	0,2883 (0,0434)	0,4411 (0,0666)

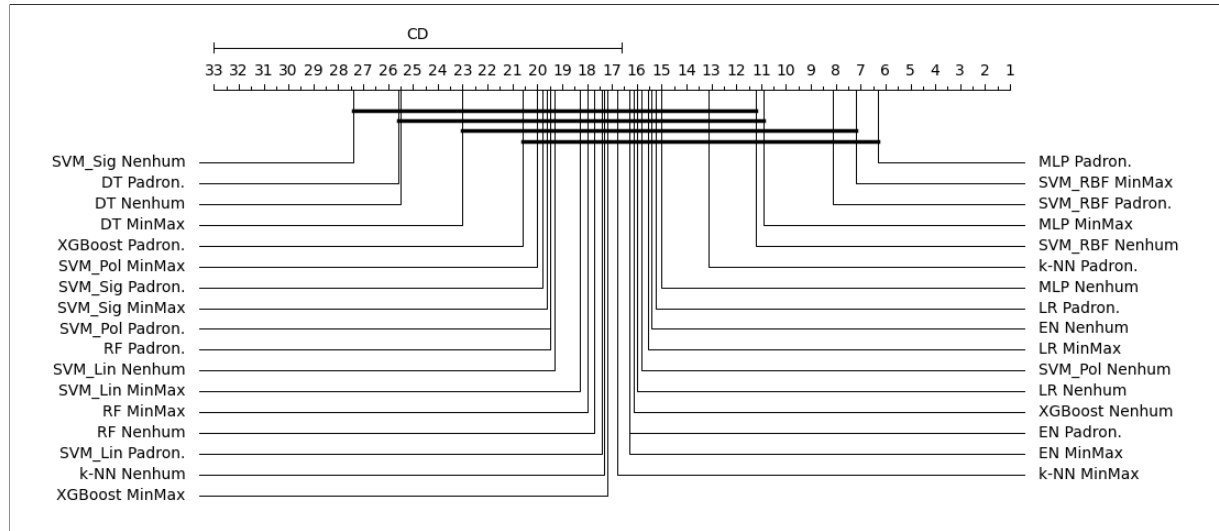
Fonte: elaborada pelo autor.

Tabela 23 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-A. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,0)	0,7 (0,04)	1 (0,0)
Nenhum	EN	2 (0,4)	0,7 (0,07)	1 (0,0)
Nenhum	DT	6 (3,4)	0,7 (0,01)	3 (0,3)
Nenhum	MLP	271 (233,2)	0,7 (0,02)	6 (0,8)
Nenhum	k-NN	1 (0,2)	0,9 (0,03)	33 (0,0)
Nenhum	SVM_Lin	57 (10,4)	0,8 (0,12)	14 (1,5)
Nenhum	SVM_RBF	30 (16,4)	0,7 (0,01)	14 (0,3)
Nenhum	SVM_Pol	26 (13,4)	0,7 (0,02)	16 (2,0)
Nenhum	SVM_Sig	23 (9,8)	0,7 (0,03)	17 (2,2)
Nenhum	RF	678 (620,6)	17,4 (6,81)	669 (536,3)
Nenhum	XGBoost	75 (117,6)	1,6 (0,42)	59 (35,5)
MinMax	LR	2 (0,0)	0,8 (0,06)	2 (0,0)
MinMax	EN	3 (0,5)	0,8 (0,03)	2 (0,0)
MinMax	DT	6 (3,1)	0,8 (0,03)	3 (0,7)
MinMax	MLP	75 (81,7)	0,8 (0,02)	6 (0,9)
MinMax	k-NN	3 (0,6)	1,0 (0,03)	33 (0,0)
MinMax	SVM_Lin	22 (9,6)	0,8 (0,02)	15 (0,4)
MinMax	SVM_RBF	30 (15,2)	0,8 (0,01)	15 (0,3)
MinMax	SVM_Pol	27 (8,8)	0,8 (0,02)	15 (0,4)
MinMax	SVM_Sig	24 (10,6)	0,8 (0,02)	15 (0,3)
MinMax	RF	571 (537,3)	13,7 (6,64)	669 (507,5)
MinMax	XGBoost	69 (124,5)	1,9 (0,23)	66 (54,3)
Padronização	LR	3 (0,1)	0,8 (0,06)	1 (0,0)
Padronização	EN	3 (0,1)	0,8 (0,03)	1 (0,0)
Padronização	DT	6 (3,9)	0,8 (0,03)	3 (0,4)
Padronização	MLP	31 (15,7)	0,8 (0,03)	5 (0,1)
Padronização	k-NN	3 (0,7)	1,0 (0,03)	33 (0,0)
Padronização	SVM_Lin	23 (9,7)	0,8 (0,02)	15 (0,4)
Padronização	SVM_RBF	28 (13,9)	0,8 (0,01)	14 (0,2)
Padronização	SVM_Pol	27 (13,7)	0,8 (0,02)	15 (0,9)
Padronização	SVM_Sig	20 (8,0)	0,8 (0,02)	15 (0,3)
Padronização	RF	671 (632,6)	14,7 (8,07)	745 (800,9)
Padronização	XGBoost	147 (331,8)	3,5 (4,50)	114 (68,5)

Fonte: elaborada pelo autor.

Figura 21 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-A. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 24 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0736 (0,0097)	0,0915 (0,0102)	6,40 (0,87)	0,2291 (0,0290)	0,2507 (0,0507)
Nenhum	EN	0,0731 (0,0096)	0,0908 (0,0103)	6,36 (0,86)	0,2242 (0,0299)	0,2623 (0,0503)
Nenhum	DT	0,0768 (0,0082)	0,0961 (0,0097)	6,68 (0,70)	0,2427 (0,0491)	0,1622 (0,1503)
Nenhum	MLP	0,0766 (0,0112)	0,0971 (0,0205)	6,65 (1,01)	0,2753 (0,1703)	0,1542 (0,2347)
Nenhum	k-NN	0,0728 (0,0093)	0,0903 (0,0102)	6,34 (0,84)	0,2223 (0,0264)	0,2702 (0,0658)
Nenhum	SVM_Lin	0,0767 (0,0121)	0,0966 (0,0145)	6,64 (1,05)	0,2473 (0,0440)	0,1511 (0,2386)
Nenhum	SVM_RBF	0,0725 (0,0091)	0,0901 (0,0104)	6,27 (0,83)	0,2300 (0,0384)	0,2699 (0,0905)
Nenhum	SVM_Pol	0,0728 (0,0101)	0,0902 (0,0116)	6,32 (0,90)	0,2308 (0,0361)	0,2722 (0,0867)
Nenhum	SVM_Sig	0,0737 (0,0105)	0,0918 (0,0113)	6,40 (0,93)	0,2288 (0,0276)	0,2475 (0,0635)
Nenhum	RF	0,0721 (0,0097)	0,0899 (0,0108)	6,25 (0,85)	0,2273 (0,0407)	0,2735 (0,0943)
Nenhum	XGBoost	0,0730 (0,0088)	0,0912 (0,0097)	6,34 (0,74)	0,2287 (0,0471)	0,2462 (0,1343)
MinMax	LR	0,0736 (0,0097)	0,0915 (0,0102)	6,40 (0,87)	0,2291 (0,0290)	0,2507 (0,0507)
MinMax	EN	0,0735 (0,0099)	0,0912 (0,0103)	6,39 (0,88)	0,2288 (0,0301)	0,2561 (0,0444)
MinMax	DT	0,0796 (0,0075)	0,0986 (0,0087)	6,90 (0,66)	0,2401 (0,0467)	0,1195 (0,1327)
MinMax	MLP	0,0746 (0,0098)	0,0923 (0,0103)	6,48 (0,86)	0,2321 (0,0262)	0,2356 (0,0847)
MinMax	k-NN	0,0800 (0,0093)	0,0998 (0,0097)	6,96 (0,87)	0,2533 (0,0446)	0,1061 (0,0784)
MinMax	SVM_Lin	0,0741 (0,0096)	0,0921 (0,0104)	6,43 (0,86)	0,2309 (0,0287)	0,2411 (0,0557)
MinMax	SVM_RBF	0,0742 (0,0094)	0,0924 (0,0101)	6,44 (0,86)	0,2331 (0,0318)	0,2357 (0,0663)
MinMax	SVM_Pol	0,0762 (0,0097)	0,0948 (0,0103)	6,60 (0,91)	0,2468 (0,0312)	0,1907 (0,0998)
MinMax	SVM_Sig	0,0738 (0,0093)	0,0920 (0,0104)	6,40 (0,86)	0,2314 (0,0266)	0,2423 (0,0590)
MinMax	RF	0,0717 (0,0093)	0,0893 (0,0107)	6,21 (0,81)	0,2198 (0,0438)	0,2827 (0,1027)
MinMax	XGBoost	0,0726 (0,0085)	0,0905 (0,0091)	6,28 (0,70)	0,2255 (0,0384)	0,2577 (0,1156)
Padronização	LR	0,0736 (0,0097)	0,0915 (0,0102)	6,40 (0,87)	0,2291 (0,0290)	0,2507 (0,0507)
Padronização	EN	0,0732 (0,0097)	0,0910 (0,0103)	6,37 (0,87)	0,2283 (0,0303)	0,2588 (0,0471)
Padronização	DT	0,0795 (0,0084)	0,0987 (0,0099)	6,91 (0,74)	0,2476 (0,0478)	0,1179 (0,1397)
Padronização	MLP	0,0734 (0,0098)	0,0912 (0,0103)	6,38 (0,86)	0,2262 (0,0287)	0,2555 (0,0514)
Padronização	k-NN	0,0768 (0,0087)	0,0957 (0,0092)	6,68 (0,81)	0,2465 (0,0395)	0,1788 (0,0487)
Padronização	SVM_Lin	0,0736 (0,0099)	0,0918 (0,0105)	6,38 (0,88)	0,2293 (0,0281)	0,2468 (0,0575)
Padronização	SVM_RBF	0,0730 (0,0092)	0,0912 (0,0102)	6,34 (0,82)	0,2316 (0,0317)	0,2560 (0,0602)
Padronização	SVM_Pol	0,0859 (0,0111)	0,1081 (0,0180)	7,47 (0,97)	0,3134 (0,1177)	-0,0536 (0,2409)
Padronização	SVM_Sig	0,0739 (0,0100)	0,0918 (0,0106)	6,41 (0,90)	0,2305 (0,0284)	0,2462 (0,0574)
Padronização	RF	0,0724 (0,0088)	0,0902 (0,0103)	6,29 (0,81)	0,2303 (0,0387)	0,2705 (0,0785)
Padronização	XGBoost	0,0726 (0,0096)	0,0904 (0,0105)	6,30 (0,81)	0,2294 (0,0467)	0,2599 (0,1324)

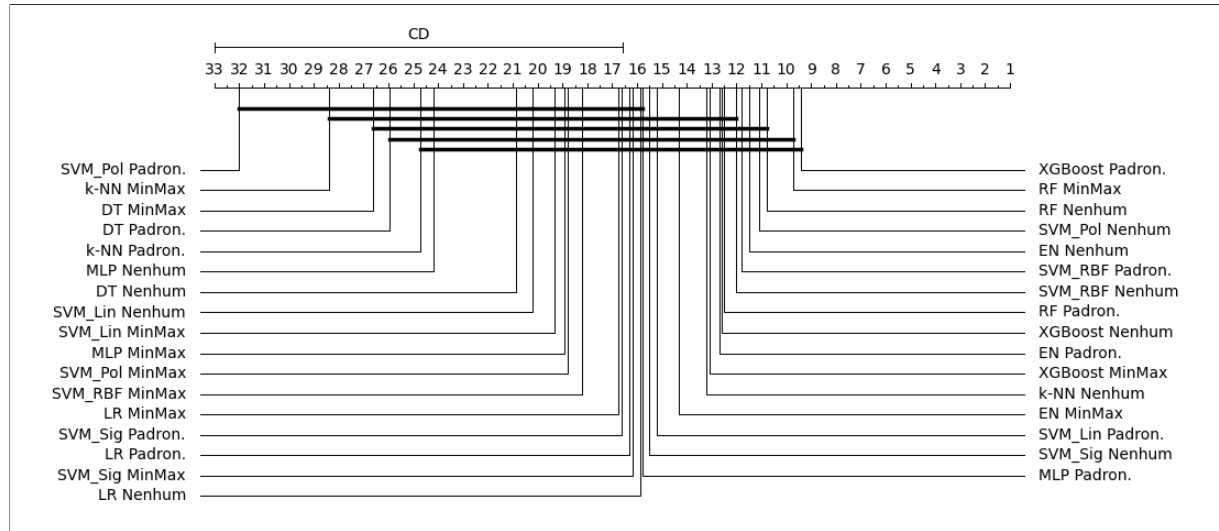
Fonte: elaborada pelo autor.

Tabela 25 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,1)	0,7 (0,03)	1 (0,0)
Nenhum	EN	2 (0,3)	0,7 (0,03)	1 (0,0)
Nenhum	DT	4 (3,1)	0,7 (0,02)	3 (0,5)
Nenhum	MLP	224 (317,3)	0,7 (0,02)	7 (2,5)
Nenhum	k-NN	1 (0,2)	0,9 (0,03)	40 (0,0)
Nenhum	SVM_Lin	26 (4,5)	0,7 (0,01)	14 (0,4)
Nenhum	SVM_RBF	10 (5,2)	0,7 (0,01)	13 (0,4)
Nenhum	SVM_Pol	14 (6,3)	0,7 (0,01)	14 (0,5)
Nenhum	SVM_Sig	9 (5,5)	0,7 (0,03)	14 (0,4)
Nenhum	RF	650 (580,3)	14,4 (7,41)	774 (693,3)
Nenhum	XGBoost	242 (377,1)	2,9 (3,70)	75 (48,6)
MinMax	LR	2 (0,1)	0,9 (0,10)	2 (0,0)
MinMax	EN	3 (0,2)	0,9 (0,13)	2 (0,0)
MinMax	DT	5 (3,0)	0,9 (0,10)	4 (0,6)
MinMax	MLP	41 (46,4)	0,9 (0,10)	7 (2,4)
MinMax	k-NN	3 (0,1)	1,1 (0,15)	41 (0,0)
MinMax	SVM_Lin	12 (6,7)	0,9 (0,14)	15 (0,5)
MinMax	SVM_RBF	11 (6,5)	0,9 (0,10)	15 (0,4)
MinMax	SVM_Pol	9 (4,3)	0,9 (0,11)	15 (0,8)
MinMax	SVM_Sig	15 (7,1)	0,9 (0,10)	15 (0,5)
MinMax	RF	812 (542,5)	20,3 (2,52)	773 (333,0)
MinMax	XGBoost	341 (684,5)	2,3 (0,86)	70 (48,0)
Padronização	LR	3 (0,2)	0,9 (0,15)	2 (0,0)
Padronização	EN	3 (0,1)	0,8 (0,03)	2 (0,0)
Padronização	DT	6 (2,7)	0,8 (0,04)	3 (0,4)
Padronização	MLP	36 (21,4)	0,8 (0,03)	9 (2,4)
Padronização	k-NN	3 (0,2)	1,1 (0,13)	40 (0,0)
Padronização	SVM_Lin	14 (7,7)	0,9 (0,10)	15 (0,5)
Padronização	SVM_RBF	13 (5,7)	0,9 (0,10)	14 (0,5)
Padronização	SVM_Pol	7 (1,6)	0,9 (0,10)	17 (0,7)
Padronização	SVM_Sig	10 (3,1)	0,9 (0,08)	14 (0,4)
Padronização	RF	330 (343,2)	11,4 (6,32)	467 (412,5)
Padronização	XGBoost	246 (363,4)	2,1 (0,38)	64 (35,3)

Fonte: elaborada pelo autor.

Figura 22 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-F. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 26 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0752 (0,0100)	0,0940 (0,0121)	6,54 (0,88)	0,2494 (0,0811)	0,2056 (0,1242)
Nenhum	EN	0,0750 (0,0096)	0,0939 (0,0114)	6,54 (0,89)	0,2535 (0,0685)	0,2092 (0,0981)
Nenhum	DT	0,0795 (0,0094)	0,0982 (0,0113)	6,90 (0,88)	0,2503 (0,0490)	0,1323 (0,1319)
Nenhum	MLP	0,0756 (0,0087)	0,0941 (0,0096)	6,55 (0,75)	0,2438 (0,0299)	0,2024 (0,0994)
Nenhum	k-NN	0,0735 (0,0097)	0,0916 (0,0113)	6,41 (0,87)	0,2234 (0,0308)	0,2465 (0,1039)
Nenhum	SVM_Lin	0,0778 (0,0117)	0,0970 (0,0144)	6,72 (0,99)	0,2680 (0,0726)	0,1546 (0,1536)
Nenhum	SVM_RBF	0,0738 (0,0104)	0,0930 (0,0125)	6,41 (0,91)	0,2633 (0,0850)	0,2150 (0,1758)
Nenhum	SVM_Pol	0,0768 (0,0138)	0,0957 (0,0171)	6,68 (1,23)	0,2527 (0,0720)	0,1819 (0,1604)
Nenhum	SVM_Sig	0,0783 (0,0097)	0,0969 (0,0106)	6,82 (0,92)	0,2539 (0,0717)	0,1497 (0,1449)
Nenhum	RF	0,0731 (0,0102)	0,0907 (0,0108)	6,35 (0,89)	0,2282 (0,0431)	0,2613 (0,0961)
Nenhum	XGBoost	0,0747 (0,0112)	0,0938 (0,0121)	6,48 (0,94)	0,2437 (0,0657)	0,2069 (0,1385)
MinMax	LR	0,0752 (0,0100)	0,0940 (0,0121)	6,54 (0,88)	0,2494 (0,0811)	0,2056 (0,1242)
MinMax	EN	0,0750 (0,0101)	0,0934 (0,0117)	6,53 (0,89)	0,2432 (0,0681)	0,2178 (0,1034)
MinMax	DT	0,0782 (0,0091)	0,0974 (0,0107)	6,82 (0,90)	0,2608 (0,0526)	0,1484 (0,0843)
MinMax	MLP	0,0741 (0,0098)	0,0931 (0,0139)	6,43 (0,83)	0,2513 (0,1086)	0,2239 (0,1149)
MinMax	k-NN	0,0737 (0,0092)	0,0916 (0,0103)	6,43 (0,80)	0,2236 (0,0283)	0,2463 (0,0864)
MinMax	SVM_Lin	0,0745 (0,0099)	0,0926 (0,0112)	6,47 (0,87)	0,2400 (0,0553)	0,2318 (0,0884)
MinMax	SVM_RBF	0,0735 (0,0092)	0,0922 (0,0115)	6,38 (0,80)	0,2565 (0,0791)	0,2290 (0,1625)
MinMax	SVM_Pol	0,0770 (0,0103)	0,0976 (0,0126)	6,69 (0,92)	0,2870 (0,0989)	0,1412 (0,1562)
MinMax	SVM_Sig	0,0742 (0,0101)	0,0924 (0,0113)	6,44 (0,88)	0,2376 (0,0483)	0,2355 (0,0866)
MinMax	RF	0,0727 (0,0110)	0,0909 (0,0120)	6,32 (0,97)	0,2356 (0,0458)	0,2611 (0,0833)
MinMax	XGBoost	0,0751 (0,0105)	0,0933 (0,0111)	6,52 (0,87)	0,2332 (0,0489)	0,2158 (0,1174)
Padronização	LR	0,0752 (0,0100)	0,0940 (0,0121)	6,54 (0,88)	0,2494 (0,0811)	0,2056 (0,1242)
Padronização	EN	0,0752 (0,0099)	0,0938 (0,0120)	6,55 (0,86)	0,2493 (0,0808)	0,2103 (0,1214)
Padronização	DT	0,0768 (0,0101)	0,0958 (0,0113)	6,68 (0,94)	0,2504 (0,0441)	0,1784 (0,0849)
Padronização	MLP	0,0764 (0,0099)	0,0953 (0,0130)	6,65 (0,88)	0,2508 (0,0729)	0,1875 (0,1080)
Padronização	k-NN	0,0742 (0,0093)	0,0923 (0,0105)	6,49 (0,81)	0,2308 (0,0304)	0,2375 (0,0732)
Padronização	SVM_Lin	0,0746 (0,0100)	0,0928 (0,0114)	6,48 (0,86)	0,2398 (0,0617)	0,2273 (0,0991)
Padronização	SVM_RBF	0,0749 (0,0107)	0,0939 (0,0125)	6,50 (0,91)	0,2599 (0,0803)	0,2022 (0,1633)
Padronização	SVM_Pol	0,0852 (0,0086)	0,1046 (0,0106)	7,40 (0,85)	0,2858 (0,0470)	0,0202 (0,0453)
Padronização	SVM_Sig	0,0746 (0,0095)	0,0931 (0,0111)	6,49 (0,84)	0,2463 (0,0624)	0,2236 (0,0907)
Padronização	RF	0,0728 (0,0101)	0,0910 (0,0109)	6,32 (0,88)	0,2320 (0,0447)	0,2570 (0,0917)
Padronização	XGBoost	0,0751 (0,0113)	0,0945 (0,0130)	6,51 (0,97)	0,2569 (0,0734)	0,1963 (0,1413)

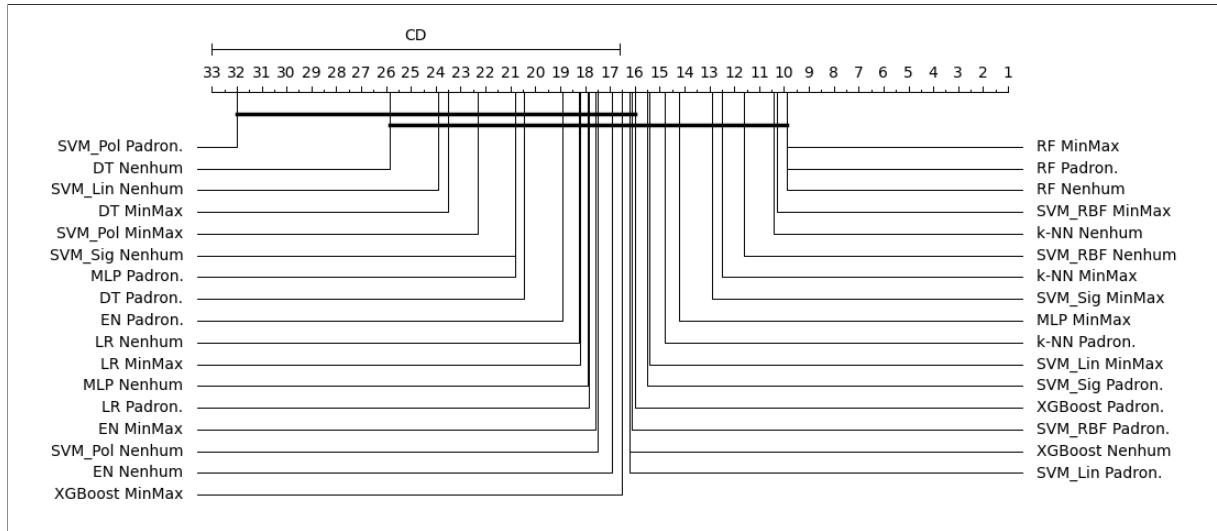
Fonte: elaborada pelo autor.

Tabela 27 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,1)	0,7 (0,05)	1 (0,0)
Nenhum	EN	2 (0,1)	0,7 (0,09)	1 (0,0)
Nenhum	DT	3 (2,6)	0,7 (0,07)	3 (0,5)
Nenhum	MLP	577 (557,0)	0,7 (0,09)	8 (2,4)
Nenhum	k-NN	2 (0,9)	1,0 (0,11)	36 (0,0)
Nenhum	SVM_Lin	35 (9,4)	0,7 (0,05)	13 (0,4)
Nenhum	SVM_RBF	7 (3,0)	0,7 (0,04)	13 (0,4)
Nenhum	SVM_Pol	15 (9,7)	0,7 (0,01)	13 (0,9)
Nenhum	SVM_Sig	6 (2,4)	0,7 (0,01)	14 (0,9)
Nenhum	RF	473 (524,5)	11,4 (8,46)	374 (294,7)
Nenhum	XGBoost	193 (309,8)	2,6 (3,13)	62 (35,7)
MinMax	LR	2 (0,1)	0,9 (0,10)	2 (0,0)
MinMax	EN	3 (0,6)	0,8 (0,10)	2 (0,0)
MinMax	DT	4 (0,3)	0,8 (0,10)	3 (0,6)
MinMax	MLP	19 (11,5)	0,8 (0,04)	6 (1,0)
MinMax	k-NN	2 (0,1)	1,1 (0,12)	37 (0,0)
MinMax	SVM_Lin	5 (1,8)	0,8 (0,03)	14 (0,4)
MinMax	SVM_RBF	7 (2,6)	0,9 (0,12)	14 (0,5)
MinMax	SVM_Pol	8 (6,0)	0,8 (0,03)	14 (0,5)
MinMax	SVM_Sig	8 (2,8)	0,9 (0,06)	14 (0,4)
MinMax	RF	796 (857,4)	14,3 (6,67)	765 (779,0)
MinMax	XGBoost	153 (255,4)	1,9 (0,08)	70 (42,0)
Padronização	LR	2 (0,2)	0,9 (0,12)	2 (0,0)
Padronização	EN	3 (0,5)	0,8 (0,05)	2 (0,0)
Padronização	DT	5 (3,3)	0,8 (0,03)	3 (0,2)
Padronização	MLP	32 (21,0)	0,8 (0,04)	7 (2,5)
Padronização	k-NN	3 (0,1)	1,0 (0,02)	37 (0,0)
Padronização	SVM_Lin	11 (2,5)	0,8 (0,03)	14 (0,4)
Padronização	SVM_RBF	10 (3,9)	0,8 (0,02)	14 (0,4)
Padronização	SVM_Pol	6 (2,6)	0,8 (0,03)	15 (0,4)
Padronização	SVM_Sig	8 (3,1)	0,8 (0,03)	14 (0,4)
Padronização	RF	466 (228,2)	14,6 (7,96)	1147 (1072,5)
Padronização	XGBoost	284 (417,1)	2,1 (0,60)	66 (43,3)

Fonte: elaborada pelo autor.

Figura 23 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-F. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 28 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0744 (0,0093)	0,0921 (0,0110)	6,47 (0,86)	0,2272 (0,0303)	0,2422 (0,0613)
Nenhum	EN	0,0744 (0,0094)	0,0921 (0,0110)	6,47 (0,86)	0,2273 (0,0304)	0,2421 (0,0612)
Nenhum	DT	0,0755 (0,0101)	0,0954 (0,0103)	6,55 (0,90)	0,2468 (0,0466)	0,1784 (0,1271)
Nenhum	MLP	0,0744 (0,0087)	0,0920 (0,0113)	6,45 (0,78)	0,2241 (0,0592)	0,2405 (0,0977)
Nenhum	k-NN	0,0725 (0,0091)	0,0900 (0,0105)	6,31 (0,84)	0,2211 (0,0269)	0,2737 (0,0730)
Nenhum	SVM_Lin	0,0749 (0,0099)	0,0935 (0,0123)	6,49 (0,87)	0,2360 (0,0323)	0,2173 (0,1023)
Nenhum	SVM_RBF	0,0734 (0,0102)	0,0906 (0,0117)	6,37 (0,91)	0,2238 (0,0348)	0,2649 (0,0840)
Nenhum	SVM_Pol	0,0737 (0,0096)	0,0912 (0,0111)	6,38 (0,88)	0,2256 (0,0387)	0,2544 (0,0835)
Nenhum	SVM_Sig	0,0763 (0,0101)	0,0944 (0,0117)	6,64 (0,91)	0,2391 (0,0357)	0,2026 (0,0907)
Nenhum	RF	0,0726 (0,0087)	0,0908 (0,0101)	6,30 (0,79)	0,2306 (0,0441)	0,2583 (0,0984)
Nenhum	XGBoost	0,0740 (0,0090)	0,0909 (0,0107)	6,42 (0,78)	0,2281 (0,0445)	0,2549 (0,1085)
MinMax	LR	0,0744 (0,0093)	0,0921 (0,0110)	6,47 (0,86)	0,2272 (0,0303)	0,2422 (0,0613)
MinMax	EN	0,0744 (0,0093)	0,0921 (0,0110)	6,47 (0,86)	0,2273 (0,0314)	0,2414 (0,0608)
MinMax	DT	0,0775 (0,0092)	0,0961 (0,0105)	6,72 (0,76)	0,2365 (0,0433)	0,1656 (0,1200)
MinMax	MLP	0,0720 (0,0093)	0,0896 (0,0116)	6,26 (0,84)	0,2279 (0,0490)	0,2801 (0,0917)
MinMax	k-NN	0,0735 (0,0087)	0,0922 (0,0100)	6,39 (0,83)	0,2309 (0,0302)	0,2380 (0,0699)
MinMax	SVM_Lin	0,0746 (0,0097)	0,0924 (0,0110)	6,47 (0,88)	0,2293 (0,0289)	0,2359 (0,0620)
MinMax	SVM_RBF	0,0721 (0,0095)	0,0893 (0,0110)	6,25 (0,86)	0,2231 (0,0326)	0,2836 (0,0931)
MinMax	SVM_Pol	0,0755 (0,0103)	0,0931 (0,0117)	6,54 (0,94)	0,2384 (0,0321)	0,2223 (0,1106)
MinMax	SVM_Sig	0,0752 (0,0096)	0,0930 (0,0108)	6,52 (0,89)	0,2317 (0,0276)	0,2268 (0,0513)
MinMax	RF	0,0736 (0,0098)	0,0917 (0,0112)	6,40 (0,89)	0,2393 (0,0409)	0,2445 (0,1028)
MinMax	XGBoost	0,0734 (0,0092)	0,0910 (0,0111)	6,37 (0,80)	0,2295 (0,0492)	0,2511 (0,1338)
Padronização	LR	0,0744 (0,0093)	0,0921 (0,0110)	6,47 (0,86)	0,2272 (0,0303)	0,2422 (0,0613)
Padronização	EN	0,0744 (0,0094)	0,0921 (0,0110)	6,47 (0,86)	0,2276 (0,0313)	0,2419 (0,0609)
Padronização	DT	0,0775 (0,0092)	0,0961 (0,0105)	6,73 (0,80)	0,2437 (0,0496)	0,1670 (0,1216)
Padronização	MLP	0,0732 (0,0091)	0,0902 (0,0109)	6,35 (0,83)	0,2274 (0,0404)	0,2700 (0,0862)
Padronização	k-NN	0,0731 (0,0099)	0,0913 (0,0109)	6,36 (0,92)	0,2268 (0,0310)	0,2542 (0,0710)
Padronização	SVM_Lin	0,0748 (0,0097)	0,0927 (0,0111)	6,47 (0,88)	0,2310 (0,0297)	0,2322 (0,0653)
Padronização	SVM_RBF	0,0722 (0,0096)	0,0894 (0,0112)	6,27 (0,86)	0,2217 (0,0401)	0,2834 (0,0918)
Padronização	SVM_Pol	0,0806 (0,0084)	0,0994 (0,0097)	6,97 (0,77)	0,2573 (0,0286)	0,1115 (0,0874)
Padronização	SVM_Sig	0,0753 (0,0093)	0,0932 (0,0105)	6,54 (0,88)	0,2285 (0,0283)	0,2225 (0,0675)
Padronização	RF	0,0732 (0,0089)	0,0909 (0,0101)	6,36 (0,81)	0,2286 (0,0375)	0,2574 (0,0881)
Padronização	XGBoost	0,0735 (0,0097)	0,0913 (0,0111)	6,37 (0,85)	0,2337 (0,0462)	0,2463 (0,1301)

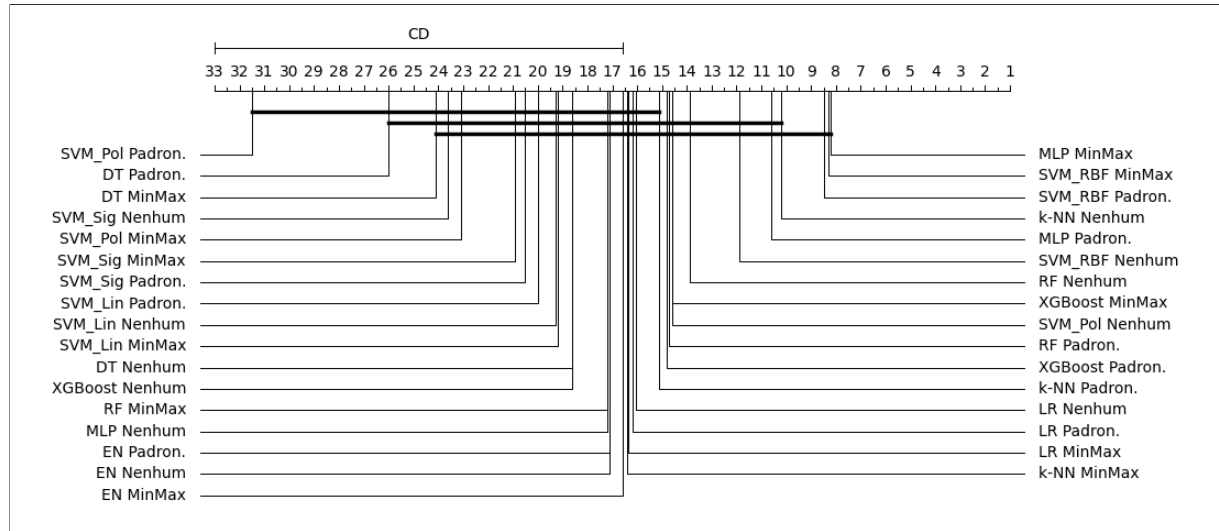
Fonte: elaborada pelo autor.

Tabela 29 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-F. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,0)	0,7 (0,07)	1 (0,0)
Nenhum	EN	2 (0,1)	0,6 (0,02)	1 (0,0)
Nenhum	DT	2 (0,6)	0,7 (0,06)	3 (0,4)
Nenhum	MLP	371 (234,0)	0,7 (0,03)	8 (2,8)
Nenhum	k-NN	1 (0,0)	0,9 (0,05)	15 (0,0)
Nenhum	SVM_Lin	25 (8,4)	0,7 (0,04)	7 (0,1)
Nenhum	SVM_RBF	11 (6,9)	0,7 (0,09)	7 (0,2)
Nenhum	SVM_Pol	16 (5,4)	0,7 (0,09)	7 (0,2)
Nenhum	SVM_Sig	6 (3,1)	0,7 (0,01)	7 (0,2)
Nenhum	RF	345 (199,2)	15,9 (3,85)	1098 (1473,3)
Nenhum	XGBoost	164 (261,5)	1,6 (0,37)	52 (31,5)
MinMax	LR	2 (0,1)	0,8 (0,15)	1 (0,0)
MinMax	EN	3 (0,2)	0,8 (0,03)	2 (0,0)
MinMax	DT	4 (1,5)	0,8 (0,10)	3 (0,4)
MinMax	MLP	32 (8,8)	0,8 (0,03)	6 (0,7)
MinMax	k-NN	2 (0,3)	1,0 (0,05)	16 (0,0)
MinMax	SVM_Lin	13 (3,9)	0,8 (0,01)	7 (0,2)
MinMax	SVM_RBF	11 (9,5)	0,8 (0,00)	7 (0,2)
MinMax	SVM_Pol	11 (6,7)	0,8 (0,02)	8 (0,2)
MinMax	SVM_Sig	13 (4,2)	0,8 (0,05)	8 (0,2)
MinMax	RF	207 (106,1)	12,5 (6,56)	387 (274,9)
MinMax	XGBoost	128 (175,2)	2,8 (2,31)	38 (10,0)
Padronização	LR	3 (0,0)	0,8 (0,07)	1 (0,0)
Padronização	EN	3 (0,5)	0,8 (0,06)	1 (0,0)
Padronização	DT	3 (0,9)	0,8 (0,06)	3 (0,7)
Padronização	MLP	13 (7,8)	0,8 (0,06)	5 (0,3)
Padronização	k-NN	3 (0,4)	1,1 (0,12)	15 (0,0)
Padronização	SVM_Lin	14 (5,8)	0,8 (0,03)	7 (0,1)
Padronização	SVM_RBF	12 (4,2)	0,8 (0,03)	7 (0,2)
Padronização	SVM_Pol	7 (4,2)	0,8 (0,02)	8 (0,2)
Padronização	SVM_Sig	12 (4,3)	0,8 (0,03)	7 (0,1)
Padronização	RF	339 (284,4)	15,4 (7,22)	519 (382,2)
Padronização	XGBoost	294 (815,9)	3,3 (3,97)	57 (32,7)

Fonte: elaborada pelo autor.

Figura 24 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-F. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 30 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E2-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0839 (0,0121)	0,1064 (0,0153)	6,52 (1,01)	0,2605 (0,0456)	0,1848 (0,1780)
Nenhum	EN	0,0835 (0,0117)	0,1060 (0,0148)	6,49 (0,98)	0,2617 (0,0440)	0,1916 (0,1684)
Nenhum	DT	0,0898 (0,0135)	0,1140 (0,0157)	6,97 (1,15)	0,2875 (0,0635)	0,0637 (0,1958)
Nenhum	MLP	0,0826 (0,0119)	0,1046 (0,0153)	6,43 (1,00)	0,2578 (0,0461)	0,2121 (0,1725)
Nenhum	k-NN	0,0860 (0,0115)	0,1078 (0,0148)	6,70 (0,99)	0,2734 (0,0528)	0,1633 (0,1821)
Nenhum	SVM_Lin	0,0877 (0,0187)	0,1109 (0,0212)	6,76 (1,41)	0,2790 (0,0524)	0,1116 (0,2505)
Nenhum	SVM_RBF	0,0840 (0,0110)	0,1048 (0,0143)	6,51 (0,92)	0,2680 (0,0514)	0,1959 (0,2320)
Nenhum	SVM_Pol	0,0828 (0,0096)	0,1043 (0,0131)	6,43 (0,82)	0,2667 (0,0472)	0,2127 (0,1663)
Nenhum	SVM_Sig	0,0863 (0,0141)	0,1091 (0,0177)	6,73 (1,24)	0,2705 (0,0539)	0,1491 (0,1615)
Nenhum	RF	0,0856 (0,0127)	0,1074 (0,0157)	6,65 (1,08)	0,2607 (0,0488)	0,1733 (0,1575)
Nenhum	XGBoost	0,0849 (0,0120)	0,1064 (0,0153)	6,60 (1,01)	0,2639 (0,0441)	0,1842 (0,1879)
MinMax	LR	0,0839 (0,0121)	0,1064 (0,0153)	6,52 (1,01)	0,2605 (0,0456)	0,1848 (0,1780)
MinMax	EN	0,0835 (0,0121)	0,1062 (0,0156)	6,50 (1,02)	0,2619 (0,0452)	0,1909 (0,1592)
MinMax	DT	0,0874 (0,0112)	0,1112 (0,0147)	6,78 (1,03)	0,2778 (0,0592)	0,1060 (0,1993)
MinMax	MLP	0,0858 (0,0095)	0,1084 (0,0126)	6,65 (0,81)	0,2702 (0,0413)	0,1455 (0,1909)
MinMax	k-NN	0,0866 (0,0134)	0,1092 (0,0177)	6,76 (1,14)	0,2681 (0,0566)	0,1542 (0,1497)
MinMax	SVM_Lin	0,0851 (0,0120)	0,1082 (0,0155)	6,60 (1,01)	0,2674 (0,0467)	0,1558 (0,1870)
MinMax	SVM_RBF	0,0839 (0,0116)	0,1063 (0,0150)	6,52 (0,99)	0,2619 (0,0423)	0,1849 (0,1783)
MinMax	SVM_Pol	0,0910 (0,0157)	0,1173 (0,0206)	7,10 (1,34)	0,3105 (0,1074)	-0,0070 (0,3454)
MinMax	SVM_Sig	0,0851 (0,0127)	0,1080 (0,0162)	6,61 (1,07)	0,2670 (0,0469)	0,1625 (0,1741)
MinMax	RF	0,0858 (0,0123)	0,1079 (0,0151)	6,67 (1,04)	0,2644 (0,0493)	0,1657 (0,1534)
MinMax	XGBoost	0,0863 (0,0109)	0,1079 (0,0132)	6,69 (0,92)	0,2668 (0,0452)	0,1572 (0,1812)
Padronização	LR	0,0839 (0,0121)	0,1064 (0,0153)	6,52 (1,01)	0,2605 (0,0456)	0,1848 (0,1780)
Padronização	EN	0,0835 (0,0123)	0,1060 (0,0154)	6,49 (1,03)	0,2622 (0,0439)	0,1934 (0,1626)
Padronização	DT	0,0877 (0,0130)	0,1109 (0,0151)	6,80 (1,11)	0,2852 (0,0495)	0,1202 (0,1539)
Padronização	MLP	0,0836 (0,0113)	0,1057 (0,0144)	6,50 (0,93)	0,2613 (0,0443)	0,1922 (0,1844)
Padronização	k-NN	0,0854 (0,0126)	0,1080 (0,0165)	6,68 (1,07)	0,2656 (0,0496)	0,1732 (0,1299)
Padronização	SVM_Lin	0,0846 (0,0117)	0,1075 (0,0152)	6,57 (0,97)	0,2633 (0,0489)	0,1617 (0,2037)
Padronização	SVM_RBF	0,0839 (0,0112)	0,1059 (0,0143)	6,50 (0,94)	0,2685 (0,0436)	0,1901 (0,1775)
Padronização	SVM_Pol	0,0981 (0,0151)	0,1233 (0,0185)	7,66 (1,31)	0,2983 (0,0446)	-0,0747 (0,1151)
Padronização	SVM_Sig	0,0838 (0,0124)	0,1065 (0,0156)	6,50 (1,03)	0,2648 (0,0416)	0,1838 (0,1819)
Padronização	RF	0,0858 (0,0120)	0,1077 (0,0150)	6,66 (1,04)	0,2654 (0,0485)	0,1672 (0,1644)
Padronização	XGBoost	0,0856 (0,0113)	0,1073 (0,0145)	6,64 (0,95)	0,2604 (0,0478)	0,1650 (0,2058)

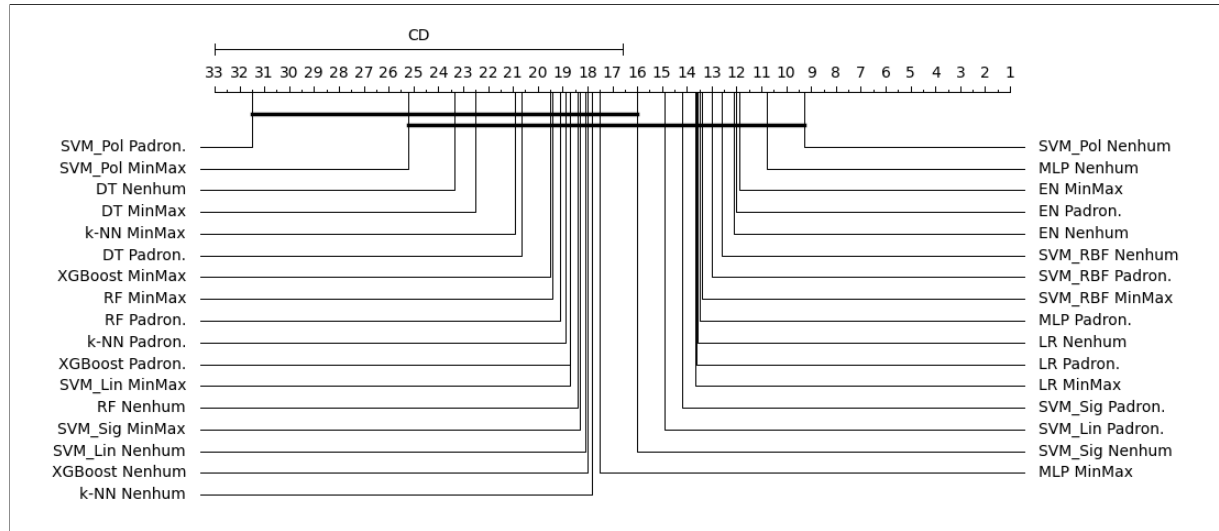
Fonte: elaborada pelo autor.

Tabela 31 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E2-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,0)	0,7 (0,05)	1 (0,0)
Nenhum	EN	2 (0,3)	0,7 (0,08)	1 (0,0)
Nenhum	DT	2 (0,7)	0,7 (0,01)	2 (0,2)
Nenhum	MLP	152 (179,5)	0,7 (0,03)	6 (2,3)
Nenhum	k-NN	1 (0,2)	0,9 (0,02)	24 (0,0)
Nenhum	SVM_Lin	20 (4,1)	0,7 (0,01)	10 (0,2)
Nenhum	SVM_RBF	9 (7,4)	0,7 (0,03)	10 (0,4)
Nenhum	SVM_Pol	8 (3,2)	0,7 (0,01)	10 (0,5)
Nenhum	SVM_Sig	5 (0,9)	0,7 (0,02)	11 (0,7)
Nenhum	RF	447 (278,1)	16,0 (6,80)	396 (320,0)
Nenhum	XGBoost	288 (575,4)	1,6 (0,15)	49 (23,7)
MinMax	LR	2 (0,2)	0,8 (0,10)	2 (0,0)
MinMax	EN	3 (0,2)	0,8 (0,05)	2 (0,0)
MinMax	DT	4 (1,7)	0,8 (0,06)	3 (0,2)
MinMax	MLP	36 (17,4)	0,8 (0,02)	6 (0,7)
MinMax	k-NN	3 (0,4)	1,0 (0,02)	25 (0,0)
MinMax	SVM_Lin	7 (4,8)	0,8 (0,01)	11 (0,3)
MinMax	SVM_RBF	9 (5,0)	0,9 (0,13)	11 (0,4)
MinMax	SVM_Pol	5 (2,2)	0,8 (0,03)	11 (0,3)
MinMax	SVM_Sig	6 (1,8)	0,8 (0,03)	11 (0,3)
MinMax	RF	347 (285,6)	10,9 (5,71)	253 (207,2)
MinMax	XGBoost	163 (385,8)	2,4 (1,60)	53 (42,0)
Padronização	LR	3 (0,2)	0,8 (0,08)	2 (0,0)
Padronização	EN	3 (0,7)	0,8 (0,04)	2 (0,0)
Padronização	DT	3 (0,8)	0,8 (0,09)	3 (0,2)
Padronização	MLP	27 (24,0)	0,8 (0,08)	6 (0,9)
Padronização	k-NN	3 (0,1)	1,0 (0,02)	24 (0,0)
Padronização	SVM_Lin	11 (4,3)	0,8 (0,04)	11 (0,3)
Padronização	SVM_RBF	10 (4,6)	0,8 (0,03)	11 (0,3)
Padronização	SVM_Pol	5 (0,4)	0,8 (0,02)	13 (0,5)
Padronização	SVM_Sig	9 (3,9)	0,8 (0,02)	11 (0,3)
Padronização	RF	524 (260,9)	17,2 (6,88)	382 (166,8)
Padronização	XGBoost	173 (250,1)	1,9 (0,19)	74 (43,5)

Fonte: elaborada pelo autor.

Figura 25 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E2-M. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 32 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E3-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0837 (0,0105)	0,1048 (0,0141)	6,51 (0,89)	0,2655 (0,0593)	0,1938 (0,2255)
Nenhum	EN	0,0835 (0,0108)	0,1048 (0,0139)	6,49 (0,91)	0,2647 (0,0553)	0,1977 (0,2107)
Nenhum	DT	0,0912 (0,0136)	0,1152 (0,0173)	7,10 (1,13)	0,2968 (0,0714)	0,0456 (0,1972)
Nenhum	MLP	0,0873 (0,0101)	0,1085 (0,0125)	6,77 (0,84)	0,2731 (0,0408)	0,1413 (0,2073)
Nenhum	k-NN	0,0864 (0,0111)	0,1085 (0,0143)	6,75 (0,96)	0,2728 (0,0499)	0,1565 (0,1472)
Nenhum	SVM_Lin	0,0875 (0,0096)	0,1102 (0,0145)	6,74 (0,77)	0,2772 (0,0661)	0,1072 (0,2553)
Nenhum	SVM_RBF	0,0826 (0,0085)	0,1035 (0,0115)	6,40 (0,74)	0,2674 (0,0508)	0,2170 (0,1913)
Nenhum	SVM_Pol	0,0832 (0,0094)	0,1046 (0,0127)	6,44 (0,81)	0,2644 (0,0510)	0,2022 (0,1905)
Nenhum	SVM_Sig	0,0930 (0,0135)	0,1171 (0,0171)	7,27 (1,14)	0,2923 (0,0373)	0,0166 (0,1915)
Nenhum	RF	0,0860 (0,0101)	0,1086 (0,0134)	6,68 (0,87)	0,2784 (0,0589)	0,1454 (0,1852)
Nenhum	XGBoost	0,0868 (0,0112)	0,1090 (0,0150)	6,73 (0,97)	0,2818 (0,0641)	0,1317 (0,2357)
MinMax	LR	0,0837 (0,0105)	0,1048 (0,0141)	6,51 (0,89)	0,2655 (0,0593)	0,1938 (0,2255)
MinMax	EN	0,0834 (0,0106)	0,1047 (0,0138)	6,48 (0,91)	0,2661 (0,0541)	0,2016 (0,1981)
MinMax	DT	0,0929 (0,0112)	0,1153 (0,0142)	7,22 (0,92)	0,2843 (0,0564)	0,0396 (0,1940)
MinMax	MLP	0,0846 (0,0107)	0,1051 (0,0134)	6,58 (0,94)	0,2725 (0,0516)	0,1921 (0,2085)
MinMax	k-NN	0,0860 (0,0112)	0,1082 (0,0142)	6,72 (0,97)	0,2732 (0,0541)	0,1594 (0,1601)
MinMax	SVM_Lin	0,0837 (0,0110)	0,1054 (0,0138)	6,47 (0,91)	0,2645 (0,0516)	0,1920 (0,2007)
MinMax	SVM_RBF	0,0836 (0,0107)	0,1040 (0,0138)	6,47 (0,90)	0,2641 (0,0486)	0,2141 (0,1787)
MinMax	SVM_Pol	0,0834 (0,0107)	0,1056 (0,0149)	6,48 (0,93)	0,2728 (0,0595)	0,1852 (0,2145)
MinMax	SVM_Sig	0,0836 (0,0107)	0,1048 (0,0133)	6,47 (0,88)	0,2642 (0,0525)	0,2037 (0,1746)
MinMax	RF	0,0861 (0,0110)	0,1083 (0,0147)	6,69 (0,94)	0,2778 (0,0621)	0,1544 (0,1714)
MinMax	XGBoost	0,0877 (0,0091)	0,1105 (0,0123)	6,79 (0,81)	0,2895 (0,0615)	0,0932 (0,2751)
Padronização	LR	0,0837 (0,0105)	0,1048 (0,0141)	6,51 (0,89)	0,2655 (0,0593)	0,1938 (0,2255)
Padronização	EN	0,0831 (0,0110)	0,1045 (0,0144)	6,46 (0,94)	0,2657 (0,0552)	0,2054 (0,2000)
Padronização	DT	0,0903 (0,0127)	0,1131 (0,0170)	7,01 (1,05)	0,2850 (0,0739)	0,0770 (0,2137)
Padronização	MLP	0,0848 (0,0107)	0,1061 (0,0135)	6,60 (0,92)	0,2760 (0,0531)	0,1765 (0,2131)
Padronização	k-NN	0,0855 (0,0115)	0,1072 (0,0151)	6,67 (1,00)	0,2707 (0,0501)	0,1766 (0,1580)
Padronização	SVM_Lin	0,0843 (0,0106)	0,1057 (0,0141)	6,52 (0,88)	0,2708 (0,0552)	0,1798 (0,2341)
Padronização	SVM_RBF	0,0827 (0,0104)	0,1036 (0,0133)	6,40 (0,88)	0,2690 (0,0578)	0,2214 (0,1754)
Padronização	SVM_Pol	0,0961 (0,0145)	0,1211 (0,0188)	7,51 (1,29)	0,3117 (0,0654)	-0,0340 (0,0896)
Padronização	SVM_Sig	0,0834 (0,0117)	0,1047 (0,0146)	6,46 (0,98)	0,2637 (0,0553)	0,2056 (0,1839)
Padronização	RF	0,0851 (0,0111)	0,1069 (0,0150)	6,62 (0,95)	0,2694 (0,0599)	0,1787 (0,1544)
Padronização	XGBoost	0,0871 (0,0113)	0,1101 (0,0163)	6,74 (0,97)	0,2875 (0,0692)	0,1168 (0,2309)

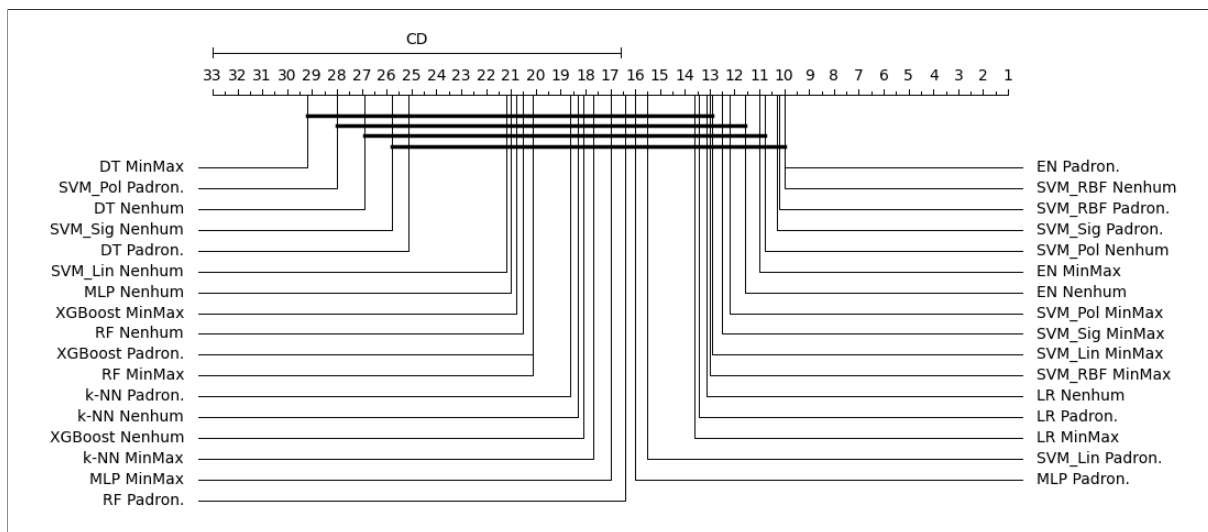
Fonte: elaborada pelo autor.

Tabela 33 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E3-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	2 (0,1)	0,7 (0,05)	1 (0,0)
Nenhum	EN	2 (0,4)	0,7 (0,03)	1 (0,0)
Nenhum	DT	3 (1,4)	0,7 (0,02)	2 (0,3)
Nenhum	MLP	356 (601,0)	0,7 (0,03)	7 (2,3)
Nenhum	k-NN	1 (0,1)	1,0 (0,03)	30 (0,0)
Nenhum	SVM_Lin	26 (10,8)	0,7 (0,03)	13 (0,5)
Nenhum	SVM_RBF	7 (2,8)	0,7 (0,01)	12 (0,2)
Nenhum	SVM_Pol	12 (8,0)	0,7 (0,03)	12 (0,3)
Nenhum	SVM_Sig	7 (5,7)	0,7 (0,02)	14 (0,9)
Nenhum	RF	677 (574,2)	14,1 (7,67)	342 (303,1)
Nenhum	XGBoost	154 (222,5)	1,5 (0,03)	89 (66,0)
MinMax	LR	2 (0,0)	0,8 (0,08)	2 (0,0)
MinMax	EN	3 (0,6)	0,8 (0,11)	2 (0,0)
MinMax	DT	5 (1,4)	0,8 (0,07)	3 (0,4)
MinMax	MLP	28 (20,4)	0,9 (0,11)	7 (2,5)
MinMax	k-NN	3 (1,3)	1,1 (0,14)	30 (0,0)
MinMax	SVM_Lin	8 (6,1)	0,9 (0,10)	13 (0,3)
MinMax	SVM_RBF	9 (5,7)	0,8 (0,02)	13 (0,3)
MinMax	SVM_Pol	6 (2,3)	0,8 (0,02)	13 (0,5)
MinMax	SVM_Sig	7 (1,8)	0,8 (0,03)	13 (0,3)
MinMax	RF	458 (209,2)	12,4 (3,75)	282 (213,4)
MinMax	XGBoost	68 (73,6)	2,0 (0,20)	64 (30,2)
Padronização	LR	2 (0,1)	0,8 (0,10)	2 (0,0)
Padronização	EN	3 (0,7)	0,8 (0,05)	2 (0,0)
Padronização	DT	5 (1,3)	0,8 (0,05)	3 (0,2)
Padronização	MLP	24 (11,8)	0,8 (0,07)	7 (1,7)
Padronização	k-NN	3 (0,6)	1,0 (0,03)	30 (0,0)
Padronização	SVM_Lin	10 (3,2)	0,8 (0,02)	13 (0,2)
Padronização	SVM_RBF	5 (1,5)	0,8 (0,02)	13 (0,3)
Padronização	SVM_Pol	4 (0,8)	0,9 (0,04)	15 (0,6)
Padronização	SVM_Sig	6 (1,5)	0,9 (0,13)	13 (0,2)
Padronização	RF	952 (439,5)	18,0 (4,80)	381 (219,2)
Padronização	XGBoost	155 (321,2)	2,6 (1,76)	70 (51,5)

Fonte: elaborada pelo autor.

Figura 26 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E3-M. (CD = 16,4)



Fonte: elaborada pelo autor.

Tabela 34 – Médias e desvios-padrão (em parênteses) das métricas de avaliação para as diferentes combinações de modelo e pré-processamento no E4-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	MAE (g/cm ²)	RMSE (g/cm ²)	MAPE (%)	ME (g/cm ²)	R ²
Nenhum	LR	0,0862 (0,0113)	0,1092 (0,0146)	6,69 (0,96)	0,2732 (0,0449)	0,1375 (0,1951)
Nenhum	EN	0,0869 (0,0112)	0,1097 (0,0144)	6,75 (0,96)	0,2768 (0,0457)	0,1302 (0,1861)
Nenhum	DT	0,0898 (0,0103)	0,1126 (0,0129)	6,96 (0,93)	0,2844 (0,0513)	0,0826 (0,1857)
Nenhum	MLP	0,0866 (0,0097)	0,1088 (0,0127)	6,72 (0,85)	0,2806 (0,0512)	0,1430 (0,1777)
Nenhum	k-NN	0,0869 (0,0112)	0,1089 (0,0143)	6,77 (0,97)	0,2765 (0,0489)	0,1507 (0,1473)
Nenhum	SVM_Lin	0,0871 (0,0126)	0,1101 (0,0167)	6,75 (1,03)	0,2710 (0,0458)	0,1271 (0,2019)
Nenhum	SVM_RBF	0,0853 (0,0102)	0,1073 (0,0130)	6,62 (0,89)	0,2754 (0,0451)	0,1687 (0,1652)
Nenhum	SVM_Pol	0,0862 (0,0114)	0,1098 (0,0142)	6,70 (0,99)	0,2813 (0,0456)	0,1329 (0,1653)
Nenhum	SVM_Sig	0,0918 (0,0136)	0,1160 (0,0184)	7,17 (1,17)	0,2863 (0,0450)	0,0318 (0,2099)
Nenhum	RF	0,0869 (0,0112)	0,1097 (0,0140)	6,73 (0,98)	0,2837 (0,0511)	0,1367 (0,1439)
Nenhum	XGBoost	0,0897 (0,0107)	0,1121 (0,0144)	6,94 (0,92)	0,2807 (0,0517)	0,0958 (0,1797)
MinMax	LR	0,0862 (0,0113)	0,1092 (0,0146)	6,69 (0,96)	0,2732 (0,0449)	0,1375 (0,1951)
MinMax	EN	0,0866 (0,0112)	0,1094 (0,0144)	6,73 (0,95)	0,2743 (0,0449)	0,1356 (0,1825)
MinMax	DT	0,0899 (0,0121)	0,1125 (0,0137)	6,98 (1,06)	0,2769 (0,0428)	0,0915 (0,1420)
MinMax	MLP	0,0867 (0,0096)	0,1086 (0,0130)	6,71 (0,83)	0,2796 (0,0532)	0,1416 (0,2005)
MinMax	k-NN	0,0878 (0,0117)	0,1102 (0,0146)	6,84 (1,02)	0,2755 (0,0489)	0,1324 (0,1424)
MinMax	SVM_Lin	0,0861 (0,0116)	0,1092 (0,0149)	6,69 (0,98)	0,2738 (0,0444)	0,1404 (0,1794)
MinMax	SVM_RBF	0,0861 (0,0113)	0,1079 (0,0143)	6,69 (0,96)	0,2788 (0,0450)	0,1616 (0,1674)
MinMax	SVM_Pol	0,0898 (0,0150)	0,1131 (0,0190)	7,01 (1,31)	0,2848 (0,0381)	0,0912 (0,1533)
MinMax	SVM_Sig	0,0861 (0,0116)	0,1089 (0,0151)	6,70 (0,97)	0,2758 (0,0431)	0,1423 (0,1965)
MinMax	RF	0,0872 (0,0110)	0,1098 (0,0140)	6,77 (0,96)	0,2836 (0,0496)	0,1366 (0,1408)
MinMax	XGBoost	0,0890 (0,0113)	0,1119 (0,0150)	6,89 (0,95)	0,2789 (0,0488)	0,0860 (0,2571)
Padronização	LR	0,0862 (0,0113)	0,1092 (0,0146)	6,69 (0,96)	0,2732 (0,0449)	0,1375 (0,1951)
Padronização	EN	0,0866 (0,0114)	0,1095 (0,0146)	6,73 (0,97)	0,2750 (0,0445)	0,1345 (0,1867)
Padronização	DT	0,0889 (0,0107)	0,1121 (0,0132)	6,90 (0,96)	0,2911 (0,0526)	0,0902 (0,1877)
Padronização	MLP	0,0858 (0,0101)	0,1085 (0,0134)	6,66 (0,88)	0,2809 (0,0499)	0,1464 (0,1950)
Padronização	k-NN	0,0880 (0,0119)	0,1101 (0,0148)	6,86 (1,03)	0,2764 (0,0470)	0,1335 (0,1411)
Padronização	SVM_Lin	0,0859 (0,0115)	0,1089 (0,0148)	6,67 (0,97)	0,2737 (0,0444)	0,1447 (0,1821)
Padronização	SVM_RBF	0,0866 (0,0106)	0,1085 (0,0139)	6,72 (0,89)	0,2804 (0,0465)	0,1484 (0,1854)
Padronização	SVM_Pol	0,0973 (0,0138)	0,1221 (0,0163)	7,60 (1,18)	0,2994 (0,0366)	-0,0612 (0,1376)
Padronização	SVM_Sig	0,0862 (0,0116)	0,1093 (0,0150)	6,70 (0,98)	0,2744 (0,0443)	0,1394 (0,1786)
Padronização	RF	0,0871 (0,0111)	0,1097 (0,0137)	6,76 (0,97)	0,2815 (0,0463)	0,1373 (0,1474)
Padronização	XGBoost	0,0887 (0,0104)	0,1109 (0,0136)	6,86 (0,92)	0,2808 (0,0458)	0,1062 (0,2100)

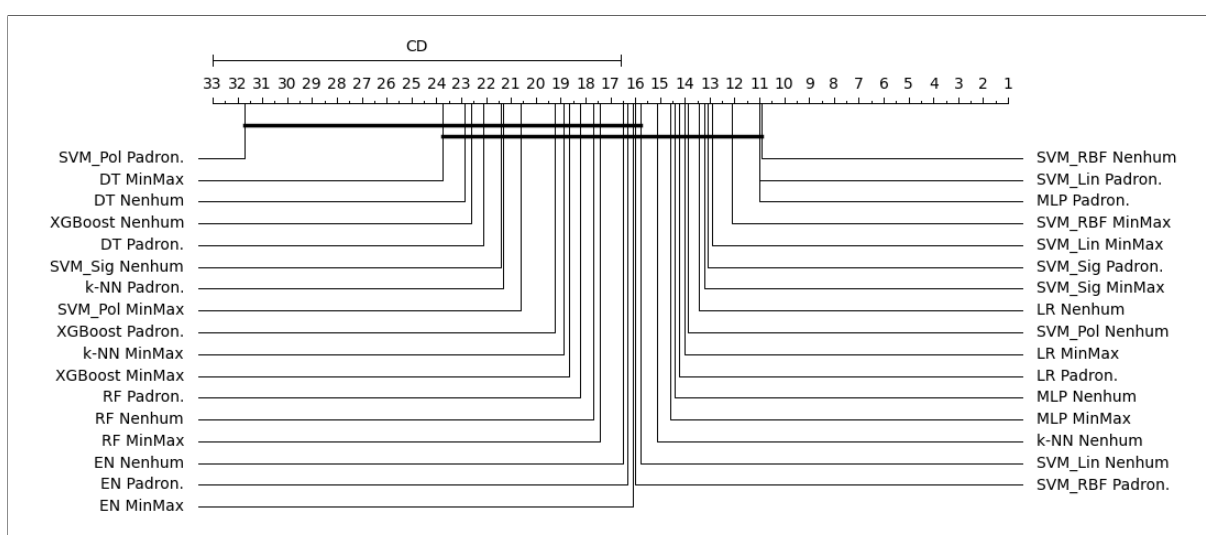
Fonte: elaborada pelo autor.

Tabela 35 – Médias e desvios-padrão (em parênteses) das métricas de recurso para as diferentes combinações de modelo e pré-processamento no E4-M. Em negrito estão as médias das métricas com valores mais próximos ao valor ideal. Em verde está a combinação escolhida como a mais adequada para este experimento.

Pré-processamento	Modelo	Ttr (ms)	Tte (ms)	TA (kB)
Nenhum	LR	1 (0,1)	0,7 (0,04)	1 (0,0)
Nenhum	EN	2 (0,1)	0,7 (0,04)	1 (0,0)
Nenhum	DT	2 (0,5)	0,7 (0,04)	2 (0,3)
Nenhum	MLP	79 (53,3)	0,7 (0,05)	5 (0,4)
Nenhum	k-NN	1 (0,2)	0,9 (0,05)	12 (0,0)
Nenhum	SVM_Lin	22 (9,0)	0,7 (0,02)	7 (0,2)
Nenhum	SVM_RBF	7 (3,4)	0,7 (0,02)	7 (0,1)
Nenhum	SVM_Pol	9 (6,4)	0,7 (0,02)	7 (0,2)
Nenhum	SVM_Sig	6 (3,3)	0,7 (0,03)	7 (0,5)
Nenhum	RF	212 (165,1)	10,1 (6,89)	137 (89,8)
Nenhum	XGBoost	180 (355,0)	1,9 (0,91)	111 (80,1)
MinMax	LR	3 (0,3)	0,8 (0,09)	1 (0,0)
MinMax	EN	3 (0,1)	0,8 (0,04)	2 (0,0)
MinMax	DT	4 (0,9)	0,8 (0,02)	2 (0,2)
MinMax	MLP	20 (13,0)	0,8 (0,03)	5 (0,7)
MinMax	k-NN	2 (0,0)	1,0 (0,03)	13 (0,0)
MinMax	SVM_Lin	7 (2,8)	0,8 (0,05)	7 (0,1)
MinMax	SVM_RBF	8 (4,2)	0,8 (0,03)	7 (0,2)
MinMax	SVM_Pol	6 (1,9)	0,8 (0,02)	8 (0,3)
MinMax	SVM_Sig	9 (4,2)	0,8 (0,03)	7 (0,1)
MinMax	RF	255 (209,2)	12,6 (7,90)	184 (151,6)
MinMax	XGBoost	47 (62,0)	1,8 (0,03)	64 (41,5)
Padronização	LR	3 (0,1)	0,8 (0,08)	1 (0,0)
Padronização	EN	3 (0,3)	0,8 (0,09)	1 (0,0)
Padronização	DT	4 (0,7)	0,8 (0,04)	2 (0,3)
Padronização	MLP	28 (20,5)	0,8 (0,03)	5 (0,4)
Padronização	k-NN	2 (0,2)	1,0 (0,05)	13 (0,0)
Padronização	SVM_Lin	7 (3,3)	0,8 (0,02)	7 (0,2)
Padronização	SVM_RBF	10 (6,3)	0,8 (0,01)	7 (0,1)
Padronização	SVM_Pol	5 (2,1)	0,8 (0,01)	8 (0,2)
Padronização	SVM_Sig	7 (2,8)	0,8 (0,02)	7 (0,2)
Padronização	RF	266 (165,3)	13,1 (5,60)	165 (69,8)
Padronização	XGBoost	222 (463,1)	2,6 (2,26)	57 (57,0)

Fonte: elaborada pelo autor.

Figura 27 – Diagrama de diferença crítica apresentando o rank médio de cada combinação, em relação aos valores de MAE do experimento E4-M. (CD = 16,4)



Fonte: elaborada pelo autor.